Running Head: Morphology and Frequency

Morphology and Frequency: Contrasting Methodologies

Michael A. Ford

William D. Marslen-Wilson

Matthew H. Davis

MRC Cognition and Brain Sciences Unit

15 Chaucer Road, Cambridge, UK.

Address for Correspondence:

Michael A Ford
MRC Cognition and Brain Sciences Unit
15 Chaucer Road
Cambridge
CB2 2EF
UK

Tel: +44 1223 355 294 (#311)
Fax: +44 1223 359 062
Email: mike.ford@mrc-cbu.cam.ac.uk

Abstract

Effects of root morpheme and whole-word frequency in lexical decision have been used to constrain theories of the representation and processing of morphologically complex words. However, results from these studies are unclear and somewhat contradictory, which may in part be due to an over-reliance on factorial designs. We present data from 2 experiments that show the utility of using correlational designs in addition to factorial designs. Experiment 1 used a regression design to explore, which of the many counts of word and morpheme frequency best predicts lexical decision response times to monomorphemic words. Results showed that word form frequency was the best predictor, though significant effects of morphological family size, cumulative morpheme frequency and the semantic coherence of the morphological family were also observed. In contrast to previous, factorially designed experiments, no clear effects of lemma frequency were found. Experiment 2 used both factorial and regression designs to probe directly for lemma frequency effects, using comparative (e.g., *stronger*) and superlative (e.g., *strongest*) adjectives. The use of the comparative also tested for affix homonymy effects, since -er is also an agentive suffix. No effects of affix homonymy were found, but clear lemma frequency effects were found in the regressions. Effects of more than one frequency count on response latencies are interpreted as reflecting the multiple sources of information that are used in making lexical decisions.

Introduction

Frequency and cognition

Frequency has been shown to strongly affect response times in a variety of linguistic tasks. With words, for example, the more frequent a word is, the faster subjects can classify it as word. In lexical decision, subjects are presented with words and non-words, either visually or auditorily and must make a speeded decision as to whether what they see or hear is a real word or not. Strong effects of word frequency have been consistently found in this task (e.g., Howes 1954). Indeed, this is one of the most robust results in experimental psycholinguistics and has played a crucial role in the development of theories of lexical representation and processing (Morton, 1969, Broadbent, 1967, Forster, 1976).

Frequency and morphology

However, there is a problem with word frequency in that it is unclear what the correct measure of frequency is. There exists a number of different ways of counting word frequency and these are not theory neutral. Word form frequency is the simplest and most obvious of these counts; a simple measure of the total number of tokens of a given word found in a language corpus. The other counts are morphemic, that is, they count the occurrence of root morphemes as opposed to surface word forms. Such counts will be emphasised by models in which morphologically complex forms are decomposed and therefore represented in terms of root morphemes.

One example is lemma frequency, this is the cumulative frequency of all the word form frequencies of words within an inflectional paradigm. The lemma frequency of the verb *help*, for example, is the sum of the word form frequencies of *help*, *helps*, *helped* and *helping*. In accounts of language processing in which regular inflectional forms are decomposed and map onto root morphemes, we would expect the frequency of the root to be more critical for determining response latencies than word form frequency and hence the lemma frequency would play a prominent role.

Accounts in which other complex forms are also decomposed (e.g., inflections, derivations and compounds) will instead emphasise the cumulative morpheme frequency, which is the sum of the frequencies of all the complex words in which a root morpheme appears. For example, the cumulative morpheme frequency of *help* would be the sum of the lemma frequency of *help* plus the lemma frequencies of *helpful, helpless, helplessness* etc. Another measure, family size, is the number of word types in which a morpheme occurs, rather than the number of tokens of it. The word 'help' has a family size of ten[1]. All of these counts have been found to influence reaction times in lexical decision. (Bradley, 1979, Meunier & Segui, 1999, Cole, Segui & Taft, 1997, Beauvillain, 1996, Baayen, Dijkstra & Schreuder 1997, Taft, 1979, Bertram, Hyönä & Laine 2000, Schreuder & Baayen 1997, De Jong, Schreuder & Baayen 2000, Giraudo & Grainger, 2000). (However, it is common for authors to emphasise the influence of one measure alone (tied a favoured lexical processing account)).

<u>Frequency and models of lexical representation</u>

[1] helpless, helplessly, helplessness, helpful, helpfulness, helpfully, helpmate, helpmeet, helping, self-help

The different frequency counts are linked with particular models of morphological processing. Models of lexical access differ in respect to which, if any, words will be decomposed into constituent morphemes and so which frequency count will determine response times. Many models make a distinction between inflectional and derivational morphology. Clahsen (2000) makes a clear distinction between inflectional and derivational processes, regular inflectional processes are symbolic, whereas derivational forms are stored in the lexicon as whole forms. Laudanna, Badecker and Caramazza (1992), claim that inflectional information is processed before derivational.

Inflectional morphology is typically seen as being regular, highly productive and governed by syntax. Derivational morphology is less paradigmatic and generally is used for creating new lexical items, often involving a change in word class, e.g., *dark-darkness*. The exact meaning of a derivational form is not necessarily as transparent as that of an inflectional form and derivational morphology is far less productive than inflectional. Although behavioural differences between inflectional and derivational morphology have been found (Feldman, 1994, Laudanna, et al, 1992), this division of morphological process is at best, prototypical, as there are cases where the categorisation of a particular process as either inflectional or derivational is not straightforward. For example there is some debate over the status of comparative and superlative adjectives and some have even questioned whether the distinction between inflection and derivation is justified (e.g., Lieber, 1981). Bertram and Baayen (2000) accept the notion that some words are processed as full forms and some are parsed, however they eschew the notion of a categorical distinction between inflection and derivation. Instead, they propose that the degree to which an affix alters the meaning of a stem, the productivity of an affix and the existence or not of another competing homonymic affix form, are the factors that determine whether a word is parsed or not, and not any simple distinction between inflection and derivation.

Sereno and Jongman (1997) propose a single system associationist account of morphology in which complex forms are represented as full forms in the mental lexicon. As words are processed as whole forms and not decomposed, only word form frequency is proposed to influence reaction time to words.

Other single system accounts take the opposing position that there is a single processing system with an obligatory morphological layer (although not all forms are decomposed). If all complex forms, inflections, derivations and compounds are represented in terms of their root morphemes then one would expect the frequency of the root to affect processing time. Meunier and Segui (1999) showed that suffixed forms in French with high cumulative morpheme frequency were responded to faster than those with low cumulative morpheme frequency (but see Baayen, Tweedie & Schreuder (2001) for a inhibitory effect of cumulative morpheme frequency).

Dual route models of lexical processing, influenced by linguistic accounts, propose that regular inflectional forms are not stored but are decomposed on-line, as they are completely predictable in form, meaning and syntactic role (Pinker, 1991, Clahsen, 2000). In such accounts, regularly inflected nouns and verbs, such as *cats* and *helped*, are not stored as full forms but are mapped onto abstract root morphemes. As these abstract morphemes are activated whenever any regularly inflected form is encountered, e.g., *help, helps, helped* and *helping*, it is the frequency of this root morpheme - its lemma frequency - that will influence response times to all the inflectional forms including stems. In such an account irregular inflections and derived forms are processed in a different manner.

Either they are stored as whole forms or they have a more complex representation of subnodes/subrules that give the relationship between the stem and irregularly inflected form (Wunderlich & Fabri, 1995). What such models would predict in terms of cumulative morpheme frequency and family size effects is unclear. Although De Jong et al (2000) showed that irregularity did not affect family size effects, proponents of dual route models have as yet not specifically addressed these phenomena.

Race models of morphology are a different kind of dual route account (Schreuder & Baayen 1997, Burani & Caramazza 1987). In these accounts there are again two processing routes but both are simultaneously involved in processing. One route processes the word as a whole form and the other route decomposes it into its constituent morphemes. The routes work in parallel and for any given word one or other route will win out. In lexical decision, responses will be made on the basis of which route wins out in the race. If the full form route wins out, then word form frequency effects should be expected, if the decompositional route wins out, then morphemic frequency effects should be expected. Such models are analogous to dual route models in other language domains, e.g., Cutler & Norris's (1979) race model of phoneme detection. As the decompositional route imposes extra processing costs, high frequency complex words will be processed as whole forms so these should show word form frequency effects. Low frequency complex words will be decomposed and are expected to show morphemic frequency effects, with for example, low frequency regular inflectional forms showing lemma effects.

The distributed connectionist approach to language processing posits a lexicon radically different from that usually conceptualised by psycholinguists. Words as such do not exist in the connectionist view of the lexicon. In such accounts (e.g., Plaut & Gonnerman, 2000), words are not single nodes but the co-activation of distributed phonological, orthographic and semantic features. The word bird is the co-activation of the phonemes /b/, /ɜː/ and /d/ along with semantic features, for example, "has wings", "flies", "has feathers" etc. When distributed connectionist simulations are trained on complex forms, the semantic and form consistency in morphologically related words has been shown to cause the hidden units, which mediate between form and semantics in such models, to represent the morphological structure in the training set (Rueckl & Raveh 1999). As models of this type are inherently sensitive to frequency (Seidenberg & McClelland, 1989), the frequency of root morphemes would therefore, be captured by such a model, so allowing for simulation of morphological frequency effects. Indeed, Davis et al (This Volume) successfully model Baayen et al's 1997 data showing word form and lemma frequency effects in singular and plural nouns within a single system distributed connectionist model.

Experimental investigations of morphology and frequency.

Having described various accounts of morphological representation and processing which make different predictions regarding the effect of word and morpheme frequency on response times in lexical decision, we now discuss methodological issues which may explain why experimental evidence to resolve this debate have so far proved somewhat inconclusive and inconsistent. These problems exist, at least in part because the different counts of frequency are highly intercorrelated.

Methodological issues with factorial designs.

Most work on morphology and frequency has used factorial designs, which test for significant differences between sets of words that differ on a single frequency variable while controlling for other frequency variables. For example, one might create two sets of words matched for lemma frequency, cumulative morpheme frequency and family size but differing in word form frequency to examine whether word form frequency affects response times, when other correlated counts are held constant. Such designs are used as they maximise the chance of finding an effect.

An intrinsic problem of this approach is that it can require the use of items that are unrepresentative of the language as a whole. A word that does not respect the correlations commonly found between the frequency variables is by definition atypical. Nevertheless, in order to create item sets for factorial experiments on word form and lemma frequency, experimenters must seek items from the extreme ends of the distribution. This approach therefore may overstate the importance of a particular variable. Lemma frequency and singular word form frequency are extremely highly correlated, $r = 0.96$, $p < 0.001$[2], therefore words where these variables are not closely matched are likely, although not necessarily, to be unrepresentative as they are at the extremes of the distribution of this correlation between the two frequency counts.

This problem of using extreme items is clearly illustrated by Bailey & Hahn (2001), who examined the role of phonotactic and lexical neighbourhood variables in the word-likeness of non-words. Compared to previous work, their results showed substantially weaker effects of trigram frequency. However, their stimuli set contrasted sharply with those of other researchers as they used a large random selection of non-words whereas previous work had used items from the ends of the trigram frequency distribution. In

[3] Based on same data as 2 above

analyses of a subset of their stimuli, from the extremes of their items' distribution, they showed results that were consistent with previous work.

Furthermore these atypical items for which the frequency counts are only weakly related naturally tend to be difficult to find, so the number of items in the experiment is often small, reducing statistical power. Clahsen (2000) found no difference in reaction times to high and low word form frequency German regular past participles, which were matched for lemma frequency. This null finding was interpreted as showing that for regular inflected forms only the lemma frequency is important. However, the cell size in this experiment was only 10 items, which could be why no significant difference was found. Similarly, Sereno and Jongman (1997) found no effect of lemma frequency in two experiments on English singular and plural nouns, based on a small sample size of 12 items per cell.

A related problem for factorial designs is that of sampling error in corpus frequency counts. To create the orthogonalized cells necessary for factorial designs it is necessary to sample from the extremes of distributions. However, as one tends toward the extreme of distributions potential sampling error increases. A word at the extreme of the distribution of a corpus frequency count (high or low) is more likely (in another corpus count) to be closer to the mean (Gernsbacher, 1984). That is, if one were to look at another corpus, it is likely that the low frequency words will have a higher frequency, that is, there will be regression to the mean.

This problem is one that could affect a number of experiments in which strong claims about the nature of lexical processing have been based on item sets which have frequencies of less than 1 per million (e.g., De Jong et al, 2000, Baayen et al, 1997). The matching of frequencies across conditions in these experiments could have been less accurate than the experimenters believed due to sampling error of the type discussed above.

A further problem is that results are often interpreted in an overly categorical manner. Baayen et al (1997) interpret their findings of word form frequency effects for plural nouns but lemma effects for singulars as showing that the frequency of the singular form has no effect on the plural form. Within the Morphological Race Model, there are full form access representations and as the singular is embedded within the plural activation of the plural will feedback to the singular but not vice versa. However, Baayen et al's results cannot exclude the possibility of there being a relative effect of the singular form frequency on plurals. That is to say the effect of the frequency of morphologically related words need not be strictly additive. In Davis et al's (This Volume) connectionist model of Baayen et al's data, although they show the differential word form and lemma frequency effects found in the behavioural data, these are relative effects with both counts playing a role in reaction times to both forms. The consistency of form and meaning between stems and plurals will mean that in a distributed model they will share representational space so the frequency of one form will necessarily influence the processing of the other.

Furthermore, Baayen, Tweedie & Schreuder (2001) find that both word form and lemma frequency are predictive of response times in reanalysis of data from Schreuder & Baayen (1997) using linear mixed effects modelling to increase the statistical power of their analyses (Pinheiro & Bates, 2000).

It is clear then, that it would be beneficial to our understanding of the roles of different frequency counts in lexical decision to use more typical items. This does not necessarily mean that we cannot make use of these more extreme items, but it does mean that we should be wary of strong claims based on such items.

By using correlational designs such as multiple regression, we can examine the degree of influence of each of the frequency variables. Moreover, by using a correlational approach we can avoid some of the methodological problems of factorial designs.

Motivation for Experiment 1

Although factorial designs are good tools for investigating whether a particular variable influences reaction time, they cannot tell us about the degree of influence of this variable. All the frequency variables may influence reaction time but to different extents, with one count being most predictive of reaction time. By using a correlational approach we can quantify which variables have the greatest influence on reaction time. Moreover, by using this kind of design, we do not need to try to maximise the difference between lemma and word form frequencies in order to increase the likelihood of finding an effect, allowing for the use of a larger and more typical item set.

In experiment 1 we will examine how well a number of variables predict response times to words: lemma frequency, word form frequency, cumulative morpheme frequency, 4 different measures of family size, and a measure of the semantic coherence of the morphological family of a word. We claim that the use of possibly atypical items and/or extreme items, which are probably subject to sampling error, has possibly caused the over-stating of the importance of some variables. For example, De Jong et al (2000) found that for a set of 66 inflected verbs family size accounted for 19% of variance in response times. These were a subset of the set of possible items and were selected such that family size and the two frequency counts did not correlate significantly. The advantage of this was that it allowed De Jong et al to examine the influence of family size uncontaminated by the usual correlation with frequency, however, it may be the case that for most such items family size would normally account for less variance.

As no attempt was made to orthogonalize the various frequency variables the items follow the language in having a high degree of intercorrelation between counts. The high degree of correlation between the frequency counts could be a problem in a regression equation as it may cause multi-collinearity (Chatterjee, Hadi and Price, 2000). However, multivariate techniques (e.g., principal components analysis) can in part overcome this problem (see below). It is important to note that this experimental approach is meant to complement and not to replace the factorial designs.

In Experiment 2, both factorial and correlational statistics are used in attempt to detect lemma frequency with optimal sensitivity.

Experiment 1

Methods

Participants

26 volunteers from the MRC-CBU volunteer panel participated in Experiment 1. All were native speakers of British English, aged 17-45 years and reported no language problems. All volunteers had normal hearing and normal or corrected to normal eyesight.

<u>Materials</u>

300 monomorphemic English nouns were selected such that they were either unequivocally nouns (no other form existed) or had greater frequency as nouns than as other forms (mean noun:verb frequency ratio, 22:1) in the CELEX database (Baayen, Piepenbrock & van Rijn, 1993)(See Appendix 1 for stimuli). Homophonic words were excluded by checking all words in the Wordsmyth on-line dictionary (1998). All words were 3-6 characters in length and were predominantly monosyllabic. Frequency data for the words was obtained from the CELEX database. For each word the Cumulative Morpheme Frequency, Lemma Frequency, and Word Form Frequency was obtained. Four family size measures were also obtained, FS 1, a count of all the derivational forms of each test item, FS 2, a count of all derivational and compound forms, FS 3, a count of derivational forms and the inflections of these derivations, and FS 4, a count of derivations, compounds and all the inflectional variants of these complex forms. Both semantically transparent and opaque morphological family members were included. FS 2 has previously been the predominant measure of family size (Schreuder & Baayen, 1997). However as with the other frequency counts what is included in family size counts is not theory-neutral. Whether inflectional variants should be included in a count of family size is in part dependent on how such forms may be represented in the lexicon. Moreover, most work on family size has been in Dutch, where compounding is a more productive feature of the language than in English. Other counts may be more predictive for

English. The count of compounds did not include compounds with spaces between the words. De Jong, Feldman, Schreuder, Pastizzo, & Baayen (2001) found that family size counts that excluded such items better predicted response times.

To provide an estimate of sampling error, an alternative frequency count was also obtained from the British National Corpus (1995). Bigram and trigram frequencies as well as orthographic neighbourhood density, Coltheart's N, were also calculated for each item using the CELEX database (Coltheart, 1977).

De Jong et al (This Volume) present data indicating that the family size effect has a strong semantic component. To explore semantic aspects of family size, we included a variable that measures the degree of relatedness of our items to their morphological families. Given that many of our items have extremely high family sizes, it was infeasible to get subjective semantic relatedness ratings of each stem to all its family members. Instead we used a measure based on Latent Semantic Analysis (LSA) vectors (Landauer & Dumais, 1997). LSA vectors are derived from corpus based co-occurrence matrices, that is, how likely sets of words are to occur in the same discourse context. The cosine of an angle between vectors representing morphologically related words in LSA space correlates well with subjective semantic relatedness ratings of word pairs (Rastle, Davis, Marslen-Wilson, & Tyler, 2000). For each word in the experiment a mean family LSA score was calculated, by taking the mean of this LSA cosine measure between the stem paired with each of its family members[3].

<Table 1 About Here>

[3] Not all family members found in the CELEX database were present in the corpus used for LSA, so this score is based on a subset of the family size. It was the lower frequency words that were not in the LSA corpus.

Table 1 gives details of the words used in the experiment. Words were chosen to represent a wide range of frequencies and family sizes in order to maximise the sensitivity of the regression design.

As there is a high degree of correlation within our frequency and family size variables, which could cause problems of collinearity in regression analyses, we collapsed these into a single frequency and a single family size variable using principal components analyses (PCA). Oblique rotation (direct oblimin) was used and for each analysis there was a single component with an eigenvalue greater than one (Kaiser, 1960)[4]. The item weightings for each of the components were saved for use as variables in the analyses.

Although family size measures and frequencies also correlate, previous work has shown that frequency and family size effects are due to separate and different underlying causes. We have used this theory driven approach to constraining our use of PCA (Chatterjee et al, 2000). Moreover, PCA on all the family size and frequency variables together produced two components, one weighted to the frequency variables and one to family size variables, so justifying our use of two separate PCA analyses.

300 non-words were created that were matched in length and bigram frequency to real words. All non-words were "monomorphemic", that is, they did not include any affixes.

Procedure

[4] If a factor has an eigenvalue of less than one, then it accounts for less variance than any one of the original variables and so is not used.

A single word lexical decision task was used. Items were presented to volunteers on a computer monitor using the DMDX experimental software, developed at Monash University and at the University of Arizona by K.I.Forster and J.C.Forster. Each item was preceded for 250 ms with a "+" fixation point followed at offset by the item itself for 500 ms in upper case 14 point Arial font. Participants were instructed to make a lexical decision to items as quickly and accurately as they could. Responses were indicated by means of pressing one of two keys on a button box. The 'yes' response key was always pressed with the dominant hand, the 'no' with the non-dominant hand. Subjects had 2000 milliseconds to respond before the system timed out and went onto the next item. There was a minimum inter-trial interval of 2000 milliseconds.

Each subject saw a different pseudo-randomised order of the list, arranged into thirty twenty-item blocks. Each block had a fixed order across subjects, but the blocks themselves were presented in a random order for each subject. This allowed for an equal spread of experimental stimuli across the length of the list while avoiding possible order effects.

Results.

2 subjects were removed from the analyses due to high error rates (over 15%). Data from 6 words was removed from the analyses due to high error rates (over 30%). This left 24 subjects and 294 words. Mean reaction times were log transformed and entered in analyses.

Analyses

Where possible variables were centred to reduce possible multicollinearity (Chatterjee et al, 2000). However, the high inter-correlation of a number of variables still created problems of multi-collinearity in the regressions when all the variables were included. This was overcome by carrying out a series of multiple regressions with the reduced number of variables created by principal components analysis and also by excluding certain variables that were neither significant predictors nor theoretically interesting to this work but caused problems of collinearity (for example, trigram frequency).

Frequency and family size variables were logged. Cumulative morpheme frequency as normally measured, that is the log of the sum of all the morphological family measures was particularly susceptible to multi-collinearity. Ridge regression, a technique that examines the stability of regression co-efficients showed that the standard measure of cumulative morpheme frequency to be unstable. The addition of a constant to the regression equation caused large changes in the regression co-efficient for cumulative morpheme frequency. However, by logging the frequencies of the family members and then, summing the logs, a single stable variable was created that was not collinear with other variables. By logging and them summing this measure will enhance the influence of the lower frequency items of a morphological family. This cumulative logged frequency variable was centred and entered into the regressions.

The model that accounted for the greatest total variance was as follows. Word form frequency was the strongest predictor response times, $R^2 = 0.33$. Cumulative logged morpheme frequency added to the explanatory power of the analysis, change in $R^2 = 0.03$, significance of F change < 0.001. However, this was a positive correlation, that is, the higher the cumulative logged frequency, the slower the reaction time. Family Size 2, the measure based on a type count of all the derived and compound members of a family was also a significant predictor of response times, $R^2 = 0.041$, significance of F change < 0.001. The LSA semantic variable again added to the explanatory power of the model, correlating negatively with response times, change in $R^2 = 0.01$, significance of F change < 0.05. There was no evidence of serious multiple-collinearity in these data, with all variables having variable inflation factor scores of less than 2.5 and the condition index for the model equal to 2.9 (Chatterjee et al, 2000).

For the frequency variables, even after centring, multicollinearity was still present when all 3 original frequency counts were included. In separate regressions, in which the other frequency variables were fitted individually, Lemma Frequency, BNC word form frequency and the PCA variable made from the three frequency variables, all accounted for less variance than did Word Form Frequency when it was fitted. In models in which the other family size variables were used these were not significant predictors of reaction time. A single family size variable created using principal components analysis was a significant predictor of response times but explained less variance than the Family Size 2 variable on its' own.

In order to include both word form and lemma frequency within the same model a form of 'partial orthogonalization' was used (Draper and Smith 1966, 1981). Word form frequency was regressed against lemma frequency and the residual values saved as measure of word form frequency adjusted for lemma frequency. Likewise, lemma frequency was regressed against word form frequency to give a measure of lemma frequency adjusted for word form frequency. Response times were then analysed using hierarchical sequentially structured regression models. By adding the adjusted and unadjusted predictors in different orders we explored the relative explanatory value of Word Form and Lemma Frequency while minimising ambiguity that might arise from collinearity. When word form frequency was fitted first, adding lemma frequency adjusted for word form frequency did not significantly increase the model's explanatory power, (change in $R^2 < 0.001$, significance of F Change = 0.714). In the second model, when lemma frequency was fitted first, adding word form frequency adjusted for lemma frequency significantly increased the explanatory power of the model, (change in $R^2 = 0.018$, significance of F change < 0.005). Again there was no evidence of serious problems of collinearity in these data, with no variable inflation factor (VIF) greater than 2.7 in either model and a condition index of 6.8 for both[5]. These data suggest that word form frequency is a better predictor of response times than lemma frequency and that the significant effects of lemma frequency when included in the absence of word form frequency are due to its correlation with word form frequency.

Discussion

Experiment 1 showed that word form frequency is the strongest single predictor of lexical decision reaction times but these are also co-determined by cumulative morpheme frequency, family size and the semantic coherence of the morphological family.

Within the frequency domain, the word form frequency is a better predictor of response times than lemma frequency. Both a model with this variable alone, and a model that includes a PCA variable created from all the frequency counts account for more variance than the model with lemma frequency alone. Hierarchical regressions suggest lemma frequency adds no explanatory power to the models, but as there are problems of multicollinearity when both measures are included in analyses, this must remain a tentative conclusion.

---

5 Inspection of the variance proportions shows some collinearity in the form variables, length, bigram frequency and trigram frequency.

The finding of a positive correlation of cumulative morpheme frequency with response times is in line with Baayen et al (2001). By using linear mixed effects modelling (Pinheiro & Bates, 2000), they increased the power of their analyses and were able to find an effect of cumulative morpheme frequency that had been undetected by their original analyses of these data (Schreuder & Baayen, 1997). The use of 294 items in our regression analysis gave us a high degree of statistical power, making the effect detectable here. Baayen et al interpreted the negative effect of cumulative morpheme frequency as deriving from lexical competition between access representations of family members.

Meunier and Segui (1999) found that the relative frequency of family members played a role in lexical decision latencies to auditory stimuli. Words with more frequent morphological family members were responded to slower, than those with lower frequency words. Although, for a small number of items, there was a derived form with higher frequency than the stem, for the items in this experiment, the stem was the highest frequency form in the morphological family. Moreover, with a logged then summed cumulative morpheme frequency count, the lower frequencies family members will have a greater influence, compared to the measure used by Meunier and Segui which will be dominated by the highest frequency members. Unlike the model proposed by Meunier and Segui, morphologically related forms do not need to be higher in frequency than the family member to cause competition.

It is unclear as yet whether this is morphological or a lexical effect. That is, is it competition in a lexical cohort with all word forms containing the root, whether morphologically related or not, activated and competing with each other, (e.g., for the root *law* both *lawn* and *lawful)*, or is competition restricted to morphologically related forms? This is clearly something that can be tested.

These data add to the growing body of work showing that morphological family size is a significant predictor of lexical decision response times. For any given word, the type count of other complex words sharing a root morpheme is predictive of response time to that word. Note that it was the Family Size 2, the count that includes only derivations and compounds, that was significant, in line with Schreuder & Baayen (1997). Including inflectional variants in the family size count does not seem warranted based on our data. The approximately 4% of variance explained by family size, was, however, substantially less than that found by De Jong et al (2000). However, in contrast to their stimulus there was a strong correlation here between word form frequency and family size, which may have led to an underestimation of the predictive power of family size. Nevertheless, it is clear that word form frequency is a stronger predictor of response times than family size.

De Jong et al (This Volume, 2001) found strong semantic effects in their work on family size. As inflections add little to the semantics of their root forms compared to derivations and compounds their data support a strong role for semantics in family size effects. In support of this, we found in addition to the family size effects, that the semantic coherence within the morphological families also significantly added to the predictive power of the regression equation, with separable effects of semantics and morphological family size. However, this effect, although reliable was quite small and may suffer from sampling error. The relatively small size of the corpus used for LSA could have caused the number of missing family members to vary over items. Nonetheless, this result suggests tools like LSA may prove useful in future investigation of the role of semantic relationships within morphological families

In contrast to previous work (Baayen et al, 1997, Taft, 1979), which found clear effects of lemma frequency, we found that word form frequency best predicted response times to singular nouns. This seems to support full listing models of morphology, as does the apparent competition between morphological relatives as indicated by the cumulative morpheme data. However, if indeed we have a separate pure semantic effect and a family size effect based solely on the sharing of a root morpheme, then these data sit best with a model of the lexicon in which there are morphemic representations. Moreover, as we did not orthogonalize our lemma and word form frequency variables, the experiment was not ideally designed to find separable effects of these counts. If as Davis et al's (This Volume) model suggests, the response times to words, although dominated by either lemma or word form frequency are in fact a product of both counts, then for typical nouns, it may be difficult to find lemma frequency effects. For typical nouns, the word form frequency of the plural adds relatively little to the lemma frequency of the word (the average frequency ratio singular to plural is approximately 3:1). Given that regression analyses were used, a small lemma effect that is highly correlated with a strong word form effect for singular nouns is unlikely to be significant, even if problems of collinearity were dealt with. Moreover, this implies that contrary to Baayen et al (1997), for typical nouns, any lemma frequency effects are more likely to be found on inflected forms. Firstly, the correlation between word form and lemma frequency is less for plurals ($r = 0.73$), than singulars ($r = 0.96$[6]). Secondly, the difference between lemma and word form frequency is greater with plurals. If there are only small, non-additive effects of lemma frequency, then it will be statistically easier to distinguish these from word form frequency effects by using noun plurals.

---

[6] Same data as footnote 2 above

Experiment 2

Although Experiment 1 showed no clear effects of lemma and word form frequency in the processing of 300 simple nouns, the items used in this experiment may not be ideal to find such effects. In order to explore differential effects of these counts researchers have generally made use of items where there is a large divergence between lemma and word form frequency. However, as discussed above, the items with maximal divergence are small in number, likely to be susceptible to sampling error and possibly atypical, so making the use of nouns here possibly problematic. In Experiment 2, we avoid these potential problems by using comparative and superlative adjectives, for which the typical pattern is that they have much lower word form frequency than their stems. On average the ratio of stem to comparative is 5.2:1 and stem to superlative is approximately 5.6:1[7]. As this is the typical pattern, it is not necessary to select from the extreme ends of the frequency distributions and we are less restricted in number of items.

In choosing to use adjectives of degree, however, we have changed the word class from nouns to adjectives and from simple to complex forms. However, the goal of this experiment is to explore whether we can find clear evidence of lemma frequency, without having to use atypical items, so adjectives of degree are well suited for this. Symbolic theories of language (e.g., Clahsen, 2000, Pinker, 1991) propose that a single mechanism underlies all regular inflectional processing. Clahsen (2000) provides evidence to support this by showing similar priming results and frequency effect results for both verbal and nominal regular inflections. By extension the same kind of processing should occur for adjectival inflections. Although for complex items there may be some processing time for stripping off the suffixes, this should be the same for both forms of the adjective.

[7] Based on 981 stem/comparative/superlative triplets, including irregulars from CELEX.

Bertram, Schreuder & Baayen (2000) set out a series of criteria by which one can predict whether a complex form will show lemma or word form frequency. They propose that complex forms that are productive, do not change the meaning of the stem, and do not have a competing homonymic affix, should show lemma frequency effects. Both forms of the adjective fulfil the first two criteria but as the comparative has a competing homonymic form it should show word form frequency effects, so making these items well suited to explore lemma and word form frequency effects.

However, there is some debate whether comparatives and superlatives are indeed inflectional. Although they fulfil some of the criteria usually given for inflection they are often argued to fall at the borderline between the inflectional and derivational categories. One criterion for inflection that is commonly used is that inflectional suffixes should occur after derivational, (Perlmutter, 1988) and in some languages (e.g., Italian) derivational affixes can attach after the comparative suffix. However, this is not the case for English in which comparative and superlative suffixes come after any derivational affix, e.g., dirtier (dirt + y + er) as do verbal and nominal inflections (e.g., dark + en + ed; govern + ment + s). Moreover, adjectives of degree serve to relate other syntactic constituents within an utterance, that is, they encode abstract "phrase level properties and relationships" between syntactic constituents (Stump 1998, Plank, 1994). Although there are two forms of the adjectives of degree, their use is obligatory in the relevant syntactic contexts and cannot be replaced by a morphologically simple form. Moreover, the choice of whether to use the affixal or periphrastic form (more/most plus adjective) of the adjectives of degree is not arbitrary but rather systematic. Plank (1994) proposes that inflectional morphological variants should represent the same basic concept as the base form. The adjectives of degree serve to represent the same concept in a relational manner. Indeed, dictionaries do not list comparative and superlative adjectives, as their meaning and form is completely predictable, as with plurals, past tenses etc.

Furthermore in behavioural work, unpublished data by Marslen-Wilson & Ford shows them to pattern like inflectional items. Marslen-Wilson, Ford, Older & Zhou (1996) showed clear priming between productive derivational affixes, such as *–ness* and *re-*, that is, *darkness* primed *toughness*, *rebuild* primed *reread*. Marslen-Wilson & Ford further examined whether this would also be the case for inflectional affixes, in particular the past tense *–ed* and superlative *–est*, asking whether *jumped* would prime *mended* or *happiest* would prime *darkest*. In 5 different priming studies, no priming between past tenses and superlatives was found. In addition, Kytö & Romaine (1997) found that in the British National Corpus the affixal form of the adjectives of degree were far more frequent than the periphrastic forms, comprising 84% of comparatives and 73% of superlatives. Moreover in comparing usage in the BNC with a number of historical corpora of English it appears that there is a diachronic trend toward the inflectional forms of the adjectives of degree, implying that such items are becoming more productive. In summary although the division of morphology into inflection and derivation is not without problems, the available behavioural and linguistic evidence leads us to conclude that the comparative and superlative adjectives fit best within the inflectional category.

These adjectival endings are also of interest because the comparative *–er* is homonymic with the deverbal agentive *–er*, whereas the superlative *–est* is a unique form. Bertram, Schreuder and Baayen (2000) surveyed frequency effects for a number of suffixed forms in Dutch. They reported that effects of word form frequency were primarily observed for items with homonymous affixes such as *–er*, which performs the same functions in Dutch as in English. The affixes that showed effects of lemma frequency were consistently those without competing homonymic interpretation, e.g., the deadjectival suffix -*heid*.

Bertram et al argue that the ambiguity due to affix homonymy causes the parsing route to be slowed allowing word form frequency effects to emerge. On the basis of this prediction, one would expect the comparative to show only word form effects but the superlative to show effects of its lemma frequency. As with Bertram et al's (2000) replication in Finnish of their findings in Dutch, the balance between the competing homonymic interpretation of English –*er* is fairly equal, with one form being inflectional and the other derivational, and the full forms belonging to a different syntactic class. Although the comparative is more frequent in terms of tokens (comparative 1401/million, superlative tokens = 423/million), they are approximately equal in terms of mean item frequency (comparative = 1.6/million, superlative = 0.5/million). Also, the comparative not only has a competing homophonic affix form, but it also has a large number of pseudo-affixed forms (approximately ten times the number of pseudo-affixed –*est* forms: 900 pseudo affixed –*er* versus forms 90 –*est* forms). The existence of many pseudo-affixed forms has also been suggested as a factor that makes full form processing more likely (Laudanna, Burani, & Cermele 1994). This makes these items ideal to test Bertram et al's claim.

Moreover, the correlation between the comparative word form and lemma frequency is r = 0.69 and between the superlative word form and lemma frequency is r = 0.50. These lower correlations should allow us to use both frequency counts as regressor variables without such acute problems of multicollinearity.

In Experiment two we use both between and within items factorial analyses and regression analyses to explore frequency effects with the adjectives of degree.

Methods

Participants

33 participants from the MRC-CBU Volunteer Panel took part in the experiment. All subjects were native British English speakers and reported no language problems. All subjects had normal hearing and normal or corrected to normal eyesight.

Materials

A nested repeated measures design was used with 3 experimental conditions. Condition 1 contained 60 comparative/superlative pairs, e.g., darker/darkest. As this manipulation was within items these items were necessarily matched for lemma frequency and root morpheme family size variables (See Appendix 2 for stimuli). All words were bisyllabic and the mean length for the comparatives was 6.3 characters and for the superlatives 7.3. CELEX lemma frequency for the adjectives was 130/million and mean word form frequency per million was 8/million for the comparatives and 3/million for the superlatives.

Two other conditions were included, one matched to the lemma frequency of the comparative and superlative adjectives and one matched to their word form. In order to match these items to the frequencies and number of characters in Condition 1 these conditions contained words of different word classes. Condition 2 contained 30 monomorphemic bisyllabic words matched in CELEX lemma frequency to the adjectives (mean lemma frequency = 127/million, mean word form frequency = 78, mean family size = 5.9) and in length to the comparatives (mean number of characters = 6.2) (25 nouns, 2 verbs, 3 adjectives). Condition 3 contained 30 monomorphemic bisyllabic words matched in CELEX word form frequency of the adjectives (mean lemma frequency = 7/million, mean word form frequency = 5/million, mean family size = 1.4) and in length to the comparatives (mean number of characters = 6.3) (22 nouns, 5 verbs, 3 adjectives). These conditions were included to compare response times to monomorphemic item matched to the different frequencies with our adjectives, allowing us to estimate how fast we might expect the adjectives should be responded to if either lemma or word frequency alone determined response times.

In addition to the test conditions there were also 120 real word fillers, consisting of 30 past tense inflected verbs, 30 plural inflected nouns and 60 monosyllabic simple words. 240 non-words were included and were created by changing one or two letters of real words. These were matched in length and syllables to the real words. There were 30 pseudo-comparative and 30 pseudo-superlative foils. In addition to this there were 60 bisyllabic non-words matched to conditions 2 and 3. 30 pseudo-past tense, 30 pseudo-plural and 60 monosyllabic "simple" non-words were also included. Condition 1 was rotated to form two lists with 30 comparative and 30 superlative items in each list. All other items were identical across the two lists. Items were pseudo-randomised. A practice block of 10 real words and 10 non-words preceded the main part of the experiment. The experiment was divided into 5 blocks of 96 items, with each block beginning with 4 dummy items (2 words/2 non-words).

Procedure

The procedure was as in Experiment 1.

Results

Data from 8 participants were removed due to high error rates (over 15%), 4 from each version, which left 13 in version 1 and 12 in version 2. Six items were removed from the analyses due to high error rates (over 30% on both versions or over 40% in one version), 5 from Condition 1 and 1 from Condition 3. As before the remaining mean response times were log transformed and entered into analyses.

<Insert Table 2 about here>

Analyses.

Table 2 shows mean RT and error rates for the three conditions. The by subjects analysis used a standard repeated measures multivariate analysis of variance. Repeated measures analyses of variance on the adjective condition alone showed no effect of affix homonymy on response times to comparatives and superlatives, $F1(1,23) < 1$, $F2(1,52) < 1$[8]. However, this was not the case with the errors, with significantly more errors for the comparatives than for the superlatives, $F1(1,23) = 3.6$, $P = 0.07$, $F2(1,52) = 4.4$, $P < 0.05$.

[8] The small but possibly significant difference in frequency between the comparatives and superlatives, was not significant as a main effect nor in any interactions when used as a covariate in an analysis of covariance.

As condition one included a within items comparison that does not occur in conditions 2 and 3, items analyses were done by means of a nested repeated measures design. There was a main effect of condition, indicating that the 3 conditions were responded to differently, $F1(3,69) = 27.5$, $P < 0.001$, $F2(2,106) = 9.1$, $P < 0.001$. This was also significant in the errors analyses by subjects, $F1(3,69) = 4.3$, $P < 0.01$ and marginal by items, $F2(2,106) = 2.9$, $P = 0.06$.

Post hoc tests on reaction time with family size as a covariate on the items data showed that the comparatives and the superlatives were responded to significantly more slowly than the lemma matched condition ($P < 0.05$) but did not differ from the word form matched condition ($P > 0.1$)

Although a relatively small number of adjectives were used in the experiment, a power analysis showed that there were enough items to produce reliable results for a regression model. Again, frequency and family size variables were centred. As in Experiment 1 hierarchical sequentially structured regression models were carried out. Again we added the adjusted and unadjusted predictors in different orders to explore the relative explanatory value of Word Form and Lemma Frequency while minimising ambiguity that might arise from collinearity. For both the comparatives and superlatives, when lemma frequency was fitted first, adding word form frequency adjusted for lemma frequency did not add to the explanatory power of the model, (comparatives - change in $R^2 < 0.005$, significance of F change > 0.1, superlatives - change in $R^2 < 0.01$, significance of F change > 0.1). Conversely, when we fit word form frequency first, adding lemma frequency adjusted for word form frequency does significantly add to the explanatory power of the model (comparatives - change in $R^2 = 0.145$, significance of F change <

0.005, superlatives - change in R^2 = 0.082, significance of F change < 0.02). In all the analyses no variable inflation factors was greater than 2, and no condition index was greater than 4. These data suggest that, for the comparatives and superlatives, it is lemma frequency that is the best predictor of response times.

Discussion

This experiment examined whether response times to low word form frequency superlative and comparative adjectives would be differentially affected by their high lemma frequency. The existence of affix homonymy for the comparatives did not influence their reaction times relative to the superlatives, although the error data is in line with Bertram et al (2000).

The between items comparison with response times to Conditions 2 and 3 show that the superlative adjectives were responded to significantly slower than might be predicted on the basis of their lemma frequency alone but not significantly different to items matched in word form frequency. However, regression analyses showed that response times to both the comparatives and superlatives were best explained by their lemma frequency. These data are in contrast to Bertram et al (2000) on the Dutch comparative –er suffix, for which they found in their regression analyses that responses were predicted only by word form frequency. The reason for the difference is unclear. One possibility is that the it is due to differences between Dutch and English, but as Bertram et al found similar results for a set of Finnish suffixes, where there are much greater language differences, this would seem unlikely. Indeed, the –er suffixes are used identically in Dutch and English, with the deverbal agentive –er being productive in both languages. However, there may be subtle differences in lexical statistics, syntax or semantics between English and Dutch. The criteria used by Bertram et al (2000) to predict full form or parsing are scalar but not necessarily linear. Small differences on these scales could lead to large differences in processing.

The discrepancy between the ANOVA results is might be thought somewhat curious. However, Conditions 2 and 3 contained simple words, predominantly nouns. Inflected forms that are marked for agreement with other words or encode relationships between other words are perhaps somewhat strange when presented in isolation. Moreover, if, as the lemma frequency effects with the adjectives suggest, these items are being parsed, then this parsing process may slow response times relative to simple forms. This is clearly something that can be further explored.

General Discussion

Experiment 1 showed that for singular nouns in visual lexical decision that the strongest predictor of response times is word frequency. Furthermore it was word form frequency and not lemma frequency that appeared to fit response times best. However, considering the data and analyses used, these data cannot rule out lemma frequency effects on singular nouns. Both family size and cumulative morpheme frequency were also significant predictors of response times, in line with other work (Baayen et al, 2001, Schreuder and Baayen, 1997). The semantic coherence of the morphological family was also a significant predictor of reaction time.

Experiment 2 showed no difference in response times to comparative and superlative adjectives. The existence of a competing homonymic form for the comparative did not cause a difference in response times relative to the superlative, although there were effects of this in the errors. Response times to both these forms were not different from a comparison condition matched to the word form frequency of the adjectives, although they were significantly slower than the condition matched to the adjectives in lemma frequency. Yet response times to both types of adjectives were predicted by lemma frequency.

Experiments 1 and 2 show frequency is by far the strongest predictor of response times (cf Whaley, 1978). Although the family size, the semantic variable and cumulative morpheme frequency in Experiment 1 they add significantly to the explained variance of the regressions, compared to the frequency alone they add relatively little. By selecting data sets that only differ on one of these two variables, one might find significant differences in response times. Nevertheless, the data presented here suggest caution in interpreting such data too strongly. Reliance on factorial designs alone may lead one to overstating of the importance of variables.

The positive correlation of cumulative morpheme frequency with response times in both experiments supports Baayen et al (2001). It would appear that we have simultaneous facilitation from morphemic and semantic levels of representation while at the same time there is competition between word form representations. Items with many frequent morphological family members were slowed relative to those with few.

Experiment 1 showed a new and interesting result, that is, a small but significant effect of the semantic coherence of the morphological family in addition to family size. De Jong et al (2000) proposed that the family size effect is both semantic and morphosyntactic, in that the type of affix attached to the stem governs which family members are activated. For example, in the context of the verbal inflectional suffix –t only verbal family members are activated. Our data suggest that the use of measures of semantic coherence within a morphological family may allow use to partial out semantic from morphological effects in family size effects.

The family size effect comes from the fact that the existence of many highly related forms mapping onto approximately the same semantics speeds response times. Interestingly this could be related to similar effects found with words with many senses (Rodd, Gaskell & Marslen-Wilson, 2000). Previous work examining the effects of homophony has produced conflicting results, although some researchers have found that response times to homophones relative to non-homophonic forms were speeded, others have found them slowed (Rubenstein, Garfield & Millikan, 1970, Azuma & van Orden, 1997). Rodd, et al (2000) examined lexical decision to homophones. However, in contrast to previous work, they distinguished between true homophones where words have two clear and unrelated meanings, e.g., bark (of a dog, of a tree) and words with multiple related senses, e.g., letter, which has many related meanings linked to a core meaning associated with writing. Words with many senses were responded to faster than otherwise matched words with few senses. A distributed connectionist model of these data (Rodd Gaskell, & Marslen-Wilson, 2001) suggested that such words had wide shallow attractor basins allowing for rapid lexical decisions. However, Rodd et al (2001) found the same effect in auditory lexical decision, whereas De Jong et al (2000) does not find the family size effect in the auditory domain. The degree to which these two effects are similar and whether a model similar to Rodd et al (2001) could model family size effects would be interesting to explore.

Strong claims about the nature of lexical representation have been made on the basis of experiments that manipulate lemma and word form frequency counts. Clahsen (2000) claimed that regularly inflected items should show only lemma effects. However, work has shown that some forms thought of as typifying inflection can show clear effects of word form frequency counts or separate effects of both lemma and word form frequency (Baayen et al, 1997, Bertram et al 2000, Bertram et al 1999). The data presented here further support the idea that the use of categories of derivational and inflectional morphology alone to predict frequency effects in lexical decision is not adequate.

The lack of lemma frequency effects in Experiment 1 does not rule out a role for lemma frequency in such words. Rather, it suggests that the effect may be relatively small for words that show the typical pattern of a high correlation between lemma frequency and word form frequency and a ratio of these frequencies of about 3:1. When there is a large divergence in the frequencies, as in Experiment 2, lemma frequency effects show up significantly in regression analyses. Although affix homonymy has been shown to cause word form frequency effects to dominate, Davis et al (This Volume) show that dominant effects of a single frequency count do not necessarily imply unique effects. Moreover, Baayen, Tweedie & Schreuder (2001), further show that there may be individual differences in which variables are responsible for response times. The contribution of the frequency of morphological variants may also vary both across word type and within word type dependent on the relative frequencies of the morphological family.

Lexical decision response times are not simply the result of the resting activation level of a single representation, but rather a result of a decision process, in which a number of sources of information, form, morphology and semantics are utilised. These sources of information can clearly be subdivided further, with morphological information comprising the various frequency/family size variables that have been shown to affect response times. It would be useful to try to integrate what we know about the role of morphology in lexical decision, with the considerable literature on the modelling of lexical decision, which has typically ignored morphology (Gomez, Ratcliff, & McKoon, 1998, Ashby, 2000). It would also be of value to explore further the time course of access to these sources of information (cf. De Jong et al, 2000, Rastle, Davis, Marslen-Wilson & Tyler, 2000).

In conclusion these results demonstrate the benefits of using correlational and factorial designs in conjunction, to detect both the existence and the strength of frequency and family size effects and to prevent the overly strong interpretation that might result from the use of a single design in isolation.

References

Ashby F.G. (2000). *A stochastic version of General Recognition Theory*. Journal of Mathematical Psychology, 44, 310-329.

Azuma T., Van Orden G.C. (1997) *Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times*. Journal of Memory and Language, 36 (4): 484-504.

Baayen R.H., Dijkstra T., Schreuder R. (1997). *Singulars and plurals in Dutch: Evidence for a parallel dual-route model.* Journal of Memory and Language, 37 (1): 94-117.

Baayen R.H., Piepenbrock, van Rijn, (1993). The CELEX lexical database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Baayen, H.R.., Tweedie, F.J., Schreuder, R.(2001). *The Subjects as a Simple Random Effect Fallacy: Subject Variability and Morphological Family Effects in the Mental Lexicon*. Brain and Language, pp. 1-11 (doi: 10.1006/brln.2001.2506).

Bailey T.M., Hahn U. (2001*). Determinants of wordlikeness: Phonotactics or lexical neighborhoods?* Journal of Memory and Language, 44 (4): 568-591.

Beauvillain, C. (1996*). The integration of morphological and whole word form information during eye fixations on prefixed and suffixed words*. Journal of Memory and Language, 35, 801-820.

Bertram R., Hyönä, J., Laine M. (2000). *The role of context in morphological processing: Evidence from Finnish*. Language and Cognitive Processes, 15, 4/5, 367-388.

Bertram R., Laine M., Baayen R.H., Schreuder R., Hyona J. (2000). *Affixal Homonymy triggers full-form storage, even with inflected words, even in a morphologically rich language*. Cognition, 74 (2): B13-B25

Bertram R.., Schreuder R.., Baayen R.H., (2000). *The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and* *productivity*. Journal of Experimental Psychology: Learning, Memory and Cognition, 26 (2): 489-511.

Bradley, D.C. (1979). *Lexical representation of derivational representation*. In Aronoff M. & Kean, M.L. (Eds.), Juncture (pp37-55), Sarasota, CA: Anma Libri.

British National Corpus (1995), Oxford: Oxford University Computing Services.

Broadbent, D.E. (1967) *Word frequency effect and response bias.* Psychological Review, 74(1): 1-15.

Burani C., Caramazza, A. (1987). *Representation and processing of derived words.* Language-and-Cognitive-Processes.; Vol. 2(3-4): 217-227.

Clahsen, H. (1999). *Lexical entries and rules of language: A multidisciplinary study of German inflection*. Behavioral and Brain Sciences, 22 (6).

Chatterjee, S, Hadi, A.S, Price, B (2000). *Regression Analysis by Example.* (3rd Edition). Wiley: New York.

Cole, Segui J., Taft, M. (1997). *Words and Morphemes as Units for Lexical Access.* Journal of memory and Language, 37, 312-330.

Coltheart, M., Davelaar, E., Jonasson, J.T., Besner, D. (1977). *Access to the internal lexicon*. In Dornic, S. (Ed.), Attention and Performance VI. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cutler A., Norris D. (1979) *Monitoring Sentence Comprehension.* In "Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett" ed. WC. Cooper & E. C. T. Walker (eds) pp.113-134 Erlbaum

De Jong N.H., Schreuder R., Baayen R.H. (2000) *The morphological family size effect and morphology*. Language and Cognitive Processes, 15 (4-5): 329-365.

De Jong, N.H., Feldman, L.B., Schreuder, R., Pastizzo, M., Baayen, R.H. (2001). *The Processing and Representation of Dutch and English Compounds: Peripheral Morphological and Central Orthographic Effects*. Brain and Language, doi:10.1006/brln.2001.2547

Draper, N.R., Smith H. (1981). *Applied regression analysis*. (2nd Edition). Wiley: New York.

Feldman, L.B., (1994). *Beyond orthography and phonology – differences between inflections and derivations*. Journal of Memory and Language, 33 (4): 442-470.

Feldman L.B. (2000) *Are morphological effects distinguishable from the effects of shared meaning and shared form?* Journal of experimental psychology-learning memory and cognition 26 (6): 1431-1444.

Forster, K.I. (1976). *Accessing the mental lexicon*. In Wales, R..J. & Walker, E. (Eds) New approaches to language mechanisms. North Holland Publishing Company.

Gernsbacher M.A. (1984). *Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy*. Journal of Experimental Psychology-General, 113 (2): 256-281

Giraudo H., Grainger J. (2000). *Effects of prime word frequency and cumulative root frequency in masked morphological priming*. Language and Cognitive Processes, 15, 4/5, 421-444.

Gomez, P., Ratcliff, R., McKoon, G. (1998). *Modelling the decision process in the lexical decision task*. Journal of Mathematical Psychology, 42 (4): 489-490.

Howes D. (1954). *On the interpretation of word frequency as a variable of affecting speed of recognition*. Journal of Experimental Psychology, 48, 106-112.

Kaiser, H. F. (1958). *The varimax criterion for analytic rotation in factor analysis*. Pyrometrical, 23, 187-200.

Kytö, M., Romaine, S. (1997). *Competing Forms of Adjective comparison in Modern English: What could be more quicker and easier and more effective?* In Nevalainen, T. & Kahlas-Tarkka, L. (Eds.) To Explain the Present: Studies in the Changing English Language in Honour of Mattie Rissanen. Helsinki.

Laudanna, A., Badecker, W., Caramazza, A. (1992). *Processing inflectional and derivational morphology*. Journal of Memory and Language, 31 (3): 333-348.

Laudanna, A., Burani, C., Cermele, A. (1994). *Prefixes as processing units*. Language and Cognitive Processes, 9 (3): 295-316

Landauer, T.K., Dumais S.T. (1997). *A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge*. Psychological Review, 104 (2): 211-240.

Lieber, R. (1981). *On the organisation of the lexicon*. (Ph.D. dissertation, MIT), Bloomington: Indiana University Linguistics Club.

Marslen-Wilson, W.D., Ford, M., Older, L., Zhou, X. (1996) *The combinatorial lexicon: Priming derivational affixes*. Proceedings of the 18th Annual Conference of the Cognitive Science Society.

Marslen-Wilson, W., Tyler, L.K., Waksler, R., Older, L., (1994). *Morphology and meaning in the English mental lexicon*. Psychological Review, 101: (1) 3-33.

Meunier, F., Segui, J. (1999). *Frequency effects in auditory word recognition*. Journal of memory and Language, 41, 327-344.

Morton, (1969). *Interaction in word recognition*. Psychological Review, 76, 165-178.

MRC Psycholinguistic Database: Machine Usable Dictionary. Version 2.00 Informatics Division Science and Engineering Research Council Rutherford Appleton Laboratory Chilton, Didcot, Oxon, OX11 0QX Michael Wilson 1987.

Parks, R., Ray, J., Bland, S. (1998). Wordsmyth English dictionary-thesaurus [ONLINE]. Available: http://www.wordsmyth.net/ [1991, February 1], University of Chicago.

Perlmutter, D. (1988). *The split morphology hypothesis: evidence from Yiddish*. In Hannan & Noonan (Eds.). Theoretical Morphology: Current Approaches in Linguistics , Academic Press: New York.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Pinker, S. (1991). *Rules of Language*. Science, 153, 530-535

Plank, F. (1994). *Inflection and derivation*. In Asher, R.E. (Ed.). Encyclopedia of Language and Linguistics. Oxford; New York: Pergamon Press.

Plaut, D., Gonnerman, L.M. (2000). *Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing?* Language and Cognitive Processes, 15, 4/5, 445-485.

Rastle, K, Davis M.H., Marslen-Wilson, W.D., Tyler, L.K. (2000). *Morphological and semantic effects in visual word recognition: A time-course study*. Language and Cognitive Processes, 15 (4-5): 507-537.

Rodd, J.M., Gaskell, M.G., Marslen-Wilson, W., (2000). *Why are words with many senses recognised faster?* International Journal of Psychology, 35 (3-4): 27-27.

Rodd, J.M., Gaskell, M.G., and Marslen-Wilson, W.D. (2001). For Better of Worse: Modelling Effects of Semantic Ambiguity. Proceedings of the Twenty Third Annual Conference of the Cognitive Science Society. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Rosenthal, O., Fusi, S., Hochstein, S. (2001). *Forming classes by stimulus frequency: Behavior and theory*. Proceedings of the National Academy of Sciences, 98 (7): 4265-4270.

Rubenstein, H., Garfield, L., Millikan, J.A. (1970). *Homographic entries in the internal lexicon*. Journal of Verbal Learning and Verbal Behavior, 9(5): 487-494.

Rueckl, J.G., Raveh, M. (1999). *The influence of morphological regularities on the dynamics of a connectionist network*. Brain and Language, 68 (1-2): 110-117.

Seidenberg, M.S., MacClelland, J.L. (1989). *A distributed, developmental model of word recognition and naming*. Psychological Review, 96(4): 523-568.

Sereno, J.A., Jongman A. (1997). *Processing of English inflectional morphology*. Memory and Cognition, 25 (4): 425-437.

Schreuder R., Baayen R.H. (1997). *How complex simplex words can be*. Journal of memory and Language, 41, 327-344.

Stump, G.T. (1998). *Inflection*. In Spencer, A and Zwicky A.M. (Eds.) The Handbook of Morphology. Blackwell: Oxford.

Taft, M. (1979). *The recognition of affixed words and the word frequency effect*. Memory and Cognition, 7, 263, 263-272.

Whaley, C.P. (1978). *Word-Nonword Classification Time*. Journal of Verbal Learning and Behavior, 17, 143-154.

Wunderlich, D. & Fabri, R. (1995) *Minimalist morphology: An approach to inflection*. Zeitschrift für Sprachwissenschaft 14:236–94.

Wurm, L.H. (1997). *Auditory processing of prefixed English words is both continuous and decompositional*. Journal of Memory and Language, 37 (3), 438-461.

Table 1. Experiment 1, Summary of Stimuli.

	Mean	Median	Min	Max	Standard Deviation
CMF	153.1	76.3	1.1	2150.3	232.8
FS1	3.4	2.0	0.0	50.0	5.2
FS2	12.0	7.0	0.0	99.0	15.0
FS3	6.5	2.0	0.0	124.0	12.2
FS4	26.2	19.0	1.0	169.0	26.3
LF	118.0	52.5	0.0	1976.0	192.0
WF	86.2	36.0	0.0	1789.0	148.8
Len	4.4	4.0	3.0	6.0	0.9
N	7.3	6.5	0.0	24.0	5.6
BF	567.2	521.0	28.0	1902.0	296.9
TF	45.6	39.0	0.0	202.0	36.7
LSA	0.5	0.5	0.0	1.0	0.2
CNC	539.7	581.0	234.0	670.0	95.8
FAM	550.0	556.0	401.0	645.0	51.6
IMG	550.9	574.0	289.0	667.0	73.8

CMF = CELEX cumulative morpheme frequency, FS1 = Family Size 1, FS2 = Family Size 2, FS3 = Family Size 3, FS4 = Family Size 4, LF = CELEX Lemma Frequency, WF = CELEX Word Form Frequency, Len = number of letters, N = number of neighbours, BF = Bigram Frequency, TF = Trigram Frequency, LSA = mean of LSA Semantic measure, CNC = Concreteness, FAM = Familiarity, IMG = Imageability.

All CELEX Frequencies are per million. Concreteness, familiarity, and imageability ratings from the MRC Psycholinguistic Database.

Table 2. Experiment 2, Harmonic Item Means.

	Comparative		Superlative	
	RT	***E%***	***RT***	***E%***
Adjectives	523	6.5	522	4.6
	RT	***E%***		
LF Matched	496	2.9		
WF Matched	537	7.1		

Appendix 1 – Experiment 1 Stimuli

frill
leash
prank
helm
bigot
kilt
tusk
gem
jot
swig
clot
grub
punk
furrow
dung
glue
canoe
frenzy
bun
matron
hoof
crest
carrot
herald
sandal
spur
den
jug
ballot
ribbon
ivory
core
lid
tomb
era
cliff
raid
poem
beard
client
cafe
camel
marble
charm
lawn
myth
thigh
pub
rival

error
trout
hint
fee
grain
slope
aunt
acre
pride
fuel
mood
cheque
finch
clam
cork
mime
pope
barge
plank
prose
swan
jazz
pork
germ
gut
axe
chalk
cod
gorge
booth
lily
barn
breeze
probe
herb
rug
ladder
linen
canal
index
cage
heap
salad
barrel
fuss
lamb
chapel
elbow
golf

guilt
rod
gap
luck
fun
chest
beer
cell
noise
lip
link
beach
pain
memory
team
scene
method
mile
chance
church
city
food
verb
quest
berry
ink
sword
herd
belly
drum
bean
basket
tube
arch
cotton
belt
cheese
nail
prize
slave
string
cycle
boot
sheep
brick
rope
ease
seed
silver

chain
load
wing
shirt
path
track
meat
spot
shoe
sky
radio
grass
coffee
king
edge
square
seat
hall
music
garden
test
price
art
value
game
wall
level
front
union
force
group
mother
year
cart
east
coat
gold
tooth
body
milk
window
law
horn
centre
knife
worm
powder
farm
coal

salt	word	boat	paper
wire	moon	gas	table
nature	block	bag	road
heat	sand	north	air
dust	cup	tea	bed
clock	glass	hole	class
bath	sex	wood	land
brain	neck	gun	foot
suit	leaf	bird	car
bomb	party	skin	war
court	bell	cross	door
motor	town	board	mind
flower	nose	dog	form
leg	note	stone	book
space	fruit	oil	side
chair	pig	horse	night
sugar	key	shop	water
lamp	storm	blood	room
hat	nut	sun	home
trade	pin	heart	head
person	cake	sign	point
sight	tail	tree	face
agent	west	fish	house
pipe	plate	field	hand
girl	bone	hair	life
star	south	fire	time

Appendix 2 – Experiment 2 Stimuli

Condition 1

Comparative

bigger	greyer	poorer	steeper
blacker	grosser	prouder	stiffer
braver	harder	quicker	stricter
brighter	harsher	rarer	surer
broader	hotter	redder	sweeter
colder	keener	richer	taller
cruder	madder	rounder	thicker
crueller	milder	sadder	tighter
darker	neater	safer	tougher
drunker	newer	sharper	vaguer
fairer	nicer	shorter	wetter
fatter	odder	sicker	whiter
fiercer	older	smaller	wilder
fonder	paler	smarter	wiser
grander	pinker	softer	younger

Superlative

biggest	greyest	poorest	steepest
blackest	grossest	proudest	stiffest
bravest	hardest	quickest	strictest
brightest	harshest	rarest	surest
broadest	hottest	reddest	sweetest
coldest	keenest	richest	tallest
crudest	maddest	roundest	thickest
cruellest	mildest	saddest	tightest
darkest	neatest	safest	toughest
drunkest	newest	sharpest	vaguest
fairest	nicest	shortest	wettest
fattest	oddest	sickest	whitest
fiercest	oldest	smallest	wildest
fondest	palest	smartest	wisest
grandest	pinkest	softest	youngest

Condition 2

woman	battle	carry	engine
market	middle	agree	missile
victim	pocket	practice	plenty
bishop	travel	merchant	spirit
cotton	angel	language	valley
museum	solid	challenge	yellow
splendid	sugar	attack	
reason	problem	crisis	

Condition 3

cherry	garlic	sullen	champion
gurgle	lizard	wallet	salute
paddle	shovel	reptile	magnet
elbow	coleslaw	raisin	tartan
pylon	merry	puddle	carpet
rifle	petal	parrot	buckle
straddle	razor	porridge	
courgette	feeble	scavenge	