John Benjamins Publishing Company

This is a contribution from *The Mental Lexicon 1:1*

Dynamics of the auditory comprehension of prefixed words

Cohort entropies and Conditional Root Uniqueness Points

Lee H. Wurm[1], Mirjam Ernestus[2], Robert Schreuder[3],
and R. Harald Baayen[2,3]
[1]Wayne State University, Detroit, USA / [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands / [3]Radboud University Nijmegen

This auditory lexical decision study shows that cohort entropies, conditional root uniqueness points, and morphological family size all contribute to the dynamics of the auditory comprehension of prefixed words. Three entropy measures calculated for different positions in the stem of Dutch prefixed words revealed facilitation for higher entropies, except at the point of disambiguation, where we observed inhibition. Morphological family size was also facilitatory, but only for prefixed words in which the conditional root uniqueness point coincided with the conventional uniqueness point. For words with early conditional disambiguation, in contrast, only the morphologically related words that were onset-aligned with the target word facilitated lexical decision.

1. Introduction

This study addresses the auditory comprehension of prefixed words. Following Marslen-Wilson (1987, 1989; Marslen-Wilson & Welsh, 1978) and Norris (1994), we take as our general point of departure that as the speech signal unfolds over time, a group of lexical candidates consistent with the acoustic input processed so far (which we will refer to as the "cohort") is reduced until a word has become identifiable.

A problem for both cohort theory (Marslen-Wilson, 1987, 1989) and the Shortlist model of Norris (1994) is that morphological complexity is not taken

The Mental Lexicon 1:1 (2006), **125–146**.
ISSN 1871–1340 / E-ISSN 1871–1375 © John Benjamins Publishing Company

into account in a principled way. In cohort theory, a word's uniqueness point (UP) is defined as the point at which that spoken word diverges from all other words in the language. Morphological complexity does not figure into the computation of the UP, and furthermore, common practice has been to exclude suffixed continuation forms of a given word when calculating its UP. The Shortlist model makes use of a lexicon that contains complex words as well as simplex words, but no affixes. In both models, morphologically complex words are handled in the same way as monomorphemic words, namely, as strings of phonemes without internal structure.

These approaches are unsatisfactory from a morphological point of view, for various reasons. First, frequency effects have been observed for complex words just as for simplex words (Baayen, McQueen, Dijkstra, & Schreuder, 2003; Wurm, Baayen, & Aycock, 2005). This suggests that it is unsatisfactory to stipulate that they are not taken into account as candidates in the recognition process. Related to this, Kemps, Wurm, Ernestus, Schreuder, and Baayen (2005) documented that morphologically related words consistent with the input up to word offset, henceforth the "morphological continuation forms," co-determine response latencies in auditory lexical decision. They applied Shannon's entropy to the probability distribution (i.e., the frequency-weighted cohort) at word offset, and used it as a measure for detecting the relevance of these continuation forms (for other recent applications of entropy in psycholinguistics, see Kostič, Markovič, & Baucal, 2003; Moscoso del Prado Martín, Kostič, & Baayen, 2004). Kemps et al. (2005) observed a negative correlation between this word-final cohort entropy and response latencies in auditory lexical decision.

A first aim of the present study is to replicate this effect of word-final cohort entropy, and to investigate the explanatory potential of two related cohort entropy measures calculated at earlier points in prefixed words. This will provide a means for gauging with greater precision the cohort reduction over time. Early in a spoken word there can be substantial ambiguity about that word's identity, because a large cohort of words might all be consistent with this minimal acoustic input. By very late in the word there should be very little ambiguity, because at some point, the only words left in the cohort will be the target word and its continuation forms. In order to measure changes in this difficulty of disambiguation as the acoustic signal unfolds over time, we calculated the entropy over the frequency-weighted cohort (which we will call the "probability distribution") of the lexical competitor sets at three positions in the word stem: At the second phoneme, just before the UP, and at word offset.

A second reason that approaches like the cohort model and Shortlist are unsatisfactory is that they fail to account for the role of morphologically related words that are not necessarily part of the cohort. For example, the word *building* is morphologically related to the word *rebuild*, but they are not in each other's cohorts because their initial phonemes do not match. The morphological family size measure allows us to capture these relationships. Family size is usually defined as the number of compounds and derived words that a given simplex word (i.e., stem) appears in as a constituent. Research has shown that family size has a facilitative effect on visual and auditory lexical decision times of monomorphemic words (Baayen, Tweedie, & Schreuder, 2002; Schreuder & Baayen, 1997) and visual lexical decision times of complex words (Bertram, Baayen, & Schreuder, 2000; de Jong, Schreuder, & Baayen, 2000), but to date there has not been a demonstration of a family size effect in the auditory processing literature for complex words. A second aim of the current study is to elucidate the role of family size in the auditory comprehension of prefixed words.

A third aim of this study is to extend the work on conditional root uniqueness points (CRUPs), to which we turn next.

2. The CRUP Construct

Wurm (1997) proposed that morphologically complex words were simultaneously analyzed by two separate routes: A continuous route that processes whole words, and a decompositional (morphemic) route that analyzes words into constituent morphemes when possible. What differentiated this proposal from the many other dual-route proposals in the literature was a unique constraint imposed on the morphemic route. Specifically, after a potential prefix is stripped off, the morphemic route attempts to match the remaining portion of the acoustic signal not to the entire lexicon, but only to a small subset of it. In particular, only free roots (e.g., the *build* in *rebuild*) are considered, not bound roots (e.g., the *-ceive* in *receive*). Furthermore, the only free roots considered are those that have in the past combined with this particular prefix. The perceptual system in this conceptualization would keep track of which prefix-plus-root combinations have occurred before. These constraints were captured in the formulation of a new kind of UP, which Wurm (1997) called the conditional root UP, or CRUP. He defined it as the UP of a root morpheme, given the prefix that has been processed.

The CRUP of a prefixed word can either be the same phoneme as its full-form UP or it can precede the full-form UP. For example, the UP of the spoken word *discredit* is the second [d]; listeners must hear this much to ensure that the word being uttered is not *discrepant, discretion, discriminate,* or a word related to these. The CRUP of *discredit,* though, is the [r], because it turns out that there are no other words with free roots that begin with [dIs'kr] besides *discredit*.[1] Following Wurm (1997; Wurm & Aycock, 2003; Wurm & Ross, 2001) we will refer to words in which the CRUP and the UP differ as CRUP words; words in which the two UPs coincide will be considered Control words (see Figure 1).

The CRUP formulation is based on the hypothesis that the perceptual system is sensitive to the different combinatorial properties of bound and free stems. Wurm (2000), replicating Taft, Hambly, and Kinoshita (1986), found that spoken pseudowords that carry genuine English bound roots (e.g., *po-*

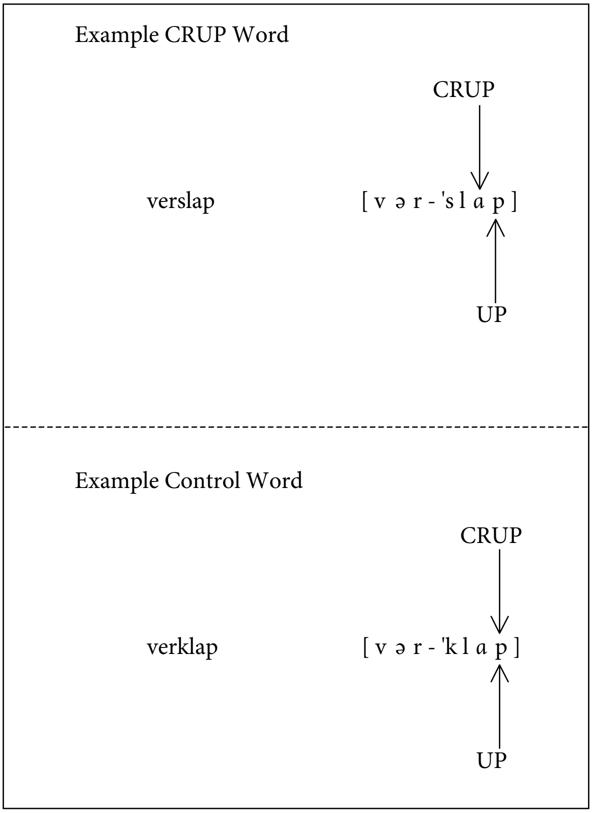

Figure 1. An example CRUP word and an example Control word. CRUP means "Conditional Root Uniqueness Point" and UP means "Uniqueness Point." CRUP words are those in which the CRUP precedes the UP.

ceive) took longer to reject in auditory lexical decision than those that do not (e.g., *po-deive*); and that those carrying both a genuine prefix and a genuine but inappropriate root (e.g., *co-ceive*) are harder still. Wurm (2000) then extended this work by showing that the effects are more pronounced when free roots are used. Wurm (2000) also found that the inhibitory effect of prefix likelihood[2] slows lexical decision times for all complex pseudowords, but significantly more for pseudowords with free as opposed to bound roots. Finally, this study showed that any free root slowed down response times by a roughly equivalent amount, but that the slowing caused by bound roots depended on root frequency. Importantly, the amount of interference caused by even high-frequency bound roots never approached the amount caused by free roots. The crucial conclusion is that bound roots have dramatically less semantic connectivity and ability to be used in new combinations, compared to free roots. Hence, compared to matched controls, CRUP words are more probable in their cohorts, given that the words with bound stems in these cohorts are less probable candidates for recognition.

Evidence from English suggests that CRUP words are recognized faster than Control words. Wurm (1997) found that CRUP words were recognized an average of 47 msec faster than Control words in a gating experiment and 46 msec faster in a lexical decision experiment.[3] Wurm and Ross (2001) contrasted twenty CRUP words and twenty morphological control words. These control words had coinciding CRUPs and UPs and were chosen to be matched to the CRUP words on as many other variables as possible. Wurm and Ross, too, found a large performance advantage for CRUP words (46 msec in a lexical decision experiment and 67 msec in a naming experiment).[4] Wurm and Aycock (2003) used tighter experimental control of some relevant variables, and replicated the RT advantage for CRUP words (henceforth the "CRUP advantage") against two other kinds of control stimuli in both a lexical decision and a naming experiment.

In summary, the CRUP distinction is a valuable predictor for English. Results indicate, on the one hand, that processing in the cohort proceeds conditional on the identification of prior constituents, such as prefixes, and on the other hand, that the distinction between free and bound roots is important for the weighting of probabilities of the lexical candidates. An important purpose of the present study is to ascertain whether the CRUP advantage is confounded with more general differences between CRUP and Control words in the size and probability structure of the cohort.

It is currently unknown whether the CRUP construct has any relation to auditory processing in a language other than English. The present study examines the usefulness of the CRUP construct for Dutch. Dutch has a lower overall proportion of CRUP words than English, so it is not obvious that there will be a CRUP advantage. Only four percent of the Dutch words starting with the prefixes *be-, ge-, ont-,* or *ver-* are CRUP words; and the overall percentage is probably even lower, because words beginning with other Dutch prefixes have fewer competitors with pseudoprefixes or bound stems. Some other differences between Dutch and English will be described below, as they are relevant to the pattern of results.

3. Selection of Critical Stimuli

In order to compute UPs and CRUPs, the appropriate cohort needs to be clearly defined. We began by selecting four productive Dutch prefixes (*be-*, *ge-*, *ont-*, and *ver-*). For each one, a computer program searched through the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995), compiling a list of the items marked as consisting of the prefix plus a monomorphemic root. The point at which the pronunciation of each root diverges from the other roots *in this prefix-specific list* was determined on the basis of the phonological transcriptions; this is the CRUP for each word. This excludes pseudoprefixed words and words with bound roots, because they are explicitly to be excluded according to the definition of the CRUP. In order to find the standard UP for each prefixed word, a program simply determined the point at which each word diverges phonologically from all of the other words in CELEX. If the CRUP and the UP are not the same phoneme, the word is a CRUP word and we selected it for use pending our ability to match it with a control word (see below). In locating CRUPs and UPs, suffixed forms related to the word in question were excluded from consideration, following Marslen-Wilson (1984), Tyler, Marslen-Wilson, Rentoul, and Hanney (1988), and our earlier studies (Wurm, 1997; Wurm & Aycock, 2003; Wurm & Ross, 2001). Thus, for example, the UP of the spoken English word *distaste* is the [eɪ], in spite of the existence of the related word *distasteful.* We will return to this point below.

Twenty CRUP words were identified using this procedure. On average, the CRUP in these words was 1.2 phonemes prior to the UP. An additional 20 prefixed words were chosen to be matched as closely as possible on phonological properties (number of syllables, vowel lengths, and number of phonetic segments).

Words in the two conditions had exactly equal numbers of phonemes, and had UPs at precisely the same phoneme positions within the words. When measured in the sound files themselves, the mean UP was 581 msec for CRUP words and 587 for Control words, a difference that did not approach significance ($p = .85$). In addition, we attempted to match items as closely as possible on a number of variables known to affect lexical processing. These are described in the next section. The 40 critical stimuli are shown in the Appendix.

4. Calculation of other independent variables

Whole-word surface frequencies and root frequencies were gathered from CELEX (Baayen et al., 1995). Family sizes were computed as the number of compounds and derived words in CELEX in which a given stimulus word's root appears as a constituent (de Jong et al., 2000; Schreuder & Baayen, 1997). Item durations in msec were measured directly from the individual sound files.

As stated above, one of the purposes of this study was to examine whether the difference between CRUP words and Control words might be confounded with or reducible to differences in the difficulty of the disambiguation process leading to lexical identification. If CRUP words have smaller cohorts or cohorts primarily consisting of low-probability candidates, the advantage observed in previous studies for CRUP words might fall out naturally as a consequence of these cohort characteristics. We opted for Shannon's entropy H for quantifying the probability distribution of a cohort:

$$H = \sum_{i=1}^{C} p_i \log_2\left(\frac{1}{p_i}\right),$$

where C equals the number of words in the current cohort. p ranges over all the words in the cohort, and represents a given word's frequency divided by the summed token frequencies from CELEX (Baayen et al., 1995) of all the cohort members. In general, the entropy of a cohort is a measure of its average amount of information, and hence its informational complexity. For the present purposes, the entropy can be thought of as a token-weighted type count. A higher entropy value indicates that there are more lexical candidates in the cohort, or candidates that are more similar in frequency (leading to a smaller probability of identification for the target word), or both. In order to capture the temporal dynamics of lexical identification, we calculated the entropy for different points in the root, in an attempt to assess how information complexity at various points (as we proceed through the word, with a continuously decreasing cohort) is reflected in the response times.

4.1 Entropy measures

We will focus on three entropy measures, which we call pre-CRUP entropy, CRUP entropy, and late entropy. Each was calculated over all words in CELEX that matched the phonological representation of the target word from onset up to one of three positions: 1) two phonemes into the root morpheme (pre-CRUP entropy); 2) the CRUP location, or matched location in Control words (CRUP entropy); and 3) the final segment of the word (late entropy, a measure first explored in Kemps et al., 2005). There are substantial entropy differences associated with the prefixes themselves, but we controlled for these by our inclusion of the same prefixes, the same number of times, in the sets of CRUP and Control words.

Readers should note that these calculations were based on whole words (not root morphemes), and so at the CRUP of *discredit,* for example, the words *discrepant, discretion,* and so on would still have been included in this

Table 1. Summary Statistics (Means and SEMs) for Critical Stimuli

	Stimulus group	
Stimulus characteristic	CRUP words	Control words
Word frequency[a]	51 (19.9)	58 (28.8)
Root frequency[a]	5412 (1669)	8914 (6870)
Duration (msec)	707 (18.2)	726 (17.1)
Pre-CRUP entropy	2.24 (0.16)	2.25 (0.13)
CRUP entropy	1.45 (0.13)	1.67 (0.16)
Late entropy	0.80 (0.14)	0.85 (0.15)
Family size	44 (10.3)	28 (6.5)

[a] per 42 million tokens, from the CELEX database (Baayen et al., 1995).
Note. None of the stimulus-group differences approached significance (all *p*s > .28).

Table 2. Regressor Intercorrelations

	Root freq.	**Family size**	**Duration**	**Entropy**		
				Pre-CRUP	CRUP	Late
Word frequency[a]	.03	.13	−.25	−.11	−.33*	−.14
Root frequency[a]		.60***	−.13	−.13	.00	.22
Family size			−.15	−.10	−.04	.18
Duration				.04	−.04	.11
Pre-CRUP entropy					.13	.29
CRUP Entropy						.49**

[a] per 42 million tokens, from the CELEX database (Baayen et al., 1995).
* $p < .05$ ** $p < .01$ *** $p < .001$

calculation. In addition, these entropy measures are calculated blindly across the CELEX database, and thus include items such as morphologically-related continuation forms that were deliberately excluded from the UP and CRUP computations described above. This is why late entropy need not be zero. Readers should also note that "pre-CRUP" and "CRUP" in our entropy labels refer simply to locations in a word. It is not appropriate to think of these quantities as characteristics or properties of CRUP words in particular. The quantities apply in the same way for Control words.

Table 1 shows mean values for the two stimulus types on several variables. Note that the values of the three entropies decrease the further into the root we measure. For none of the measures in Table 1 did any group difference approach significance (smallest $p = .28$).[5] This means that any differences in processing between CRUP words and Control words cannot be due to differences in entropy, frequency, or family size that exist among the stimuli, because there are no such differences. We also ran a logistic regression analysis to determine whether membership in the group of CRUP words could be predicted by any of the variables shown in Table 1. No variable was retained as a predictor of group membership (smallest $p = .31$), so it appears that the information carried by the listed variables, plus any information carried by status as a CRUP word, are complementary. Table 2 shows the regressor intercorrelations.

5. Auditory lexical decision experiment

5.1 Method

5.1.1 *Participant*

Participants were 40 members of the subject pool at the Max Planck Institute for Psycholinguistics. All were native speakers of Dutch with normal hearing. Each received five euros for his or her participation.

5.1.2 *Materials*

The 40 critical stimuli, described in the previous section, are listed in the Appendix. An additional 114 filler words were included to provide variety in the morphological structure of the stimuli. Some of the filler items were included as part of a separate study. The filler words consisted of 20 verb stems, 20 infinitives (verb stems plus the suffix *-en*), 34 past tense third person singular forms (verb stems plus the suffix *-te* or *-de*), 20 monomorphemic nouns, and 20 plural nouns (noun stems plus the suffix *-en* or *-s*). An equal number of pseudowords were constructed by taking existing Dutch words with the same morphological structures as our 154 real word stimuli, and changing a single sound equally often in the beginning, the middle, and the end of the word.

Each stimulus was read aloud by a native speaker of Dutch who was unfamiliar with the purpose and hypotheses of the study. Stimuli were digitized at a sampling rate of 16 kHz, low-pass filtered at 7.8 kHz, and stored in individual computer files.

5.1.3 *Procedure*

Participants were tested individually in a quiet room. The 308 stimuli were presented in a different random order for each participant. The digitized speech files were played for the participants over headphones at a comfortable listening level. Participants were directed to make a speeded lexical decision about each item. Each participant made responses on a button box, pressing one button for words and another for pseudowords. The "Word" response button was always pressed with the participant's preferred hand. Reaction times (RTs) were measured from the UP of each word. This allowed direct comparison to the related studies conducted in English (Wurm, 1997; Wurm & Aycock, 2003; Wurm & Ross, 2001), but there are empirical and theoretical reasons for doing this as well. First, the UP is the earliest moment at which a participant can be sure of what he or she is hearing. Second, Tyler, Moss, Galpin, and Voice (2002) found that a word's semantic information is already fully activated by the identification point, a recognition measure from the gating paradigm that is strongly correlated with the UP. Finally, Gaskell and Marslen-Wilson (1997) demonstrated with simulations that semantic ambiguity is reduced to zero at the UP of a word.

Before the main experiment, participants heard a practice list of similar composition and performed lexical decisions. The 10 practice items were not used in the main experiment. The duration of the experiment was approximately 25 minutes.

5.2 Results and Discussion

Two participants were excluded from the analyses for inaccurate responding (> 50% errors). Three items (two Control words and one CRUP word) were excluded from the analyses. One was included in the experiment because of a transcription error, and two were categorized as pseudowords by more than half of the participants. Finally, we excluded from the analysis extremely fast RTs (less than 150 msec after the UP; < 1% of the data) and extremely slow RTs (longer than 2000 msec post-onset; 3.0% of the data). Analyses reported below were conducted on the 1207 RTs that remained in the data set following these exclusions.

We analyzed the data set with a linear mixed-effect (multilevel) analysis of covariance with log RT as the dependent variable and Subject and Item as crossed random effects (Bates, 2005; Bates & Sarkar, 2005; Pinheiro & Bates, 2000). In addition to the numerical predictors listed in Table 1 and the Stimulus type factor (i.e., CRUP or Control), we included as a predictor the position of a stimulus in the (randomized) presentation lists, in an effort to minimize variance attributable to practice or fatigue effects. In our analysis, we did not impose linearity *a priori,* but explored potential non-linearities (cf. Baayen, 2005; Harrell, 2001) by allowing quadratic terms into the model. The frequency and family size variables were transformed logarithmically in order to remove the skewness in their distributions and to minimize the effect of atypical outliers.

Table 3 shows the results of this stepwise analysis. Figure 2 visualizes these results, showing the partial effect of each predictor when the other predictors are held constant at their medians. The effects for CRUP words and Control words coincide in all panels except for family size, where the dashed line

Table 3. Summary of Analysis for Variables Predicting Log Reaction Time (All Items)

Variable	***B***	**SE B**	***β***	***t***
Main effects				
Stimulus type	0.4175	.1300	–	0.55
Position	−0.0003	.0001	−0.008	−2.91**
Word frequency	−0.0534	.0162	−0.251	−5.03***
Pre-CRUP entropy: linear	−0.8951	.2602	−1.357	−3.44***
quadratic	0.1893	.0582	1.284	3.26**
CRUP entropy	0.1773	.0407	0.259	4.36***
Late entropy: linear	0.0609	.1069	0.089	0.57
quadratic	−0.1389	.0555	−0.394	−2.50*
Family Size	0.0293	.0262	0.073	1.12
Interaction with Stimulus type[a]				
Family size	−0.1377	.0039	–	3.51***

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Note. The degrees of freedom for each *t*-test = 1196.

[a] Contrast coding was used for the Stimulus type factor. CRUP words were the reference group by default, so the family size effect for these items equals the family size main effect coefficient (0.0293). This is not significantly different from zero ($p = 0.26$). The coefficient for the Control words equals this same coefficient plus the coefficient for the interaction with stimulus type (i.e., 0.0293 + (−0.1377) = −0.1084). This facilitative effect is significantly different from that of the CRUP words ($p < .001$; see center-left panel of Figure 2).

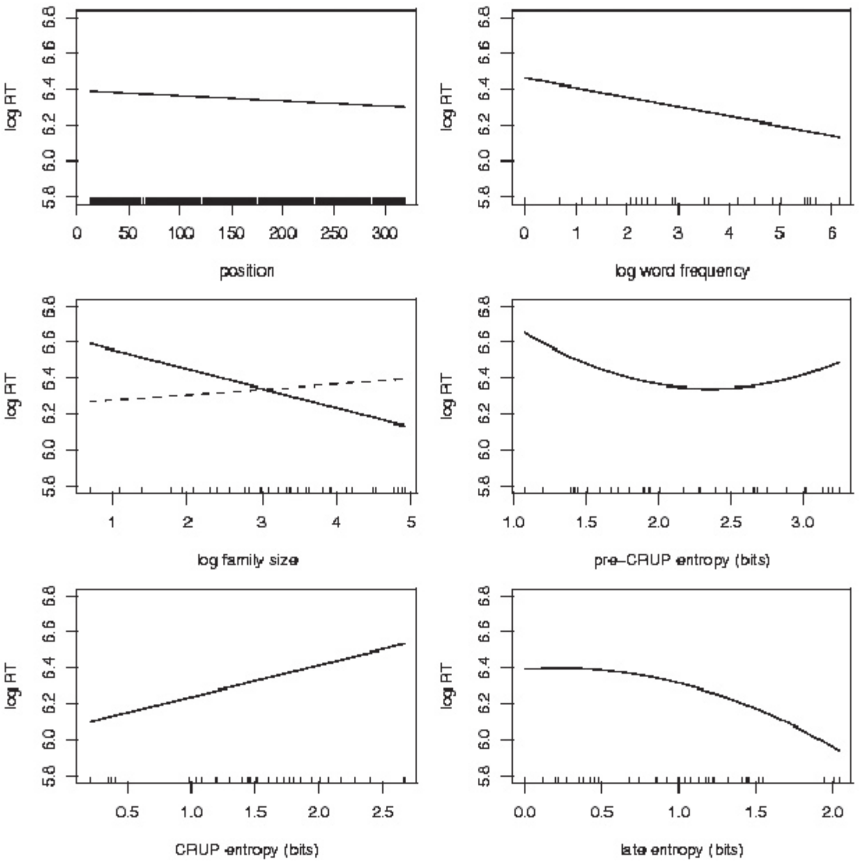

Figure 2. Partial effects of the predictors (when the other predictors are held constant at the median). Effects for CRUP words and Control words coincide in all panels, except for log family size, where the dashed line represents the CRUP words.

represents the CRUP words. As the vertical axis plots the same range of (log) RTs for each predictor, Figure 2 provides immediate insight into the relative effect sizes of the predictors. The standard deviations estimated for the random effects of subject and item were 0.206 and 0.099 respectively, and the estimate for the residual error was 0.301.

As shown in the upper left panel of Figure 2, participants responded faster as the experiment progressed. Log word frequency revealed the expected facilitative relationship with RTs (upper right panel). The higher the *a priori,* unconditional probability of a word is, the faster it can be responded to (cf. Rubenstein & Pollack, 1963, Baayen et al., 2003; Wurm, 1997; Wurm et al., 2005).

The three measures with which we hoped to probe the temporal dynamics of developing cohorts all turned out to be significant predictors. Pre-CRUP entropy was present with a U-shaped non-linear effect, as can be seen in the center right panel of Figure 2. For pre-CRUP entropies up to 2.5 bits, facilitation is observed, but for higher entropies, inhibition (a re-analysis using restricted cubic splines instead of a quadratic polynomial confirmed the presence of both facilitation and inhibition). Such a U-shaped pattern has recently been observed for the family size effect as well (Baayen, 2005) and suggests that two opposing forces are at work in lexical decision (cf. also Grainger & Jacobs, 1996). On the one hand, large cohorts with many equiprobable cohort members contribute to the general likelihood that the incoming stimulus is a word. On the other hand, the larger a cohort is, and the more equiprobable its members, the smaller the probability of the target in the cohort becomes. At our early probe position these two opposing forces (high general lexicality but also high difficulty of disambiguation) seem to find an optimum at 2.5 bits. It is at this point that RTs are fastest.

When we measure entropy at later points in time, the relevances of general lexicality and difficulty of disambiguation (as reflected in the response times) appear to change. Entropy at the CRUP position (or the matched position in the Control words) inhibits RTs, as shown in the lower left panel of Figure 2. At this critical point near the uniqueness point, winnowing down the cohort to the most probable candidate is all-important. Larger cohorts with more equiprobable members render this selection process more difficult, resulting in longer reaction times. The benefits of general lexicality that earlier in the word allowed facilitation are apparently outweighed by the importance of narrowing the cohort down to the target word.

The effect of entropy is again facilitative when measured at word offset, as shown in the lower right panel of Figure 2. One way to understand this effect is to suppose that the prefix and stem have been identified by this point, and what remains in the cohort are the word itself and its morphological continuation forms. Higher late entropy implies more of such continuation forms, and more activation in this particular group of related words. Note that the entropies themselves are fairly low here. At this point in the stimulus, the perceptual system can be quite certain which morphological family is in question, so once again general lexicality information can speed responding. Disambiguation is just about completed, and the remaining cohort members begin to affect RTs in a way much like pre-CRUP entropy did in the range of 0 to 2 bits (exactly the range for which we have late entropy values).

The only significant interaction in our analysis is shown in the left center panel of Figure 2. The effect of family size for CRUP words was not significant, whereas Control words showed the expected facilitation from larger families. The facilitatory effect for the Control words, which represent the overwhelming majority of Dutch prefixed words, is in line with the family size effects observed for the visual modality in previous studies for simplex words (e.g., Schreuder & Baayen, 1997) and suffixed words (e.g., Bertram et al., 2000; de Jong et al., 2000).

The present experiment is the first to reveal a family size effect in auditory comprehension for complex words. Because words with a large family provide partial matches with many morphologically related words in lexical memory, the joint probability distribution of the family members adds to the general probability that the incoming stimulus is indeed an existing word. Surprisingly, a large family size does not offer such an advantage for CRUP words. Why would this be the case?

Wurm's (2000) results suggest that bound stems are less easily detected in the auditory stream than free stems. CRUP words differ from matched Control words in that their cohorts only contain words with pseudoprefixes and with bound stems before the actual UP is reached, whereas the cohorts for the matched controls contain free stems up to the UP. The independently attested reduced probabilities of identification for bound stems (Wurm, 2000) imply that CRUP words have higher probabilities in their cohorts compared to their controls. Therefore, CRUP words are more likely to allow a yes-response in auditory lexical decision on their own, independently of any facilitation that could be provided by the other members of the morphological family.

This may explain the interaction of stimulus type and family size. In the case of Control words, the joint probability of the morphological family members apparently contributes to raising the probability of the target stem. Crucially, family size is a significant predictor over and above late entropy for these words. Recall that late entropy accounts for the facilitation by the morphological continuation forms. Hence, the effect of family size is carried primarily by the morphological family members that are not continuation forms. For example, words such as *workable* and *unworkable* belong to the same morphological family as *rework,* and contribute to its probability of identification; but they are not continuation forms of *rework.* For CRUP words, the identification process has proceeded further, apparently leaving no room for additional disambiguation from the joint probability of the family members.

A question that arises at this point is to what extent the statistical model on which this discussion is based overfits the data. For a data set with only 37 different words, a model with 11 coefficients for the fixed effects, 10 of which describe properties of these 37 words, is suspect. In order to investigate potential overfitting, we made use of bootstrap validation. Since we are not interested in validation across subjects, we obtained the by-item RTs averaged over subjects, and reran the analysis using ordinary least squares regression. This regression supported all conclusions reached on the basis of the overall multilevel analysis reported in Table 3 (all p-values $<= 0.01$). Bootstrap validation with 200 runs revealed an optimism of 0.1518, indicating that the initial R^2 of 0.7954 should be adjusted downward to a bootstrap-validated R^2 of $0.7954 - 0.1518 = 0.6436$. In 70 of the 200 runs, all predictors were retained, and in 98 additional runs, only one parameter was dropped. This shows that the ordinary least squares model overfits the data only slightly. Hence, we have confidence that the observed effects are replicable. With respect to the multilevel model, we note that inclusion of word as random effect leads to a conservative analysis for lexical covariates that also guards against overfitting.

We performed some additional analyses specific to the CRUP words, as there is one covariate that is specific to this subset: the CRUP-to-UP distance. This is the difference (in milliseconds) between the CRUP and UP locations within a CRUP word. As this distance is 0 for Control words, a regression analysis including CRUP-to-UP distance for both CRUP and Control words makes no sense. CRUP-to-UP distance was found to have a linear relationship with overall item RT in English (Wurm & Ross, 2001), and we wanted to see

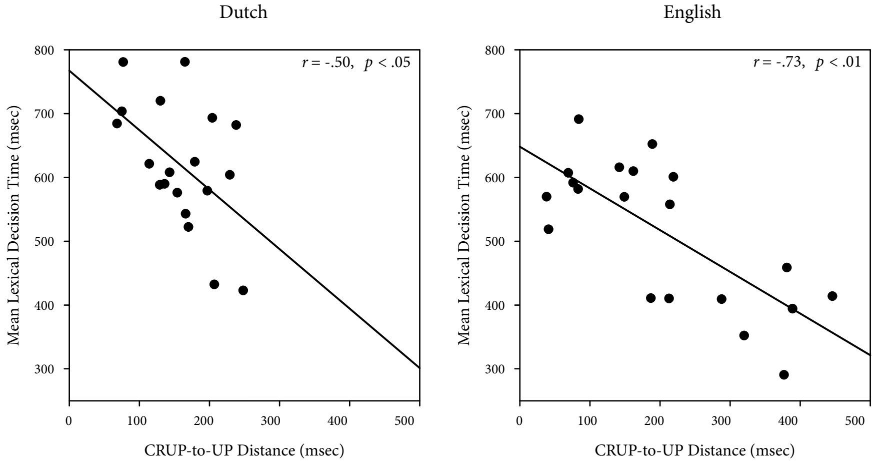

Figure 3. Mean lexical decision time (in msec) as a function of CRUP-to-UP distance (in msec). The left panel shows data from the current study. The right panel shows the lexical decision data from Wurm and Ross (2001). Each data point represents one item.

if this also held for Dutch. The left panel of Figure 3 shows that the simple bivariate relationship does hold. As is clear from the figure, there is a significant correlation between these two variables: Greater CRUP-to-UP distances are associated with faster RTs. We ran a multilevel regression analysis analogous to that described for the full data set, except that the CRUP vs. Control factor was obviously removed, and the CRUP-to-UP distance was added. The analysis showed that CRUP-to-UP distance was not significant as a main effect ($p = .27$), but it did interact with word frequency as shown in Figure 4 ($t(624) = -3.84$, $p < .001$). This conditioning plot (Cleveland, 1993) graphs log RTs as a function

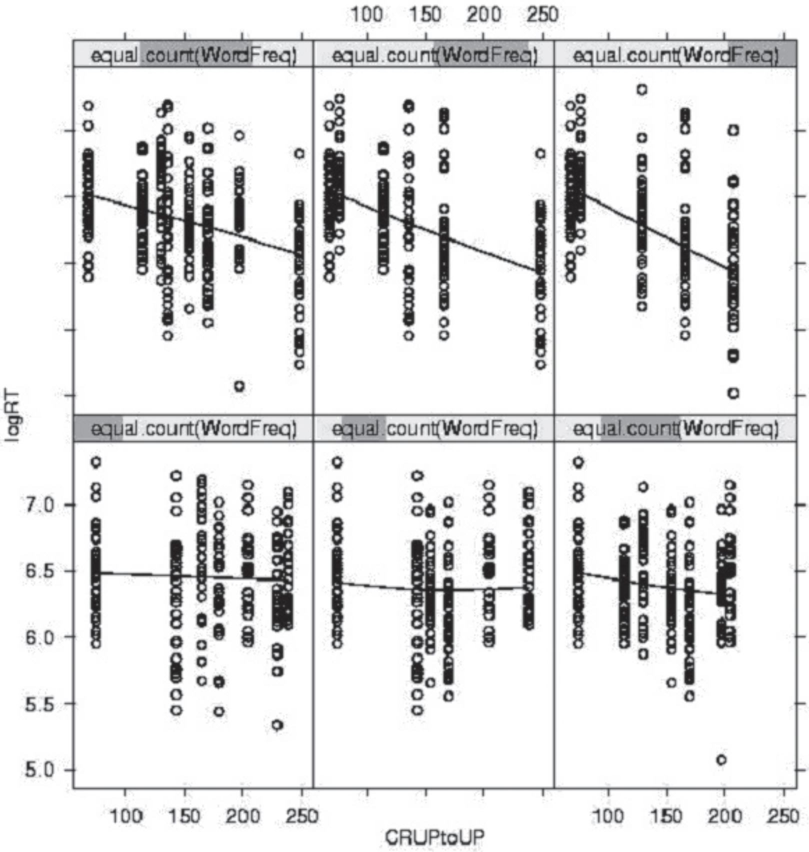

Figure 4. Log reaction time (in msec) as a function of word frequency and CRUP-to-UP distance (in msec). Data are for CRUP words only. In this conditioning plot log RT is graphed as a function of CRUP-to-UP distance, for subsets of the data that differ on word frequency. Frequency increases as one proceeds from the lower left panel to the right and then on the upper row of panels from left to right.

of CRUP-to-UP distance, for subsets of the data (with equal numbers of data points) that differ on word frequency. Frequency increases as one proceeds from the lower left panel to the right and then on the upper row of panels from left to right. The interaction indicates that there is very little effect of CRUP-to-UP distance for words with the lowest frequencies, but that as word frequency increases, a strong facilitative effect of CRUP-to-UP distance emerges. This result shows that the CRUP-to-UP effect demonstrated for English does hold in Dutch, at least for words that participants are presumably fairly familiar with.

Returning to Figure 3, the right panel plots RT against CRUP-to-UP distance observed for English by Wurm and Ross (2001). Comparing the two panels, we can see that the values for CRUP-to-UP distance in Dutch are only about half as large as those in English. This may help explain why we did not find an overall CRUP advantage (i.e., a main effect of stimulus type) in Dutch. It is possible that the main effect observed in English requires more time to build than is allowed by these shorter temporal distances that characterize Dutch. We were able to assess this hypothesis by reanalyzing the Wurm and Ross (2001) lexical decision time data, restricting the analysis to those CRUP words with CRUP-to-UP distances less than 250 msec. When we did this, their previously significant CRUP advantage shrank by two-thirds and was no longer significant.

6. General discussion

This study addressed the auditory processing of prefixed words. We probed this topic by means of measures gauging the entropy of the cohort (Kemps et al., 2005) at three positions in the prefixed word, by means of morphological family size (Schreuder & Baayen, 1997), and by means of the CRUP construct (Wurm 1997; Wurm & Aycock, 2003; Wurm & Ross, 2001).

Our study with Dutch prefixed words revealed that the three measures of cohort entropy are significant predictors of auditory lexical decision times. When the entropy of the cohort is measured early in the stem, as well as when it is measured at word offset, a negative correlation is observed: A higher entropy is associated with shorter response latencies. However, at the conditional root uniqueness point (or matched point in Control words), the effect of cohort entropy reverses, such that a higher entropy gives rise to longer response latencies.

Our results suggest that for lexical decision, two types of information provided by the cohort influence response latencies. One is of course the entropy itself, which is a function of the size of the cohort and its distribution of frequencies. The other is the point within the auditory signal at which this entropy is calculated. High entropy relatively early in a word is not necessarily a bad thing, because a large cohort is an index of general lexicality supporting a yes-response in lexical decision. Indeed, our results showed that two segments into a word's stem, this general lexicality outweighs the cost of disambiguation (up to a pre-CRUP entropy of 2.5 bits; as entropy continues to rise past this level, though, higher entropies do slow processing times). High entropy at the end of a word can be seen as beneficial, as well: At this point, the only alternatives left in the cohort are morphological continuation forms of the same word. Thus our late entropy effect, too, was facilitative.

A cohort that has nearly converged to a single high-probability candidate also allows a yes-response in lexical decision, as uncertainty about what word has been heard has been reduced considerably. Our results showed that this is the state of affairs nearer the uniqueness point of a word (specifically, at the CRUP or the matched point in a Control word). Here, as the information carried by the acoustic signal is nearly sufficient to narrow the cohort down to a single candidate, winnowing out irrelevant candidates is much more important than general lexicality. Higher entropy at this point in the word (our CRUP entropy) inhibits response speed.

Further insight in the temporal dynamics of auditory word recognition was provided by the morphological family size measure. The present study is the first to reveal a family size effect in auditory comprehension for complex words, a facilitative effect that is consistent with those found in previous research in visual word recognition. However, it was found that larger family sizes led to shorter response latencies only for Control words. CRUP words are high-probability lexical candidates already one or two segments before the uniqueness point. Apparently, they receive no additional benefit from the joint probability distribution of their morphological relatives, possibly because their probability of identification is already near ceiling. Alternatively, early disambiguation might give rise to earlier integration of the stem with the prefix. As a consequence, all family members beginning with a different prefix, as well as family members in which the stem is word-initial, might have been already discarded as irrelevant, leaving only the word's own morphological continuation forms as lexical candidates (as witnessed by the cohort entropy measured at word offset). In other words, due to rapid disambiguation and morphological integration for CRUP words, the overall family size counts may no longer be accurate, and hence they are not predictive.

Although there was no main effect of stimulus type in our study, from the interaction of stimulus type with family size, and the interaction of CRUP-to-

UP distance with word frequency, we conclude that the relevance of the CRUP construct is not restricted to English. Moreover, it is also not confounded with changes in the entropy of the cohort over time, as the CRUP words and their Controls show similar distributions of the three entropy measures. There was no trace of an interaction of stimulus type with these measures.

Compared to English, the temporal distance between CRUP and UP in Dutch is quite short. This difference may explain why in English CRUP effects appear to be much stronger. Given that there are so few CRUP words in Dutch, and given that the temporal distance between CRUP and UP is so small, the emergence of any effect of the CRUP remains surprising. Apparently, listeners are highly sensitive to the temporal dynamics of cohort reduction, even for reduction unfolding over short periods of time, and involving subtle processing differences between bound and free stems.

This study extends earlier empirical work on the role of cohort entropies (e.g., Kemps et al., 2005), morphological family size (e.g., Schreuder & Baayen, 1997), and the CRUP (e.g., Wurm, 1997), showing that all of these variables contribute to the auditory comprehension of prefixed words. Morphological family size is being included in an increasing number of empirical studies, but at present no formal theoretical models of speech perception or morphological parsing include it. Similarly, no formal model includes either the CRUP construct or the various cohort entropies that were included here.

In principle either the cohort model or the Shortlist model could presumably be modified to take quantities such as our entropy values into consideration, by some mechanism consistent with their respective architectures. However, our data also show that suffixed continuation forms of words exert an influence on their recognition, and it is difficult to reconcile this finding with the standard practice of excluding these words from the uniqueness point calculations. Similarly, a large body of data (the current study included) demonstrates that morphological structure matters perceptually, and as we noted in the Introduction, neither of these models does an adequate job of dealing with such structure. In summary, existing models are still far from accounting for the temporal dynamics of the auditory comprehension of prefixed words.

Notes

1. The CRUP account does not imply that listeners would fail to activate the root *cradle* upon hearing a neologism such as *dis-cradle.* Rather, it says that *cradle* is not marked as one of the roots that has combined with *dis-* in the past, and so it would not be on the special restricted list of roots that gets activated when the prefix *dis-* is heard. The root *cradle* would of course be activated upon the first encounter with this new word, but probably not until quite late (perhaps after the perceptual system has concluded that this acoustic pattern does not match any previously encountered).

2. Prefix likelihood is the frequency-weighted likelihood that a phoneme string is functioning as a prefix rather than a pseudoprefix (i.e., a string of letters or phonemes that sometimes functions as a prefix, but is not doing so in a given instance; for example, in English *re-* is a prefix, but the string does not serve as a prefix in the word *realize*). High values of prefix likelihood make a word an especially good candidate for decomposition.

3. The mean RTs in these two experiments were 441 msec and 610 msec, respectively. The mean item duration was 787 msec.

4. The mean RTs in these two experiments were 572 msec and 558 msec, respectively. The mean item duration was 860 msec.

5. We subsequently discarded three items from the analysis of the RT data in the main experiment. The smallest *p*-value is .17 when these three items are excluded.

References

Baayen, R.H. (2005). Data mining at the intersection of psychology and linguistics. To appear in A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 69–83). Hillsdale, NJ: Erlbaum.

Baayen, R. H., McQueen, J., 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. 355–390). Berlin: Mouton de Gruyter.

Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Baayen, R.H., Tweedie, F.J., & Schreuder, R. (2002). The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon. *Brain and Language, 81,* 55–65.

Bates, D.M. (2005). Fitting linear mixed models in R. *R News, 5,* 27–30.

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

Bertram, R., Baayen, R.H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language, 42,* 390–405.

Cleveland, W.S. (1993). *Visualizing data.* Summit, NJ: Hobart Press.

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

Gaskell, M.G., & Marslen-Wilson, W.D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

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

Harrell, F.E. (2001). *Regression modeling strategies.* New York: Springer.

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

Kostič, A., Markovič, T., & Baucal, A. (2003). Inflectional morphology and word meaning: Orthogonal or co-implicative cognitive domains? In R.H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 1–43). Berlin: Mouton de Gruyter.

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

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71–102.

Marslen-Wilson, W.D. (1989). Access and integration: Projecting sound onto meaning. In W.D. Marslen-Wilson (Ed.), *Lexical representation and processing* (pp.3–24). Cambridge, MA: MIT Press.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29–63.

Moscoso del Prado Martín, F., Kostič, A., & Baayen, R.H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition, 94,* 1–18.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.

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

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

Schreuder, R., & Baayen, R.H. (1997). How complex simplex words can be. *Journal of Memory and Language, 37,* 118–139.

Taft, M., Hambly, G., & Kinoshita, S. (1986). Visual and auditory recognition of prefixed words. *Quarterly Journal of Experimental Psychology, 38A,* 351–366.

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.

Tyler, L.K., Moss, H.E., Galpin, A., & Voice, J.K. (2002). Activating meaning in time: The role of imageability and form-class. *Language and Cognitive Processes, 17,* 471–502.

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. (2000). Auditory processing of polymorphemic pseudowords. *Journal of Memory and Language, 42,* 255–271.

Wurm, L.H., & Aycock, J. (2003). Recognition of spoken prefixed words: The role of early Conditional Root Uniqueness Points. In R.H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 259–286). Berlin: Mouton de Gruyter.

Wurm, L.H., Baayen, R.H., & Aycock, J. (2005). Surface frequency effects below the threshold: Comparing types, tasks, and modalities. Manuscript submitted for publication.

Wurm, L.H., & Ross, S.E. (2001). Conditional root uniqueness points: Psychological validity and perceptual consequences. *Journal of Memory and Language, 45,* 39–57.

Author's address:

Lee H. Wurm
Department of Psychology
Wayne State University
5057 Woodward Ave. (7th floor)
Detroit, MI 48202, USA

E-mail: lee.wurm@wayne.edu

Appendix: Critical Stimuli

CRUP words

behelp	bə-'hɛlp	'make do'	gewed	xə-'ʋɛt	'bet'
benut	bə-'nʏt	'utilize'	gezind	xə-'zɪnt	'disposed'
bevecht	bə-'vɛxt	'fight'	ontval	ɔnt-'fɑl	'let slip'
bewoon	bə-'ʋon	'inhabit'	vergok	vər-'xɔk	'gamble away'
bezuip	bə-'zœyp	'get drunk'	vergun[a]	vər-'xʏn	'permit'
gebeten	xə-'betə	'bitten'	verhonger	vər-'hɔŋər	'starve'
gedram	xə-'drɑm	'nagging'	verschoon	vər-'sxon	'change'
genummerd	xə-'nʏmərt	'numbered'	verslap	vər-'slɑp	'relax'
gepost	xə-'pɔst	'posted'	versleep	vər-'slep	'drag'
geprikt	xə-'prɪkt	'pricked'	verwelk	vər-'ʋɛlk	'wilt'

Control words

bespan	bə-'spɑn	'stretch'	gebel	xə-'bɛl	'ringing'
belet	bə-'lɛt	'prevent'	gemoord[a]	xə-'mort	'murdered'
beheks	bə-'hɛks	'bewitch'	ontzeg	ɔnt-'sɛx	'refuse'
bewaak	bə-'ʋak	'guard'	verkas	vər-'kɑs	'relocate'
behaag	bə-'hax	'please'	vergal	vər-'xɑl	'spoil'
gebeden	xə-'bedə	'prayed'	verbitter	vər-'bɪtər	'embitter'
geklit	xə-'klɪt	'sticking together'	verkleur	vər-'klør	'discolor'
gewaggeld	xə-'ʋɑxəlt	'tottered'	verklap	vər-'klɑp	'give away'
gespit	xə-'spɪt	'dug'	verslaap[a]	vər-'slap	'sleep away'
geknipt	xə-'knɪpt	'cut'	verwond	vər-'ʋɔnt	'injure'

[a] Item was excluded from the analyses (see the Results section).