

TITLE: Morphological effects in auditory word recognition: Evidence from Danish

RUNNING HEAD: Morphological effects in Danish

Laura Winther Balling
Department of English
University of Aarhus
Jens Chr. Skous Vej 5
8000 Aarhus C
DENMARK
E-mail: englkb@hum.au.dk
Phone: +45 51740757. Fax: +45 89426540

R. Harald Baayen
Department of Linguistics
4-26 Assiniboia Hall
University of Alberta
Edmonton T6G 2E5
CANADA
E-mail: baayen@ualberta.ca
Phone: 001 780 492-5697. Fax: 001 780 492-0806

Corresponding author: Laura Winther Balling

ABSTRACT

In this study, we investigate the processing of morphologically complex words in Danish using auditory lexical decision. We document a second critical point in auditory comprehension in addition to the Uniqueness Point (UP), namely the point at which competing morphological continuation forms of the base cease to be compatible with the input, henceforth the Complex Uniqueness Point (CUP). Suffixed words with later CUP elicited longer response latencies. We also observed an interaction between suffix frequency and whole-word frequency. Both suffix and whole-word frequency were facilitatory, except for words for which both frequencies are high. For such words, we observed inhibition, and most clearly so for female compared to male participants. Finally, a comparison of complex with simple words revealed that, other things being equal, complex words have a processing advantage compared to simple words. We discuss the consequences of these findings for models of morphological processing.

ACKNOWLEDGMENTS

We are grateful to Ocke-Schwen Bohn for helpful advice and discussion, to Lee Wurm for useful input on the UP-effects and to Jonathan Grainger and two anonymous reviewers for constructive comments on an earlier version of this paper.

INTRODUCTION

This study contributes to the understanding of morphological effects in auditory word recognition in three ways: First, we introduce a new uniqueness point measure which accounts for the contribution of morphological continuation forms to lexical competition in suffixed words, complementary to the traditional Uniqueness Point (UP) measure. Secondly, we investigate the role of whole-word and constituent frequencies and document a facilitatory effect of suffix frequency in the absence of a base form frequency effect, together with an interaction of suffix frequency with whole-word frequency. Thirdly, we compare the processing of morphologically complex with simple words and find a solid advantage for the complex words when other properties are held constant. The language under investigation is Danish; to our knowledge, this is the first study that addresses morphological processing in Danish.

Danish is a North Germanic language with a relatively simple inflectional system, similar to English, though verbs are not marked for person-number agreement. Adjectives are inflected for gender and syntactic position as in Dutch. Nouns inflect for definiteness with a distinction between neuter and non-neuter, as well as for number. Danish has productive derivation but productivity is most clearly manifested in compounding. Compounding is so productive in Danish that morphological families are much larger than those in two other Germanic languages, English and Dutch, though smaller than those reported for Finnish, see table 1.

(Table 1 around here)

The UP is the phoneme at which a word deviates from all other words that share the same phonemes up to and including the phoneme preceding the UP. This is the point where lexical

competition is resolved and the word has become uniquely distinguishable from its competitors. Words and nonwords with later UP tend to elicit longer reaction times (e.g. Marslen-Wilson, 1984, 1990). Radeau, Morais, Mousty, & Bertelson (2000) observed for French that the effect of UP was restricted to slower speech rates. Possibly, this is due to faster speech rates being characterized by more acoustic reduction. For laboratory speech, UP effects have been replicated repeatedly, see for instance Wurm (1997) for American English and Wurm, Ernestus, Schreuder, & Baayen (2006) for Dutch.

The definition of the UP is a-morphological: It is based on a cohort of lexical competitors from which all suffixed forms and compounds have been excluded (Marslen-Wilson, 1984). If suffixed words and compounds would be allowed into the cohort, the vast majority of words in languages like English and Danish would have their UP at word offset, rendering the concept of UP vacuous. As a consequence, the UP construct does not capture the competition between morphological continuation forms that takes place after the uniqueness point.

For prefixed words, early studies (Tyler, Marslen-Wilson, Rentoul, & Hanney, 1988; Schriefers, Zwitserlood, & Roelofs, 1991; Wurm, 1997) compared the effects of the UP of the whole word and the UP of the base as an independent word and found effects only of the former. However, the UP of the base as an independent word is arguably an inappropriate measure of lexical competition in prefixed words where the prefix has already been encountered before the base is processed. Therefore, Wurm (1997) defined a context-sensitive base UP: the UP of the root (or base) given the prefix it occurs with. Wurm (1997) and Wurm & Ross (2001) found that words where this conditional root uniqueness point (CRUP) preceded the standard UP were processed significantly faster than those where they coincided.

The CRUP construct adds a refinement to the UP for prefixed words. Wurm et al. (2006) added a further refinement to the UP-research programme by considering the role of the lexical cohort at word offset. They calculated Shannon's entropy across the morphological continuation forms that constitute the cohort at word offset and found that words with higher entropy, indicating more continuation forms or more equiprobable ones, were recognised faster in auditory lexical decision. Since inflectional suffixed words usually have few or no continuation forms, the continuation entropy measure is not relevant for suffixed inflectional forms. However, this does not imply that the UP would capture all the lexical competition for suffixed inflected words. For example, if morphological continuation forms of the base are excluded, an inflected word like *duft-ede* ('smelled', past tense) in Danish becomes unique at the third phoneme, /f/, although a large number of words are still compatible with the input after this phoneme, including both inflectional forms like *duft-et* ('smelled', past participle) and compounds like *duft-e-vand* ('perfume'). What we need is a supplementary UP measure, henceforth the Complex UP (CUP), for the point at which the complex word deviates from its morphological competitors, including continuation forms of the base, but excluding continuation forms of the complex word itself. In the example above, the CUP occurs at the final phoneme of *duftede*, /ə/. Continuation forms of the whole word are excluded to avoid that the CUP would be located at the offset just as for the classic UP. The CUP applies not only to inflected words, but also to derivations and compounds. To give an example from English: for a derived word such as *kind-ness*, the UP occurs at the first /n/ where the base *kind* deviates from other words in the language like *captain* and *kite*. However, although the base can be identified at the UP, the competition is not fully resolved since morphological continuations of *kind-* are still compatible with the input. The CUP measures the point where this morphological competition is resolved; in the case of *kind-ness*, the CUP occurs at the second /n/ where *kind-ness*

deviates from other continuations like *kind-hearted* and *kind-ly*. A first goal of this paper is to provide experimental support for the CUP-construct.

A second goal of this study is to improve our understanding of frequency effects in lexical processing. It is well-known that a higher frequency of occurrence affords shorter processing latencies across a wide range of cognitive domains (see Hasher & Zacks, 1984), and auditory comprehension is no exception (Rubenstein & Pollock, 1963). Although it has been argued that frequency effects in auditory word recognition would arise at the level of auditory access representations (Schreuder & Baayen, 1995; Baayen, McQueen, Dijkstra, & Schreuder, 2003), other researchers argue that there are no such auditory access representations (Marslen-Wilson & Zhou, 1999; Marslen-Wilson, Zhou, & Ford, 1996) and that frequency effects probably reflect central semantic/conceptual probabilities. The latter position finds support also in studies of visual word recognition: Baayen, Feldman, & Schreuder (2006) found that word frequency effects were much stronger in a more semantic task like lexical decision than in naming where form-related variables played a larger role. Furthermore, in a cluster analysis of the predictors used in their analysis, Baayen and colleagues found that frequency clustered with the semantic rather than the form variables. From this perspective, whole-word frequency is a measure of familiarity with the concept or meaning of the word, irrespective of whether the word is complex or simple. For instance, the frequency of English *speed* and the frequency of *hast-ig-hed*, its Danish trimorphemic equivalent, would both reflect familiarity with the same concept, irrespectively of morphological complexity.

Alongside whole-word frequency, base form frequency has been the subject of considerable attention in the literature. While whole-word frequency has been understood as a measure of word-based processing, base form frequency has been seen as a primary measure of morphological

processing (e.g. Taft, 1979; Baayen, Dijkstra, & Schreuder, 1997; Caramazza, Laudanna, & Burani, 1988). However, base form frequency effects have failed to emerge in a number of studies (Bertram, Schreuder, & Baayen, 2000; Sereno & Jongman, 1997; Wurm et al., 2006) in contrast to the morphological family size effect. The morphological family size is the type count of words in which a given target word (or, in the case of complex words, its base) appears as a constituent. Across a range of languages, this type count has emerged as a crucial predictor for response latencies, to the exclusion of the corresponding cumulative token count (e.g. De Jong, Schreuder & Baayen, 2000; Moscoso del Prado Martín, Bertram, Häikiö, Schreuder, & Baayen, 2004). This suggests that base form frequency might not be the best measure of morphological processing. There may be several reasons for this: Firstly, base form frequency measures the unigram probability of the base occurring, without reference to the specific structural position and morphological context in the given complex target. Secondly, there is evidence from auditory word recognition that listeners are sensitive to subphonemic cues that differentiate a base in a complex word from the same base occurring as an independent word (Kemps, Ernestus, Schreuder, & Baayen, 2005; Kemps, Wurm, Ernestus, Schreuder, & Baayen, 2005; see also Davis, Marslen-Wilson, & Gaskell, 2002); again the base form frequency may be too coarse a measure to capture the sensitivity of the processing system, in this case to subphonemic cues. Thirdly, there is the issue of statistical power: Baayen, Wurm, & Aycock (2007) sampled randomly from a large dataset from the English Lexicon Project (Balota et al., 2002) and found that at least 1000 items were necessary to detect base form frequency effects in more than 50% of samples. Together these considerations help explain why base form frequency effects are not systematically detected in standard non-primed word recognition studies. Given the estimates of Baayen et al. (2007), the likelihood of detecting a base form frequency effect for the dataset of the present study, with 160 words, is below 0.25.

Therefore, our primary focus is on the frequency of the complementary constituent, the suffix. Affixal properties have been studied by Laudanna & Burani (1995) and Burani & Thornton (2003) for Italian. We consider affixal token frequency (which, as we shall see, can be exchanged with affixal type frequency). What makes affixal token frequency interesting compared to base frequency measures is the much wider range of values covered by the affixal measure. Hence, it promises to be a more powerful measure to gauge constituent-driven lexical processes. A second goal of the present study is to establish whether suffix frequency is indeed a significant predictor of auditory word recognition. One of the issues that needs to be addressed is whether suffix frequency might interact with other frequency measures, as interactions between constituent and whole-word frequencies were found in the visual modality by Baayen et al. (2007) and by Kuperman, Bertram, & Baayen (2007, 2008). For instance, Baayen et al. found that the effect of base word frequency is facilitatory for words with low whole-word frequency, but that it becomes slightly inhibitory for words with high whole-word frequency. Another issue that should be kept in mind when studying frequency effects is the possibility that frequency effects might be stronger for females than for males (cf. Ullman et al., 2002).

A third goal of this study is to evaluate the processing advantage or disadvantage arising from morphological complexity by comparing simple and complex words. It is conceivable that simple words are easier to recognise than complex words, precisely because they have no internal structure. On this view, internal structure requires time-costly processing. However, if this is true one may wonder why languages would have morphologically complex words at all. It is also conceivable that internal structure reduces memory load, by providing partial motivation both at the level of form and at the level of meaning. A reduction of memory load would render complex words easier

to process. By comparing simple and complex words in a regression analysis, matching statistically for a range of distributional properties, we hope to obtain further insight into this issue.

In the following, we report the results of an auditory lexical decision experiment designed to address these three goals. We made use of multiple regression in order to avoid the problems known to arise when continuous predictors are dichotomized for the purposes of analysis of variance (MacCallum, Zhang, Preacher, & Rucker, 2002; Cohen, 1983).

EXPERIMENT

Method

Materials

90 derivational and 70 inflectional Danish forms were presented in the experiment along with 72 simple words and 232 nonwords. We used suffixed words in order to investigate the role of Complex UP. Seven inflectional and nine derivational suffixes were represented by 10 words each; the suffixes are listed in table 2. All inflectional forms in the experiment were fully regular. The experimental words were all bimorphemic with no allomorphic variation, other than regular schwa-deletion and the usual variations in stress and *stød* (creaky voice on a portion of a voiced phoneme, which is phonemic in some contexts). Homonymy with other suffixes was unavoidable for the group of inflectional suffixes, because of the reduced nature of the Danish inflectional system, but given the preceding stems, the suffixes were always unambiguous. Homonymous and strongly polysemous words were not included. The nonwords were constructed by changing one to three phonemes in each word, while retaining the suffixes, so that these should not in themselves be cues to wordness.

(Table 2 about here)

Danish has no standardised lexical resources like CELEX (Baayen et al. 1995); crucially, the existing corpora are not marked up for morphological structure in derivations and compounds. Therefore, our initial task was to modify existing resources so that the critical distributional variables could be calculated. We downloaded two corpora together making up 43.5 million words (see <http://korpus.dsl.dk/e-resurser/>). The corpora consist of texts from a variety of genres and were collected from 1988 to 1992 and from 1998 to 2002, mainly for lexicographic purposes and in order to track the development of Danish in these periods. Both corpora are lemmatised and tagged for inflection and word type, but mark no other morphological structure. The downloaded text files were parsed into a database run under pgAdminIII which could be queried for string matches and for specific tags using Structured Query Language (SQL). Base form and whole-word frequencies were extracted based on the lemmatisation of the corpus, as were the type and token frequencies of the inflectional suffixes. Because of the absence of derivational and compounding structure in the tagging of the corpora, derivational suffix frequencies, morphological families and continuation forms were extracted using regular expressions and then manually cleaned in order to avoid accidental string matches. The suffix frequency counts included only tokens where the suffix occurred in the outermost level of morphological structure, i.e. appearing at the very end of the word and added as the last morpheme during the composition of the word. This ensures maximum comparability of the counts for inflected words (for which the suffix is most often word-final in Danish) and derived words. The whole-word frequencies are the frequencies of the whole words actually presented and do not include other inflectional variants. The same logic applies in the case of the base form frequency: It is the frequency of the base in its citation form, rather than the whole

lemma. For the base, the frequencies of the citation form and of the whole lemma are highly correlated, and the results presented below hold also if lemma frequency is used.

For the UPs, we phonologically transcribed each word and then queried the corpus for spellings compatible with the transcription in order to identify the phoneme of the UP and the phoneme of the CUP for each word. We then located these phonemes in the speech signal, and defined the middle of the time segment straddled by the phoneme as the uniqueness point. For stops, however, the beginning of the release noise was defined as the uniqueness point.

The derivations varied considerably in semantic transparency, which was therefore assessed for all the complex words in a pre-test, applying the procedure of e.g. Wurm (1997). This also provided us with some control for the difference between inflected and derived words. Participants in the rating study were asked to rate the semantic relatedness between each of the complex words and its base on a scale from 1 to 7, with 1 indicating no relation and 7 indicating a close relation. 36 native speakers of Danish completed the ratings in an online questionnaire, with each base-suffixed pair presented with the base first for half the participants and the suffixed derivation or inflection first for the other half. Each participant received a different random order of the pairs. The mean rating for each word was used as a predictor in the regression analysis.

The variables that we collected for our materials are summarised in table 3. The collinearity between these lexical variables is very high, indicated by a κ of 152.2. For analytical clarity, various lexical variables were decorrelated by fitting regression models where one variable was predicted by those variables with which it was highly correlated, and then using the residuals of these decorrelating regression models as predictors. In this way, semantic transparency was decorrelated

from morphological family size, morphological family size was decorrelated from suffix and whole-word frequency, and suffix frequency was decorrelated from whole-word frequency. Further, following Baayen et al. (2007), length was recalculated as distance in msec from CUP to word offset, considerably reducing the collinearity of the model. The κ of 45.6 of the model still indicates some collinearity between the variables, primarily between UP, CUP and length. We made sure that this collinearity was not harmful by rerunning the model with a less collinear division of the acoustic signal, into three parts: the distance from onset to UP, from UP to CUP, and from CUP to offset. The model with this partitioning showed the same pattern of significant effects and had a low collinearity, indicated by a κ of 12.6. We present the model (see table 5) with the original UP- and CUP-measures for conceptual clarity. The raw correlations between the decorrelated predictors are shown in table 4.

(Table 3 about here)

(Table 4 about here)

De Vaan, Schreuder, & Baayen (2007) and Baayen et al. (2007) found that the reaction time to a given target entered into strong correlations with the reaction times to the items immediately preceding in the experiment. The four preceding reaction times are also highly collinear, and were therefore orthogonalized using principal components analysis.

The stimuli were recorded by a female native speaker of Danish, on a Sony DAT-recorder (model TCD-D8) and using a Sony electret condenser microphone (model EC-959a). The reading lists were digitized at a sampling rate of 22 KHz. The items were segmented from the reading lists and normalised for peak intensity.

Participants

22 volunteers (13 females, nine males) participated in the experiment. The participants were between the ages of 19 and 37; all were native speakers of Danish and reported normal hearing. The participants were tested separately in a sound-attenuating room. None of the participants in the rating study participated in the lexical decision experiment.

Procedure

The experiment was run on a portable computer using DMDX version 3.1.4.5 (Forster & Forster, 2006). The stimuli were presented over headphones and the participants adjusted the volume to a comfortable level. Each stimulus was preceded by a fixation point (a plus sign) displayed for 500 msec in the middle of the screen. A trial ended when participants responded or at time-out which occurred 2500 msec after the beginning of each stimulus. Reaction times were measured from stimulus onset.

The stimuli occurred in a different pseudo-random order for each participant, with no more than three words or nonwords in a row and with no one suffix appearing in two consecutive trials. The orders were generated using Mix (Van Casteren, 2006).

Participants received standard lexical decision instructions and were allowed to ask questions after training on 20 items whose composition was similar to that to the experimental items. Two breaks were inserted one third and two thirds through the 464 experimental items; six warm-up items appeared at the beginning of the experiment and two after each break. The experiment lasted 15 to

20 minutes. Participants pressed a key marked “YES” with their dominant hand, a key marked “NO” with the other hand.

Results and discussion

For the analysis of the experimental data, we made use of linear mixed-effects models. These models are mixed in the sense that they combine fixed-effects factors (factors and covariates) with random-effect factors (factors the levels of which sample only a small subset of the population of factor levels). In repeated measures designs, typical random effect factors are subject and item. Traditionally, these are analysed by means of separate by-subject and by-item analyses. Linear mixed-effects models make it possible to fit one overall model to the data, including both subject and item as independent (crossed) random effect factors. This offers many advantages, including the possibility to bring longitudinal effects in the experiment into the model.

In our analyses, we included not only subject and word as random effects, but also suffix, as we sampled only a subset of Danish suffixes. The simplest way in which a random effect is instantiated in the model is through random intercepts. For subjects, random intercepts calibrate the model for slow versus fast subjects. For items, random intercepts provide adjustment to the population means by allowing for difficult versus easy items. Sometimes, random slopes (or contrasts) are required in addition to random intercepts. Whether the presence of such random slopes is justified in the model is ascertained by means of likelihood ratio tests, which evaluate whether the increase in the number of parameters is justified given the increase in goodness of fit.

The significance of fixed-effects predictors can be evaluated by means of the usual t-test for the coefficients, but there is no good solution for determining the appropriate degrees of freedom. We

therefore rely on the Bayesian method of Markov chain Monte Carlo sampling of the posterior distribution of the parameters, which provides both the functional equivalent of 95% confidence intervals in the form of 95% Highest Posterior Density intervals, and corresponding p-values. The reader is referred to Baayen (2008, chapter 7), and to Baayen, Davidson & Bates (2008) for introductions to mixed-effects modelling. We made use of the statistical computing environment R (version 2.4.1) and the lme4 package (Bates & Sarkar, 2005). The errors were fitted using a logistic link function and binomial variance.

This section focuses on the analyses of the morphologically complex items, which investigate the variables relevant to the complex items, including those pertaining specifically to the morphological structure of the items.

One derived word was removed because of missing semantic transparency ratings, while two inflected words were removed because of unintended whole-word homonymy and atypical pronunciation by the speaker, leaving 157 items. For the reaction time analysis, one very low-frequency and semantically opaque derived item was removed because of an error rate over 30%. Additionally, all error responses were removed for the RT-analysis, all in all, 3.7% of the responses. We fitted an initial regression model, using backwards stepwise elimination of non-significant predictors. For this initial regression model, we inspected the residuals, removed potentially overly influential outliers and refitted the model. As a criterion for removal of outliers, we used a cut-off point of 2.5 standard deviations, which amounts to 2.8% of the correct responses being excluded. We then refitted the model to this trimmed dataset. The coefficients and corresponding statistics reported below pertain to this refitted model. With this approach, fewer datapoints are removed than

with filtering based on means and standard deviation cut-offs, and a better model fit is obtained. For the error analysis, all responses to the 157 items were retained.

(Table 5 about here).

Table 5 summarises the estimated coefficients for the regression model in the first column and the associated p-values based on the t-distribution with as degrees of freedom the upper bound (the number of observations minus the number of fixed effects parameters). Table 5 also summarises the posterior distribution of the fixed effect coefficients, using Markov Chain Monte Carlo sampling (MCMC) with 10,000 samples (see Baayen, Davidson, & Bates, 2008). The second column of table 5 shows that the means of these posterior distributions are close to the model estimates listed in the first column. The third and fourth columns list the lower and upper bound of the 95% highest posterior density (HPD) intervals which provide superior accuracy to the standard confidence intervals based on the t-distribution. The p-values listed in the fifth column provide the corresponding MCMC significance levels.

To complete the specification of the model, we report the estimates for the random effect parameters: For the RT-model, the standard deviation for the by subject random intercepts was estimated at 0.0639, the standard deviation for the by word random intercepts at 0.0431 and the standard deviation for the by suffix random intercepts at 0.0279. We also checked whether random slopes might be required. This turned out to be the case only for word frequency, the effect of which varied significantly between subjects. The standard deviation for the by-subject random slopes for word frequency was estimated at 0.0048. No other random slopes reached significance. The residual standard error of the model was estimated at 0.1303. Inspection of the HPD intervals showed that

these parameter estimates were properly bounded. The inclusion of these parameters in the model was supported by likelihood ratio tests (all $p < 0.05$). In what follows, we discuss the various fixed-effects predictors in the model group by group.

Table 6 sums up the logistic regression model used to analyse the error data, with estimated coefficients in the second column, and their associated standard errors, z- and p-values in the subsequent columns. The standard deviation for the by subject random effect was 0.3494 and for the by item random effect 0.9887; a suffix random effect was not included because the variance accounted for was so small that it did not justify the extra parameter.

(Table 6 around here)

Context-related control variables

The RT-analysis showed strong effects of a number of control predictors that all relate to the experimental context. These are shown in fig.1, using the same y-axis scale as in figures 2 to 4 and 6 in order to make the magnitude of the effects directly comparable. The top two panels of fig. 1 show the effect of reaction times to the previous four items, orthogonalized using principal components analysis, as described above. PC1, the principal component that accounts for most of the variance in the four preceding reaction times, was significant in both the RT and error analyses, while PC2 had a significant effect only on RTs. PC1 was negatively correlated with the RTs at the preceding trials, hence the negative slope in the upper right panel of fig.1 indicates that longer RTs at the preceding trials predict a longer RT for the current trial. Inspection of the factor loadings for PC2 indicates that PC2 likewise captures the inhibitory effect of elongated preceding RTs, with the RT to the immediately preceding trial most highly correlated with PC2.

(Fig. 1 about here)

The effect of PC1 was modulated by the number of times a suffix had been repeated, indicated by the interaction between Previous PC1 and Suffix Repetition (see table 5): The effect of the previous RTs was stronger at the first occurrence of a suffix than at the last. This is shown in the top left panel of fig. 1 where the dotted line shows the effect of PC1 for the first occurrence of a suffix and the dashed line the effect for the last occurrence. The bottom left panel of fig. 1 illustrates the effect of suffix repetition, i.e. an effect of suffix priming, which varies with previous RT. The plot shows that this priming effect is noticeable mainly for cases where the PC1 was low, corresponding to fast reactions to the previous items. This indicates that suffix repetition primarily plays a role when responses are fast.

Finally, the correctness of the previous response was significant, with an error response to the immediately preceding item resulting in slower RTs, as shown in the bottom right panel of fig. 1.

These four variables are more informative about the nature of the task than about the structure of the mental lexicon, but the effects, particularly of the previous RTs, were strong and highly significant and as such constitute an important source of noise in the experiment that should be brought under statistical control.

Uniqueness points

We found significant effects both of the traditional UP and of the new Complex UP, the CUP: The expected inhibitory effect of the UP is shown in the top left panel of figure 2. The effect of the

CUP, displayed in the top right panel of fig. 2, is also inhibitory, as expected, but non-linear, with the inhibition becoming stronger for later CUPs. The fact that both uniqueness points are significant shows that there are two key points during the winnowing of competitors from the cohort. The UP pinpoints in time the elimination of early competition between morphologically unrelated lemmas in the mental lexicon. The CUP represents the later point in time at which remaining morphologically related base continuations are eliminated from the cohort. There was no additional effect of the number of post-offset continuation forms of the whole word which Wurm et al. (2006) found to be significant. This could be because the suffixed inflected forms, which make up almost half the items, have few or no continuation forms.

(Fig. 2 about here)

The nonlinear nature of the CUP effect may indicate that lexical candidates that have been in the cohort together with the target become increasingly strong competitors as time proceeds and more and more candidates have already dropped out of the cohort. Moreover, a late CUP generally indicates not only late ambiguity but also more morphologically related competitors, which could explain why the effect becomes stronger as the CUP becomes later.

Interestingly, words with later CUPs elicited more correct responses than did words with early CUPs. Apparently, the elongated competition characteristic of words with longer CUPs is a reliable index of the input being a real word, and hence allows more correct responses.

Another form-related measure which is known to have significant effects in auditory processing is length, calculated here as msec from the CUP to the end of the word. As expected, this length variable was inhibitory, as can be seen in the bottom panel of fig. 2.

The UP and CUP overlap in that they both include the distance from word onset to UP. In order to make sure that the effects of these predictors are not confounded, we also ran the analysis with the acoustic signal partitioned into three non-overlapping regions: word onset to UP, UP to CUP, and CUP to word offset. All three durational measures turned out to be significant predictors, and all were inhibitory, as expected.

Morphological family size and semantic transparency

The effect of morphological family size is plotted in the left panel of fig. 3; it is a linear facilitatory effect with larger families corresponding to shorter RTs. The presence of a family size effect indicates that the morphological structure of the word is playing a role.

The right panel of the same figure shows the effect of semantic transparency ratings: the more transparent the combination of base and suffix, the faster the recognition. The significance of the semantic transparency of the base-suffix combination is parallel to the results of Wurm (1997). It is also in line with the results of Marslen-Wilson, Tyler, Waksler, & Older (1994) who dichotomized the semantic transparency variable and observed priming effects for transparent but not for opaque words. Since we did not dichotomize the variable (for the statistical reasons given above), the results are not directly comparable; however, we did not observe interactions between semantic transparency and either constituent frequency, which would have corresponded to the difference in priming effects observed by Marslen-Wilson and colleagues. The difference may be the result of the

present opaque items being less opaque than those employed by Marslen-Wilson et al., or of the possibility that in unprimed processing opaque words can undergo decomposition.

(Fig. 3 somewhere here)

Frequency effects

The frequency of the base form did not reach significance in our model. We also did not observe any interactions involving base form frequency. This holds both when base frequency was residualised on whole-word frequency (with which it was correlated) and vice versa. Although we anticipated this non-result on the basis of the power analyses presented in Baayen et al. (2007), it remains striking that the family size measure is apparently so robust that it can emerge as a significant diagnostic of morphological processing where base form frequency fails to do so. We refer to Baayen et al. (2007) for a probabilistic explanation of why base form frequency fails to be predictive across many lexical decision tasks. We also investigated whether the absence of base form frequency effects was due to affix homonymy, as Bertram, Schreuder, & Baayen's (2000) reported that base form frequency effects were not observed for complex words whose affix was homonymous with another affix. However, in the present study, base form frequency did not differ between items with homonymous vs. non-homonymous affixes, nor were there a main effect or any other interactions involving affix homonymy. The fact that Bertram et al. studied visual word recognition could be responsible for this difference to the present auditory experiment: Bertram et al.'s items were of a length that can be read in one fixation, so that the homonymous suffixes are processed at the same time as the base which selects specifically for one of the suffixes. In contrast, the base always precedes the suffix in auditory processing, making interference from a

homonymous suffix less likely when it is selected for by a base of a different word class than that which has actually occurred.

We did observe significant effects for the frequency of the whole word, the frequency of the suffix, as well as a significant interaction between these two predictors. Interestingly, the pattern of frequency effects also varied as a function of the sex of the participant.

Fig. 4 illustrates this interaction between suffix frequency, whole-word frequency, and sex. The top panels show the effect of suffix frequency for words with minimum (dotted), median (solid) and maximum (dashed) whole-word frequency, for females (left) and males (right). For words with median whole-word frequency, the effect of suffix frequency is practically absent. However, there is a clearly facilitatory effect of suffix frequency for words with low whole-word frequency, and a clearly inhibitory effect for words with high whole-word frequency. The effects are strongest for females.

The bottom panels of fig. 4 demonstrate for females (left) and males (right) how the whole-word frequency effect is modulated by suffix frequency: The solid line represents the whole-word frequency effect for words with median suffix frequency, a straightforwardly facilitatory effect. For words whose suffixes have the lowest token frequency, represented by the dotted line, the facilitatory effect of whole-word frequency is noticeably stronger than for the median frequency suffixes. In contrast, for words with maximum suffix frequency, shown by the dashed line, the effect of whole-word frequency reverses to become slightly inhibitory for the women (the left panel) and flat for the men (the right panel).

The difference between the sexes is best discerned by comparing the left panels of fig. 4 with the right panels: The effects and their interactions are more pronounced for females than for males. This is in line with the results of Ullman et al. (2002), who found stronger word frequency effects for females under various conditions. However, in contrast to Ullman et al., we found that the word frequency effects were significant for the males, although they were weaker than for the females.

When studying a group difference, such as the sex difference in the present analysis, it is important to make sure that it does not arise because individual differences were ignored. Following the general logic of analysis of variance, what we need to show is that the group difference is solid when all relevant individual differences (here with respect to the slopes of word frequency and suffix frequency) have been brought into the model. We therefore checked whether by-subject random slopes for word frequency and suffix frequency were required. As mentioned above, by-subject random slopes were explanatory. By-subject random slopes for suffix frequency were not. In the model with by-subject random slopes, the three-way interaction of word frequency, affix frequency, and sex remained significant. From this we conclude that the group difference is real and not an artefact of individual differences.

The effect of sex in the present data is small, a general characteristic of male/female differences (see, e.g., Kimura, 2000). Of course, given that our subjects are all highly educated, highly literate speakers, generalizations to larger populations (Danish males or females, or males and females in general) are not warranted. Further investigations showed that the effect is not confounded with age or handedness. We can therefore conclude that our data do provide support for the hypothesis of Ullman et al. (2002) that females have an advantage with respect to declarative memory, at least for our subject population of educated, literate speakers. In the light of the fact that, in our experience,

most subjects that volunteer for psycholinguistic experiments are female, it is worth keeping in mind that results for (unspecified, hence predominantly female) subject groups may be biased towards whole-word processing. As sex differences are not the main topic of our results, and function merely as a subject-specific control, we now return to the discussion of the interaction of word frequency by suffix frequency.

The multiplicative interaction underlying fig. 4 imposes a highly restrictive geometric form on the joint effect of suffix and word frequency on the response latencies. In order to verify that this functional form is indeed justified, we also inspected the data with the help of conditioning plots. The conditioning plot for reaction time as a function of word frequency for different intervals of suffix frequency is shown as fig. 5. Suffix frequency increases from left to right and from bottom to top, as indicated by the highlighted areas of the strips above the panels. As we move through these panels, we see that the facilitatory effect of frequency becomes successively smaller.

Overall, when one of the two frequency measures assumes a low value, the other frequency measure is facilitatory. In contrast, when both measures assume a high value, the frequency effect disappears (word frequency for men) or even becomes inhibitory (word frequency for women and suffix frequency for both sexes). As facilitation is expected for frequency effects, this pattern leaves us with the question why the effects disappear or become inhibitory when both frequencies are high. The interaction observed here for auditory lexical decision is very similar to the interaction observed for visual lexical decision by Baayen et al. (2007) between base form and whole-word frequency: these authors found that the facilitatory effect of whole-word frequency was greatest for words with the lowest base form frequency, and that a facilitatory effect of base form frequency found for words with the lowest whole-word frequencies turned into slight inhibition for words with

the highest whole-word frequencies. The interaction is also similar to the interactions observed by Kuperman, Bertram, & Baayen (2008) in an eye-movement study of Finnish compound reading; in that study, compound frequency interacted with both left and right constituent family size.

We take these interactions to indicate that the parsing and whole-word routes in a dual-route framework do not operate independently. Whereas dual-route race models such as Frauenfelder & Schreuder (1992), Schreuder & Baayen (1995) and Bertram, Laine, & Karvinen (1999) predict maximal efficiency for words with high constituent and high whole-word frequency due to statistical facilitation, we observe delays in processing for such words. Independent evidence for delayed processing when the parsing route and the whole-word route are effective simultaneously is reported by Baayen & Plag (2007): Average response latencies were low for unproductive affixes for which the role of the parsing route is negligible, as well as for high-productivity affixes where the whole-word route is attenuated. Affixes occupying the middle ground where both routes are active simultaneously show the longest response latencies. The question is then why we observe delayed processing when both suffix and whole-word frequencies are high.

For words that have a high word frequency and high constituent frequencies, the two processing routes might be roughly equally fast in delivering an interpretation for the complex word. This might give rise to ambiguity, the resolution of which leads to a processing delay. This ambiguity could be structural in nature, i.e. [*hast*] + [*ig*] + [*hed*] vs. [*hastighed*] (speed). Alternatively, the ambiguity could be semantic: Monomorphemic words benefit from having multiple senses (Rodd, Gaskell, & Marslen-Wilson, 2002), but this may not hold for complex words, since complex words often lexicalize for only one sense of their base. For instance, the base *brygge* (brew) also has a figurative sense (concoct), but the derivation *brygg-eri* (brewery) is generally used in relation only

to the literal sense. Such lexicalizations of complex words are presumably stored and accessed as part of the whole-word representation, while the parsing route is likely to activate the meanings of the parts and the transparent composition(s) of those. When both whole-word and morpheme representations are strong (as indexed by high frequencies) and therefore activated simultaneously, the system needs to resolve any conflicts between the activated senses, leading to a processing delay. The activated senses that come into conflict may be more or less distinct, stemming from differences that range from relative opacity to collocational differences. Finally, it is possible the processing delay arises because information from two different neural substrates (see Ullman, 2001, 2004) has to be integrated at roughly the same point in time.

A second question is why there is an asymmetry such that the suffix frequency effect becomes more inhibitory when word frequency is high, than the word frequency does when suffix frequency is high. This is likely to be because the word has more bottom-up support than the suffix and is encountered earlier than the suffix in auditory processing, making the whole-word route more resistant to interference from the suffix, with the result that the whole-word frequency is not reversed but merely attenuated. In contrast, the suffix frequency becomes directly inhibitory because this has less bottom-up support than the word and hence is more prone to interference. This difference in amount of bottom-up support may also help explain why the reversal of the base form frequency effect in the study of Baayen et al. (2007) is weaker than the reversal of the suffix frequency effect in this study but stronger than the attenuation of the whole-word frequency effect in both studies: the bottom-up support that the base form receives is intermediate between that of the suffix and the whole-word.

In our analysis we have used suffix token frequency as a predictor. Suffix token frequency is highly correlated with suffix type frequency in our data, and the token measure can be replaced by the type measure without loss of significance. However, the effects seem slightly more robust for the token measure. We therefore used suffix token frequency as the clearest index of suffix strength. A number of potentially confounding factors were tested but did not reach significance; these include the P and P* measures of productivity (Baayen 1992, 1994), affix homonymy, and the difference between inflection and derivation. Since these factors do not reach significance and since semantic transparency is already in the model, we believe that the interaction of suffix frequency by whole-word frequency cannot be reduced to confounds with other variables. Moreover, similar interactions are found in other languages using other paradigms (visual lexical decision in English in Baayen et al., 2007; eye-movement measurements for Finnish compounds in Kuperman, Bertram, & Baayen 2008).

Comparing simple and complex words

In the above analyses we examined only the complex words, but the experiment also included simple words. In order to address the third main goal of this study, the assessment of the relative advantage or disadvantage of morphological complexity, we report a final analysis where we include the 72 simple words that were also presented in the experiment, using as predictors only those variables that make sense for both simple and complex items and excluding variables that are specific to the complex words, such as suffix frequency. For the UPs, we used the three-way division of the acoustic signal into distance from onset to UP, from UP to CUP, and from CUP to offset. Both UP and CUP exclude continuations of the whole word, and for simple words there are no base continuations which are not also whole-word continuations. Therefore, the distance from UP to CUP is zero for the simple words, but the three-way division allows us to control for CUP for

the complex words. The fixed effects in the model that we fitted to the data are summarised in table 7. The model also included random intercepts for item (s.d. 0.0536) and participant (s.d. 0.0641); the residual standard error was estimated at 0.1311.

(Table 7 around here)

With the other variables in the model functioning as controls, the contrast coefficients for Affix Type in table 7 show that inflected and derived words are both processed faster than simple words (the reference level). Other things (such as length, UP, word category) being equal, complex words have a processing advantage over simple words, as visualised in fig. 6. Fig. 6 shows the partial effect of the difference between the three types, given that all other predictors are held constant at their medians. In other words, we are considering the imaginary situation in which simple and complex words are equally frequent, equally long, have the same UPs, etc. Since actual simple words tend to be shorter and more frequent, they tend to be processed faster than complex words, but this advantage is due to various distributional properties that are not intrinsic to the distinction between simple and complex words. After these extrinsic covariates have been partialled out (as in fig. 6 and in the regression model summarised in table 7), we see that, other things being equal, complex words have a processing advantage.

(fig. 6 around here)

This processing advantage is especially interesting in the light of the finding that when both whole-word frequency and suffix frequency are very high, processing is delayed. Apparently, such extreme cases are rare. We conclude that the majority of complex words benefit from the

simultaneous availability of two processing routes, memory and parsing. This benefit is likely to be both formally and semantically motivated. Importantly, the effect of repeated exposure to the suffixes was controlled by the inclusion of a variable indexing how many times the suffix had been encountered in the experiment previous to each trial. This rules out the possibility that the complexity advantage observed is an artefact of affix priming.

In the overall analysis of simple and complex items, the effect of sex did not reach significance. This is not surprising, as the interaction of sex by word frequency is expected primarily for complex words, where according to the dual mechanism theory of Pinker (1997) storage is highly redundant. Hence it is here that differences between males and females are most likely to be observed. By adding a large set of words for which no interaction with sex is expected, the simple words, a small effect is swamped and no longer reaches significance.

Whereas there was a clear difference between simple and complex words in the analysis of all items, the analysis of the complex items provided no support for a processing difference between inflected and derived words, neither in the form of main effects nor in the form of interactions, contrary to what the dual mechanism model of Pinker (1997), with its fundamental distinction between regular inflectional morphology and less regular derivational morphology, would lead one to expect. We found significant effects of variables such as affix frequency and semantic transparency that vary in a graded way between prototypical inflection and prototypical derivation, but no categorical difference between inflected and derived words. When the categorical difference between inflected and derived words was added to the model summarised in table 5, the graded effects remained significant, while the categorical contrast was solidly non-significant ($p > .4$). In other words, it seems that the differences between regular and less regular morphology is best

accounted for as continuous rather than categorical, and thus not easily fitted into a framework with two distinct mechanisms. However, such a categorical difference might be found when a much larger set of affixes is considered across a much larger group of words, see Baayen et al. (2007) for an analysis of many thousands of words in which the effect of affix type survives that of affix frequency.

GENERAL DISCUSSION

This study explored three interrelated questions concerning the lexical processing of complex words in auditory comprehension. First, is there a uniqueness point after the uniqueness point? More specifically, given the evidence for lexical entries for complex words as witnessed by robust and ubiquitous whole-word frequency effects, one would expect to find evidence for a second uniqueness point – after the standard uniqueness point – at which morphologically related competitors have dropped from the cohort. Our auditory lexical decision experiment with Danish complex words provided evidence for the usefulness of such a second Complex Uniqueness Point (CUP). Words with a late CUP elicited longer response latencies.

The CUP effect shows that the incoming speech input is matched against a lexicon that contains entries for complex words, both inflected and derived. Unsurprisingly, a whole-word frequency effect also supported the hypothesis of a comprehensive lexicon. Does this imply that the internal structure of complex words is irrelevant? The absence of a base form frequency effect suggests an affirmative answer. However, the presence of a family size effect and the presence of an effect of suffix frequency suggest that the internal structure of complex words also plays a role. Of special interest is the interaction of whole-word frequency and affix frequency that emerged from our data.

When one of these two variables has a low value, the effect of the other predictor is facilitatory. But when one of the variables assumes higher values, this facilitation for the other predictor reverses and may even become inhibitory. Apparently, when both retrieval of the full form from memory and constituent-driven comprehension are running simultaneously, processing is delayed instead of enhanced. The frequency effects and their interaction were significantly enhanced for females as compared to males. Thanks to their superior verbal memory (Kimura, 2000; Ullman et al., 2002), females are probably better at retrieving both full forms as well as their constituents from the lexicon. Hence they suffer most when the whole word and the constituents are highly frequent. Methodologically, this result indicates that it is important to take the sex of the participants into account when attempting to generalise from experimental work.

Our third question addressed the overall advantage or disadvantage of morphological complexity. High frequency words with high frequency affixes have a processing disadvantage compared to words for which one of these frequencies is not high. How does the average balance of storage and computation fall out for complex words? To answer this question, we compared simple and complex words. Other things being equal (i.e. statistically controlled in the regression model), complex words emerged with a processing advantage compared to simple words. Although at the extreme high end of the frequency ranges, the whole-word and decompositional routes function suboptimally, on average the joint availability of word and morpheme representations allows for more optimal processing.

Our findings have several consequences for models of morphological processing. Frauenfelder & Schreuder (1992) hypothesized that in a race model with two independent routes, complex words might have a processing advantage compared to simple words thanks to statistical facilitation:

When the normally fast direct route happens to be slow, the normally slow parsing route might happen to be fast, and complete access first. Although we do see, on average, a processing advantage for complex words, the interaction of suffix frequency by whole-word frequency suggests the two routes are not independent and that optimal conditions for each route separately do not guarantee optimal processing for the system as a whole. Since the idea of statistical facilitation is based on the assumption that the two routes race independently, while our data suggest that the two routes are not independent, statistical facilitation may not be the correct interpretation for the processing advantage of complex words.

The CUP measure provides further support for the importance of properly conditioning of processing measures. The CRUP construct developed by Wurm (1997) provides further refinement of the UP by conditioning the cohort for the base on the preceding prefix. The present CUP construct conditions the uniqueness point on the presence of the base. The CRUP narrows down the traditional cohort by considering only the bases that can occur in combination with the given prefix, while the CUP draws in more words than those considered by traditional cohort theory. Both in the work of Wurm and colleagues and in this study, the pattern of uniqueness effects shows how sensitive auditory processing is to the context of the different morphemes: for prefixed words, the base is processed given its prefix; for suffixed words, it seems that the same is happening for the suffix given its base. Interestingly, this sensitivity emerges although the phoneme-based UPs are relatively crude measures, given the amount of subphonemic detail that listeners are sensitive to (see e.g. Goldinger, 1998; Kemps, Wurm, Ernestus, Schreuder, & Baayen, 2005; Davis et al., 2002). A further possibility in relation to the CUP measure is that what matters is not only the existence of morphological competitors up until the CUP, but also their frequency relative to the target, similar to the pattern of morphological family effects found by Meunier & Segui (1999),

who observed an inhibitory effect of a larger number of family members with a higher frequency than the target. However, the number of family members of a higher frequency than the target, which was inhibitory in the French experiment of Meunier & Segui, was not predictive for the present Danish experiment. This could be because the larger morphological families in Danish make the frequency rank in the family less distinctive than it is in French. This suggests that the relative frequency of the target in its pre-CUP cohort would not be predictive for languages with large families, like Danish, but could play a role for languages with smaller families. Alternatively, the fact that Danish is stress-timed, while French is syllable-timed, could mean that there are durational cues in Danish which help the listener identify precisely which of the family members is the target, thus reducing the possibility of inhibition from higher-frequency family members. The absence of such durational cues in French could make other cues, such as the relative frequency in the family, more important. Further research is required to clarify this.

The CUP effect and the whole-word frequency effect bear witness to the importance of knowledge of the lexical entries for complex words. These two effects challenge models that reject pervasive storage of whole complex forms and that rely instead on rule-based comprehension (e.g., Pinker, 1997; Pinker & Ullman, 2002). The presence of whole-word frequency and CUP effects for regular inflection side by side with semi-regular derivation adds to this challenge. At the same time, our data also argue against models in which the role of morphological structure is marginalized. Except for extreme values of whole-word and suffix frequency, we observe synergy between the two routes, a synergy without which we suspect it would not be profitable for languages to have morphology at all.

REFERENCES

- Baayen, R.H. (1992). Quantitative aspects of morphological productivity. In G. Booij & J. van Marle, *Yearbook of Morphology 1991*, pp. 109-149. Dordrecht: Kluwer.
- Baayen, R.H. (1994). Productivity in Language Production. *Language and Cognitive Processes*, 9, pp. 447-469.
- Baayen, R.H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge, UK: Cambridge University Press.
- Baayen, R.H., Piepenbrock, R., & Gullikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Linguistic Data Consortium. Philadelphia: University of Pennsylvania
- 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, 94–117.
- Baayen, R.H., McQueen, J.M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R.H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 463–484). Berlin: Mouton de Gruyter.
- Baayen, R.H., Feldman, L.B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55, 290–313.

Baayen, R.H., Wurm, L.H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, 2, pp. 419-463.

Baayen, R. H., Davidson, D., & Bates, D.H. (2008). Mixed-effects modelling with crossed random effects for subjects and items. Revised manuscript under review for *Journal of Memory and Language*.

Baayen, R.H. & Plag, I. (2007). Parsing is not weaknessless: suffix ordering revisited. Manuscript under review for *Language*.

Balota, D.A., Cortese, M.J., Hutchison, K.A., Neely, J.H., Nelson, D., Simpson, G.B., & Treiman, R. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. <http://lexicon.wustl.edu/>.

Bates, D. M. & Sarkar, D. (2005). *The lme4 library*. <http://lib.stat.cmu.edu/R/CRAN/>

Bertram, R., Laine, M., & Karvinen, K. (1999). The Interplay of Word Formation Type, Affixal Homonymy, and Productivity in Lexical Processing: Evidence from a Morphologically Rich Language. *Journal of Psycholinguistic Research*, 28, 213–226.

Bertram, R., Schreuder, R., and 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, 489–511.

Burani, C. & Thornton, A.M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. In R.H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 157–207). Berlin: Mouton de Gruyter.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28, 297–332.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, 249–254.

Davis, M.H., Marslen-Wilson, W.D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218–244.

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

De Vaan, L., Schreuder, R., & Baayen, R.H. (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon*, 2, 1–24.

Forster, K. I. & Forster, J.C. (2006). *DMDX version 3.1.4.5*.

<http://www.u.arizona.edu/~jforster/dmdx.htm>.

Frauenfelder, U.H. & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: the role of productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1991*, (pp. 165–183). Dordrecht: Kluwer.

Goldinger, S.D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Hasher, L. & Zacks, R. T. (1984). Automatic processing of fundamental information. The case of frequency of occurrence. *American Psychologist*, *39*, 1372–1388.

Kemps, R.J.J.K., Ernestus, M., Schreuder, R., & Baayen, R.H. (2005a). Prosodic cues for morphological complexity: the case of Dutch plural nouns. *Memory & Cognition*, *33*, 430–446.

Kemps, R.J.J.K., Wurm, L.H., Ernestus, M., Schreuder, R., & Baayen, R.H. (2005b). Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes*, *20*, 43–73.

Kimura, D. (2000). *Sex and Cognition*. Cambridge, MA: The MIT Press.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R.H. (2008). Reading of multimorphemic Dutch compounds: towards a multiple route model of lexical processing. Manuscript under revision, *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Kuperman, V., Bertram, R., & Baayen, R.H. (2008). Morphological Dynamics in Compound Processing. Manuscript accepted for publication, *Language and Cognitive Processes* (this volume).

Laudanna, A. & Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 345-364). Hillsdale, NJ: Lawrence Erlbaum Associates.

MacCallum, R.C., Zhang, S., Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19–40.

Marslen-Wilson, W.D. (1984). Function and process in spoken word recognition. In H. Bouma & D. Bouwhuis (Eds.), *Attention and Performance X: Control of language processes*, (pp.125–150). Hillsdale, NJ: Lawrence Erlbaum Associates.

Marslen-Wilson, W.D. (1990). Activation, competition and frequency in lexical access. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, (pp. 148–172). Cambridge, MA: MIT Press.

Marslen-Wilson, W.D., Tyler, L.K., Waksler, R., & Older, L. (1994). Morphology and Meaning in the English Mental Lexicon. *Psychological Review*, 191, 3–33.

Marslen-Wilson, W.D., Zhou, X., & Ford, M. (1996). Morphology, modality, and lexical architecture. In G. Booij, & J. van Marle (Eds.), *Yearbook of Morphology 1996*, (pp. 117–134). Dordrecht: Kluwer.

Marslen-Wilson, W. & Zhou, X. (1999). Abstractness, Allomorphy, and Lexical Architecture. *Language and Cognitive Processes*, 14, 321–352.

Meunier, F. & Segui, J. (1999). Frequency Effects in Auditory Word Recognition: The Case of Suffixed Words. *Journal of Memory and Language*, 41, pp. 327-344.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R.H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1271–1278.

Pinker, S. (1997). Words and rules in the human brain. *Nature* 387, 547-548.

Pinker, S. & Ullman, M.T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6, 456–463.

Radeau, M., Morais, J., Mousty, P., & Bertelson, P. (2000). The Effect of Speaking Rate on the Role of the Uniqueness Point in Spoken Word Recognition.

Rodd, J., Gaskell, G., & Marslen-Wilson, W.D. (2002). Making Sense of Semantic Ambiguity: Semantic Competition in Lexical Access. *Journal of Memory and Language*, 46, pp. 245-266.

Rubenstein, H. & Pollack, I. (1963). Word predictability and intelligibility. *Journal of Verbal Learning and Verbal Behavior*, 2, 147–158.

Schreuder, R. & Baayen, R.H. (1995). Modeling morphological processing. In L.B. Feldman (Ed.), *Morphological Aspects of Language Processing*, (pp. 131–154). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schriefers, H., Zwitserlood, P., & Roelofs, A. (1991). The Identification of Morphologically Complex Spoken Words: Continuous Processing or Decomposition. *Journal of Memory and Language*, 30, 26–47.

Sereno, J. & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*, 25, 425–437.

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

Tyler, L.K., Marslen-Wilson, W.D., Rentoul, J., & Hanney, P. (1988). Continuous and discontinuous access in spoken word-recognition: The role of derivational prefixes. *Journal of Memory and Language*, 27, 368–381.

Ullman, M.T. (2001). The Declarative/Procedural Model of Lexicon and Grammar. *Journal of Psycholinguistic Research*, 30, 37–66.

Ullman, M.T., Estabrooke, I.V., Steinhauer, K., Brovotto, C., Pancheva, R., Ozawa, K., Mordecai, K., & Maki, P. (2002). Sex differences in the neurocognition of language. *Brain and Language*, 83, 141–143.

Ullman, M.T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231–270.

Van Casteren, M. (2006). *Mix*. <http://www.mrc-cbu.cam.ac.uk/maarten/Mix.htm>.

Wurm, L.H. (1997). Auditory Processing of Prefixed English Words Is Both Continuous and Decompositional. *Journal of Memory and Language*, 37, 438–461.

Wurm, L.H. & Ross, S.E. (2001). Conditional Root Uniqueness Points: Psychological Validity and Perceptual Consequences. *Journal of Memory and Language*, 45, 39–57.

Wurm, L.H., Ernestus, M., Schreuder, R., & Baayen, R.H. (2006). Dynamics of the auditory comprehension of prefixed words. *The Mental Lexicon*, 1, 125–146.

Table 1

Comparison of largest morphological family sizes attested in three Germanic languages and Finnish. Note that for English spaced compounds are not counted and that for Finnish, the largest count cited is the largest family of an item used in the experiment of Moscoso del Prado Martín, Bertram, Häikiö, Schreuder & Baayen (2004).

Language	Max	Source
English	187	CELEX (Baayen, Piepenbrock, & Gullikers 1995), 18 million words
Dutch	550	CELEX (Baayen et al. 1995), 42 million words (Schreuder & Baayen 1997)
Danish	3476	Danish corpus described below, 43.5 million words
Finnish	6029	Turun Sanomat newspaper corpus, 22.7 million (Moscoso del Prado Martín et al. 2004)

Table 2

Properties of the suffixes used on the complex items in the experiment. Word class covers adjectives (A), nouns (N) and verbs (V). Homonymy refers to homonymy with other affixes.

	Suffix	Word class	Meaning/function (X = root)	Type freq./mill	Token freq./mill	Homonymy
Derivational	-bar	A	Able to be X'ed	3	199	N
	-ere	V	To make X	29	5,975	N*
		N	Place for/state of/case of X	13	458	N
	-eri					
	-hed	N	Quality of being X	51	4,445	N
		A	Like X/characterised by property X	34	5,115	N
	-isk					
		A	Characterised by property X	5	1,380	N
	-iv					
		A	Characterised by property X	14	15,260	N
	-lig					
	-ning	N	Result of X	45	5,661	N
	A	Characterised by property of X	1	577	N	
-som						
Inflectional	-ede	V	Past	153	6,888	Y
		N	Common gender definite	2,337	29,327	Y
	-en					
	-ende	V	Present participle	141	2,834	N
	-er	N	Plural	1,915	26,908	Y
	-est	A	Superlative	8	1,141	N
	-et	V	Past participle	225	15,081	Y
	-s	V	Passive	149	5,834	Y**

* Homographic but not homonymic with adjectival comparative suffix.

** Homonymic with genitive -s which is clitical rather than affixal.

Table 3

Lexical predictors for the complex items. The variables marked with an asterisk are frequency counts per million.

	Median (s.d.)	Range
Whole-word frequency*	2.9 (45.9)	0 to 388.2
Base form frequency*	15.8 (83.8)	0.1 to 691
Suffix token frequency*	5397.9 (8753.1)	199 to 29327
Morphological family size	184 (451.1)	1 to 2165
Continuation forms	2 (40.8)	0 to 458
Semantic transparency rating	5.26 (1.59)	1.25 to 6.97
Length in msec	590 (91)	364 to 820
UP in msec	288 (77)	154 to 523
Complex UP in msec	458 (95)	203 to 682

Table 4

Correlations between the significant predictors for the complex items, with previous RTs included as principal components (PCs), semantic transparency residualised from morphological family size, family size from word and suffix frequency, and suffix frequency from word frequency as described in the text. All pairwise (absolute) correlations are smaller than 0.51, and hence stay well below the traditional most simple warning of collinearity.

	Previous PC1	Previous PC2	Affix rep.	Log UP	Log CUP	Log CUP to offset	Sem. trans- parency	Family size	Log word freq.	Log suffix freq.
Previous PC1	1.00	0.03	0.10	0.02	0.03	-0.03	0.02	-0.01	0.01	0.01
Previous PC2	0.03	1.00	0.01	0.01	0.02	-0.01	-0.02	0.03	0.03	0.01
Affix repetition	0.10	0.01	1.00	0.04	0.05	-0.01	-0.01	0.01	0.01	0.01
Log UP	0.02	0.01	0.04	1.00	0.51	-0.15	0.16	-0.02	0.09	0.15
Log CUP	0.03	0.02	0.05	0.51	1.00	-0.40	0.34	-0.02	0.11	0.06
Log CUP to offset	-0.03	-0.01	-0.01	-0.15	-0.40	1.00	-0.42	-0.01	-0.23	-0.30
Sem. transparency	0.02	-0.02	-0.01	0.16	0.34	-0.42	1.00	-0.07	0.25	0.35
Family size	-0.01	0.03	0.01	-0.02	-0.02	-0.01	-0.07	1.00	-0.01	-0.01
Log word freq.	0.01	0.03	0.01	0.09	0.11	-0.23	0.25	-0.01	1.00	0.44
Log suffix freq.	0.01	0.01	0.01	0.15	0.06	-0.30	0.35	-0.01	0.44	1.00

Table 5

Summary of the mixed-effects analysis of covariance for variables predicting auditory lexical decision time to complex items only. Df = 3214.

	Estimate	MCMC mean	HPD lower	HPD Upper	p MCMC	p (t)
Intercept	13.1803	13.0321	7.4414	18.4922	0.0002	0.0000
Previous RT PC1	-0.0002	-0.0002	-0.0002	-0.0001	0.0001	0.0000
Affix repetition	-0.0037	-0.0037	-0.0053	-0.0019	0.0001	0.0000
Prev. PC1*Affix repetition	0.0000	0.0000	0.0000	0.0000	0.0008	0.0003
Previous RT PC2	0.0000	0.0000	0.0000	0.0001	0.0008	0.0008
Previous response: error	0.0343	0.0329	0.0113	0.0568	0.0060	0.0038
Log UP	0.0970	0.0985	0.0603	0.1388	0.0001	0.0000
Log CUP (linear)	-2.5289	-2.4809	-4.3458	-0.7103	0.0082	0.0101
Log CUP (quadratic)	0.2269	0.2229	0.0746	0.3737	0.0036	0.0049
Log msec from CUP to offset	0.0154	0.0152	0.0086	0.0215	0.0001	0.0000
Residualised semantic transparency	-0.0086	-0.0084	-0.0163	-0.0007	0.0352	0.0474
Residualised family size	-0.0085	-0.0084	-0.0146	-0.0024	0.0072	0.0093
Log word frequency	-0.0127	-0.0128	-0.0396	0.0145	0.3320	0.0001

SexM	0.0057	0.0056	-0.0403	0.0516	0.8102	0.8530
SexM*Log word frequency	0.0015	0.0015	-0.0389	0.0455	0.9494	0.6366
Residualised suffix frequency	-0.0429	-0.0430	-0.0671	-0.0199	0.0004	0.0008
SexM*Resid. suffix frequency	0.0255	0.0251	0.0036	0.0468	0.0236	0.0190
Resid. suffix freq*Log word freq.	0.0095	0.0096	0.0047	0.0140	0.0002	0.0001
SexM*suffix freq.*word freq.	-0.0063	-0.0063	-0.0107	-0.0016	0.0066	0.0056

Table 6

Summary of binomial regression model for correctness of the responses to the complex items, with correct responses scored as 0 and incorrect as 1. Df = 3454.

	Estimate	Std. Error	z	p (z)
Intercept	6.7261	4.2124	1.597	0.1103
Previous RT PC1	0.0008	0.0004	2.029	0.0425
Log Complex UP	-1.6260	0.6962	-2.336	0.0195
Log whole-word frequency	-0.1880	0.0751	-2.503	0.0123

Table 7

Summary of the mixed-effects analysis of covariance for variables predicting auditory lexical decision time for both complex and simple items. Df = 4710.

	Estimate	MCMC mean	HPD lower	HPD Upper	P MCMC	P(t)
Intercept	9.8999	9.9113	6.9224	12.8035	0.0001	0.0000
Previous RT PC1	-0.0001	-0.0001	-0.0001	-0.0001	0.0001	0.0000
Affix repetition	-0.0042	-0.0042	-0.0059	-0.0026	0.0001	0.0000
Prev. PC1*Affix repetition	0.0000	0.0000	0.0000	0.0000	0.0188	0.0220
Previous RT PC2	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001
Previous response: error	0.0288	0.0287	0.0087	0.0476	0.0020	0.0030
Log UP (linear)	-1.3669	-1.3709	-2.4213	-0.3594	0.0104	0.0090
Log UP (quadratic)	0.1413	0.1416	0.0515	0.2331	0.0028	0.0021
Log UP to CUP	0.0729	0.0730	0.0531	0.0928	0.0001	0.0000
Log msec CUP to offset (linear)	-0.0160	-0.0161	-0.0372	0.0036	0.1204	0.1259
Log msec CUP to offset (quadratic)	0.0051	0.0051	0.0012	0.0092	0.0116	0.0122
Log whole-word frequency	-0.0149	-0.0149	-0.0193	-0.0101	0.0001	0.0000
Affix type: Derivation	-0.2609	-0.2610	-0.3600	-0.1579	0.0001	0.0000
Affix type: Inflection	-0.2781	-0.2782	-0.3831	-0.1768	0.0001	0.0000
Word class: Noun	0.0200	0.0198	0.0016	0.0399	0.0594	0.0556
Word class: Verb	0.0287	0.0285	0.0061	0.0509	0.0116	0.0128

Figure 1

Partial effects of the context-related control predictors on auditory lexical decision times, adjusted to the median values of the other significant continuous predictors and the reference levels CORRECT for the factor PREVIOUS RESPONSE and FEMALE for the factor SEX. Previous PC1 and Previous PC2 are the two strongest of four principal components based on the four previous reaction times.

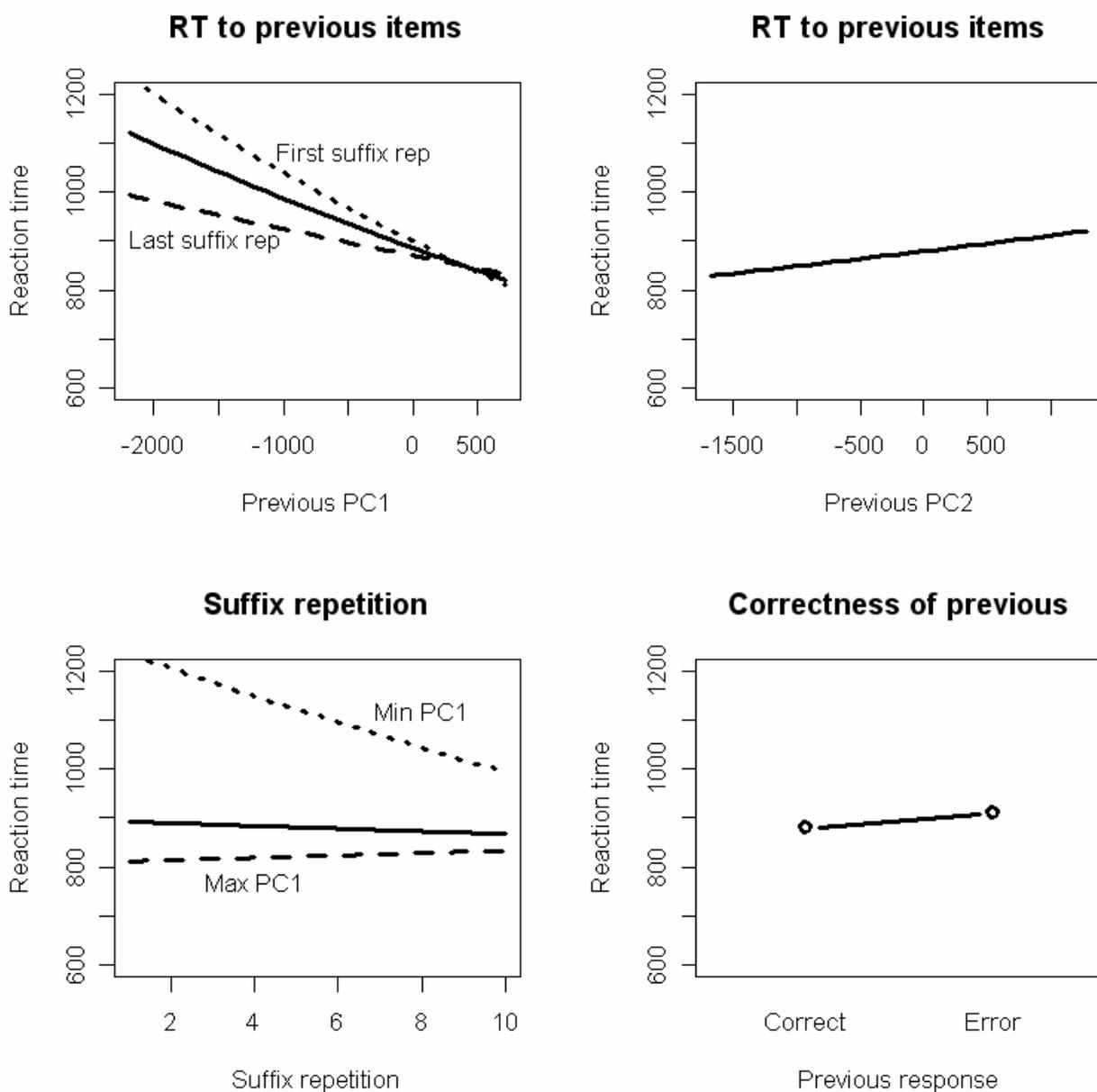


Figure 2

Partial effects of the form-related predictors on auditory lexical decision times, adjusted to the median values of the other significant continuous predictors and the reference levels CORRECT for the factor PREVIOUS RESPONSE and FEMALE for the factor SEX.

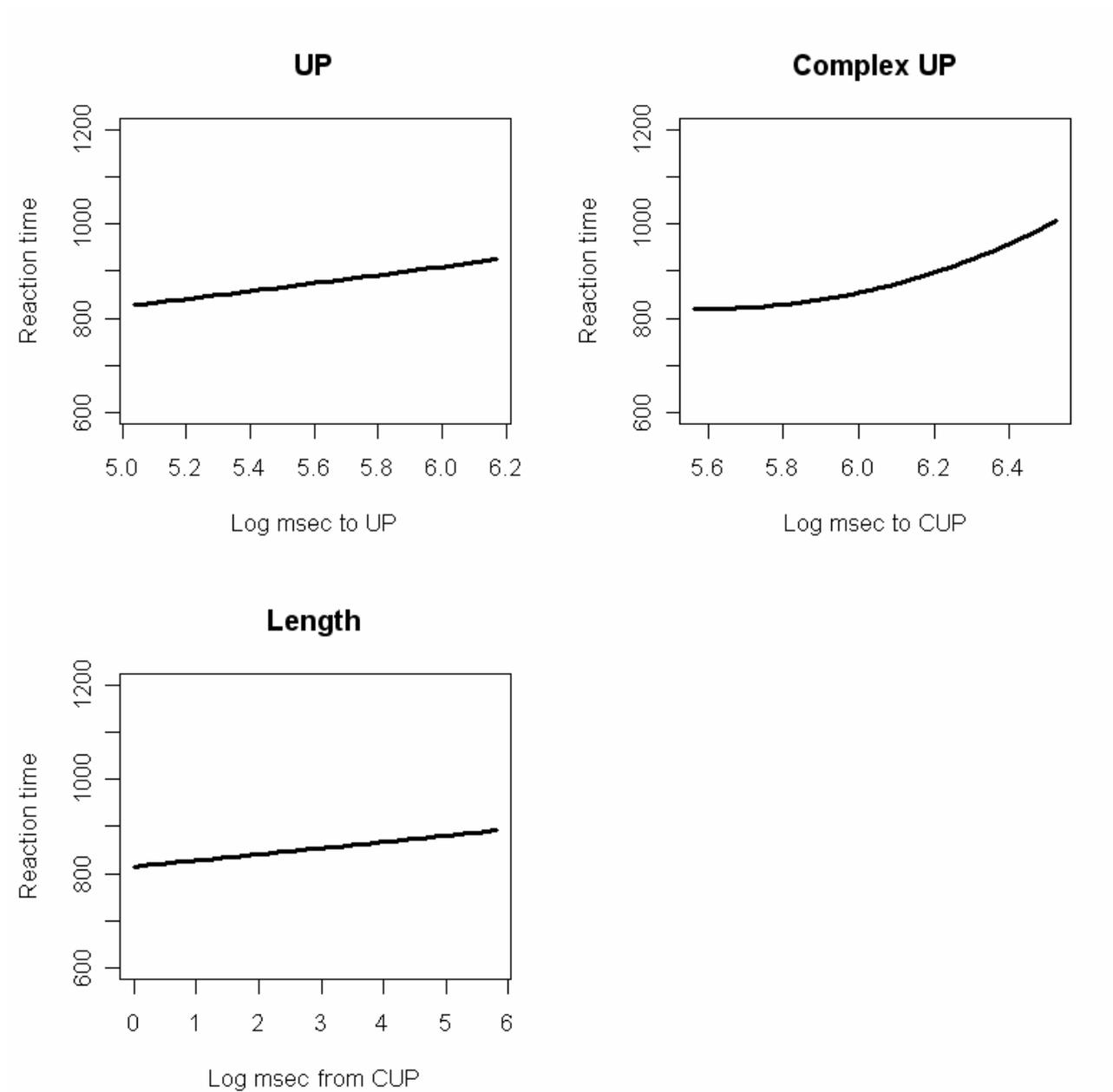


Figure 3

Partial effects of the semantic predictors on auditory lexical decision times, adjusted to the median values of the other significant continuous predictors and the reference levels CORRECT for the factor PREVIOUS RESPONSE and FEMALE for the factor SEX.

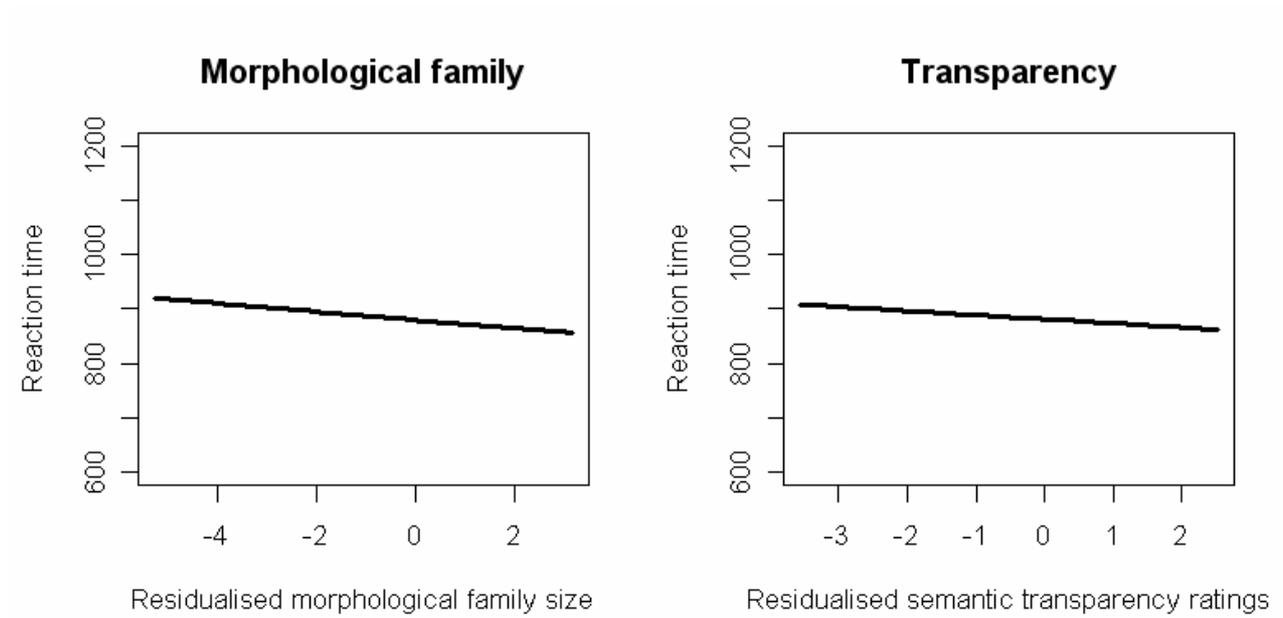


Figure 4

Effects of word and suffix frequency split between the two sexes. The left panels show the effects for female participants, the right panels, the effects for male participants. The top panels show the partial effect of whole-word frequency on RT, split between words with minimum (dotted), median (solid) and maximum (dashed) suffix frequency, and adjusted to the median values of the other significant continuous predictors and the reference level CORRECT for the factor PREVIOUS RESPONSE. The bottom panels show the partial effect of suffix frequency on lexical decision times, split between words with minimum (dotted), median (solid) and maximum (dashed) word frequency, similarly adjusted for the other variables in the model.

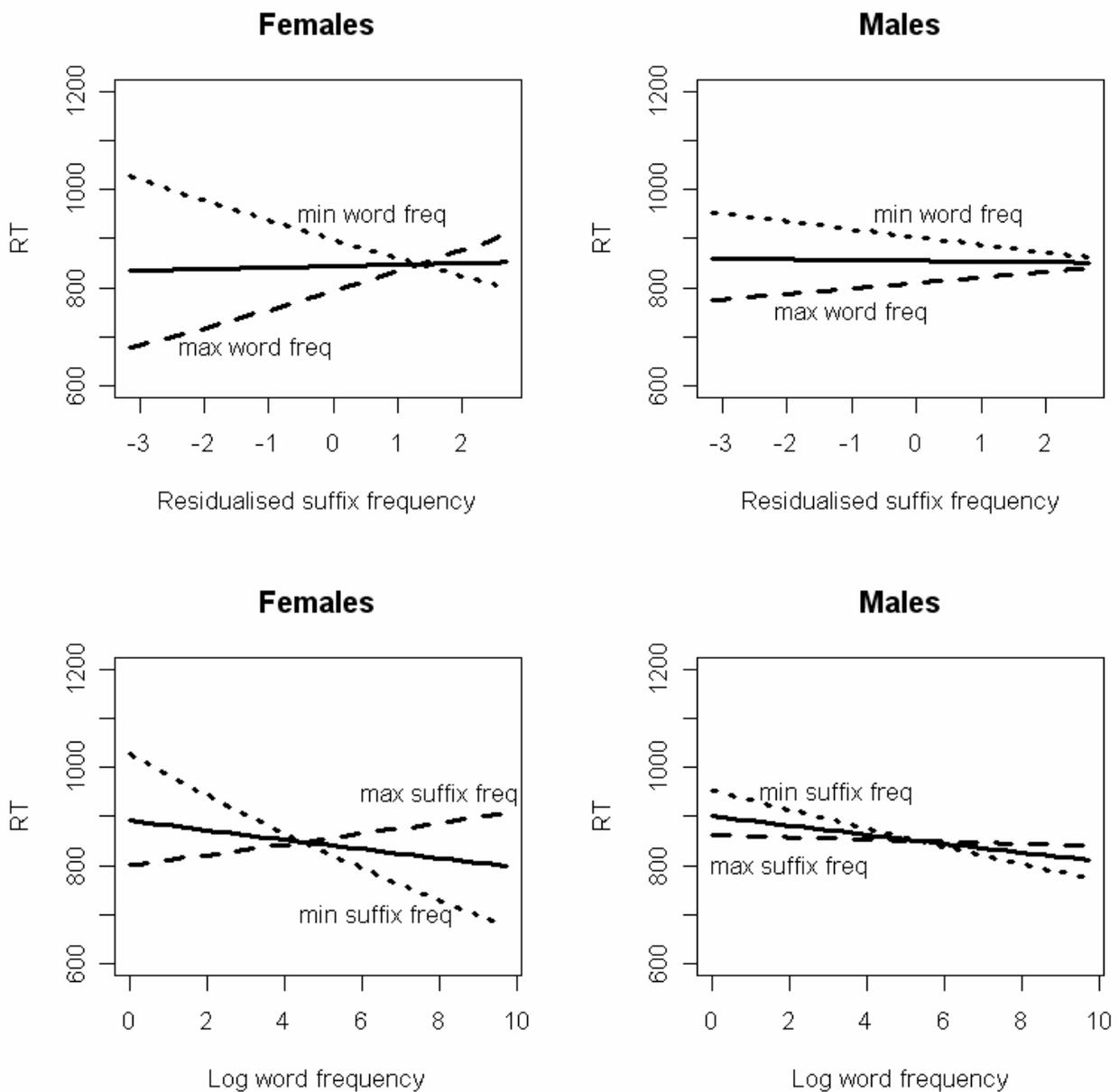


Figure 5

The effect of whole-word frequency for different intervals of suffix frequency, with suffix frequency increasing from bottom left to top right. The intervals are determined by dividing the items into six groups of the same size according to residualised suffix frequency and plotting the effect of whole-word frequency for each of these intervals.

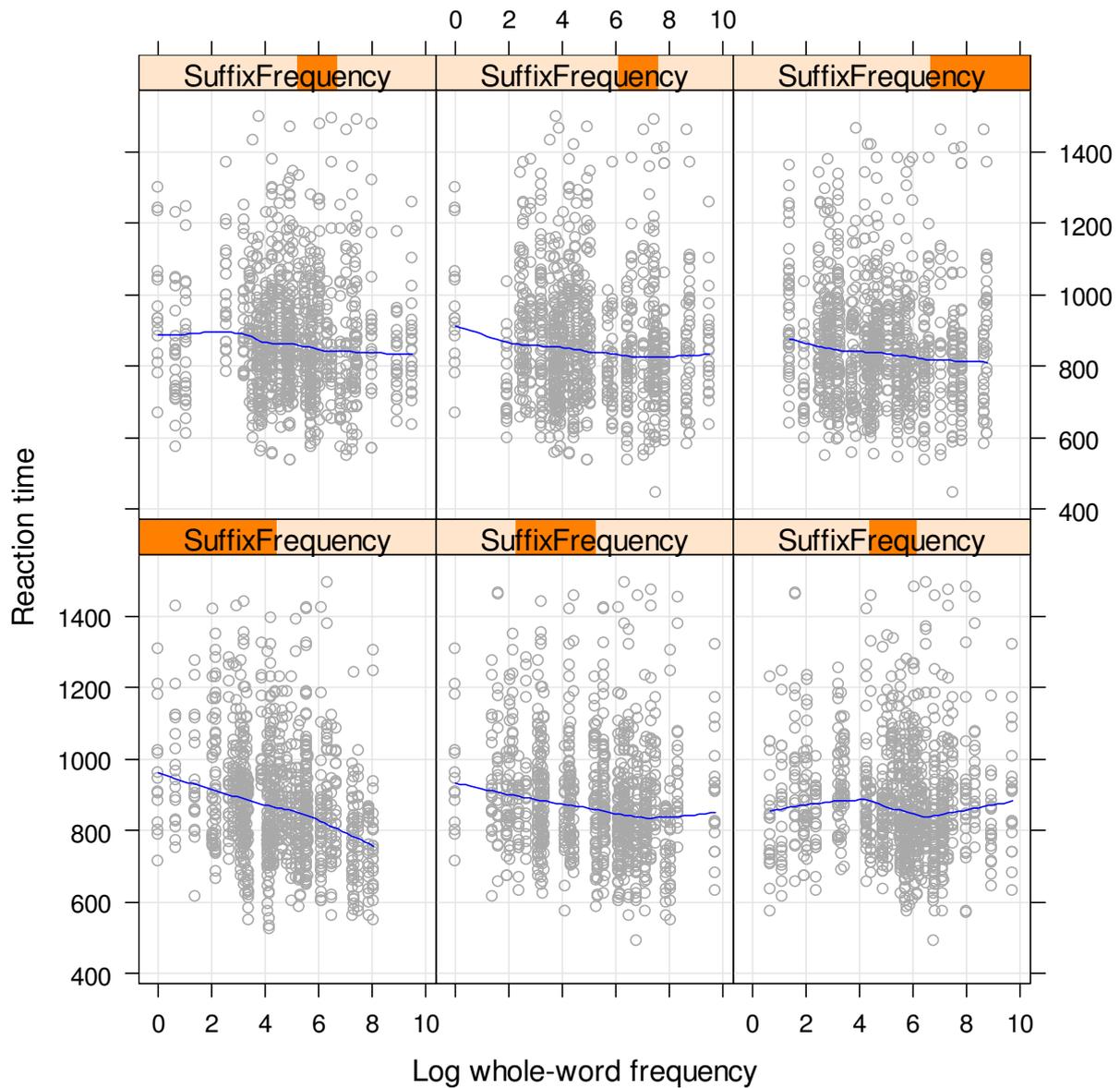


Figure 6

Partial effects of morphological complexity on auditory lexical decision times, adjusted to the median values of the other significant continuous predictors and the reference levels CORRECT for the factor PREVIOUS RESPONSE and ADJECTIVE for the factor WORD CLASS.

