

Evidence for early morphological decomposition in visual word recognition: A single-trial correlational MEG study

Olla Solomyak¹ and Alec Marantz¹

Corresponding Author:

Olla Solomyak
olla.solomyak@nyu.edu

New York University
Department of Linguistics
726 Broadway, 7th floor
New York, NY 10003

¹ New York University

Abstract

We employ a single-trial correlational MEG analysis technique to investigate early processing in the visual recognition of morphologically complex words. Three classes of affixed words were presented in a lexical decision task: free stems (e.g. taxable), bound roots (e.g. tolerable) and unique root words (e.g. vulnerable, the root of which does not appear elsewhere). Analysis was focused on brain responses within 100-200ms post-stimulus onset, in the previously identified letter string and visual wordform areas. MEG data were analyzed using cortically constrained minimum-norm estimation. Correlations were computed between activity at functionally defined regions of interest and continuous measures of the words' morphological properties. ROIs were identified across subjects on a reference brain and then morphed back onto each individual subject's brain (N=9). We find evidence of decomposition for all three classes of affixed words at the M170 stage in processing. The M170 response is shown to be sensitive to morphological properties such as affix frequency and the conditional probability of encountering each word given its stem. These morphological properties are contrasted with orthographic form features (letter string frequency, transition probability from one string to the next), which exert effects on earlier stages in processing (~130ms). We find that effects of decomposition at the M170 can in fact be attributed to morphological properties of complex words, rather than purely orthographic and form related properties. Our data support a model of word-recognition in which decomposition is attempted, and possibly utilized, for complex words of all three classes.

1. Introduction

An extensive psycholinguistic tradition has studied morphological decomposition, exploiting in particular the lexical decision paradigm with different modes of stimulus presentation including masked and unmasked priming. This tradition has led to models of visual word recognition and storage that emphasize a number of complex statistical properties of stems, affixes and families of morphologically related words that might influence word recognition at different stages. For the last decade or so, however, neurolinguistic debate on morphology has centered around Pinker's "words and rules" model (Pinker, 2000) and the issue of single vs. dual route models of the access to complex words, where the dual route model has been explicitly linked to the contrast between procedural and declarative memory systems (Embick & Marantz, 2005; Pinker & Ullman, 2002; Ullman, 2004). While the psycholinguistic and computational modeling literature (Taft & Forster, 1975; Marslen-Wilson, Bozic, & Randall, 2008; Hay & Baayen, 2005) has examined sophisticated models of lexical access that attempt to come to terms with a wide range of conflicting data, the neurolinguistic literature has tended to investigate a more simplified view. Here frequency effects on behavior are associated with the declarative (associative) memory system and complexity effects with the procedural system. Where we see surface frequency effects for morphologically complex words — RT correlating with frequency of the complex form itself — this is taken to be evidence for whole word access through the declarative memory system. Where we see base frequency effects — RT correlating with the frequency of the stem of a complex word rather than the frequency of the complex form — this is taken as evidence for the declarative memory system accessing the base of the word and the

procedural memory system putting the pieces of the word together (see the recent critique of this view in Baayen, Wurm, & Aycocock, 2007).

Evidence from, e.g., masked priming (Rastle, Davis, & New, 2004) indicates that RT in lexical decision experiments may not always be the best evidence for the mechanisms underlying lexical decomposition, since decomposition effects are evident for morphologically complex words for which complexity itself does not directly modulate RT (see e.g. Zweig & Pylkkänen, in press, and below). MEG techniques allow us to follow (putative) stages of visual word recognition to evaluate models of complex word recognition. The MEG literature on word recognition clearly indicates that the brain can exhibit sensitivity to variables that do not significantly influence RT, even showing cases where speed up in brain responses correlates with a slow down in RT, reflecting processing costs at later stages (Pylkkänen, Stringfellow, & Marantz, 2002).

This study provides additional evidence from MEG for obligatory morphological decomposition of derived words in visual word recognition, independent of RT. We extend the range of constructions examined to include most of the common derivational morphemes of English and to encompass an investigation of bound roots (like *toler-* in *tolerable* and *tolerate*) and unique bound roots (like *ten-* in *tenable*, which do not occur in other English words). We employ a single trial correlational approach (following Hauk, Davis, Ford, Pulvermuller, & Marslen-Wilson, 2006) to analyzing distributed source data from MEG as calculated on the cortical surface of each subject extracted from his/her structural MRI data. This approach allows us to evaluate the effect of two continuous variables implicated in decomposition — the

transition probability between stem and affix and affix frequency — and compare them to strictly orthographic but related (and correlated) variables — the orthographic transition between the last letters of the stem and the first letters of the affix and the positional letter frequency of the affix.

Taft and Forster published the most influential work on morphological decomposition in word recognition starting in the 1970's (e.g. Taft & Forster, 1975). Their model supposed a stage of blind affix stripping, during which letter strings that corresponded to a possible affix were stripped from the stem and the resulting stem subjected to lexical look-up. Successful discovery of a lexical entry for the stem led to a stage of recombination, where grammatical and lexical knowledge was consulted to evaluate the combination of stem and affix.

Despite the force of Taft and Forster's experiments, the behavioral literature since the 1970's paints a mixed picture on morphological decomposition. In conscious priming experiments, the apparent lack of significant behavioral priming for, e.g., pairs like "darkness" and "darkly" or irregular past tense to stem pairs like "taught" and "teach," despite the close semantic relation between the members of the pairs and the robustness of immediate semantic priming, pointed to interference effects between lexical access and, e.g., lexical decision (see the discussion in Stockall & Marantz, 2008). In languages like English, there seems to be no simple effect of morphological complexity on RT in naming or lexical decision, while in languages like Finnish, in some cases there is a robust inhibitory effect of complexity associated with inflectional morphology (Lehtonen et al., 2007; Lehtonen, Vorobyev, Hugdahl, Tuokkola, & Laine, 2006). For psycholinguists proposing models responsive to the literature, the complexity of the results points to theories in which decomposition is at least always attempted, perhaps in

parallel with whole word recognition, and in which distinct processing stages leading up to word recognition are recognized, allowing contextual task variables to influence processing at different levels.

The evoked response literature has elucidated the early stages of reading, with MEG studies providing spatial and temporal information of early processing consistent with the results from ERP and fMRI. Concerning the spread of stimulus-evoked brain activity in word recognition, the ERP and MEG literature is generally consistent with the description in Marinkovic (2004). Although there is much agreement in the literature on the spreading activation stimulated by visually presented words, the evoked response components consistently observed in visual word recognition have not been unambiguously identified with the task-specific computations they index, if any, within computational models. The visual M100 response (the “Type I” response of Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999), from visual cortices in occipital brain regions, shows sensitivity to features of visually presented words but has not been shown to differentiate between letter and symbol strings. A response component that peaks prior to 150ms after word presentation, called the Type II response in (Tarkiainen et al., 1999), does differentiate between letter and symbol strings, but did not appear in that work to be sensitive to the frequency properties of letter strings. Around 170ms post stimulus onset, the M170 (the MEG counterpart to the N170 evoked by visually presented objects) from the ventral temporal lobe shows sensitivity to a set of frequency properties of letter strings. This response corresponds to the fMRI response from the so-called visual word form area (McCandliss, Cohen, & Dehaene, 2003). In Solomyak & Marantz (in press), we present some evidence that this response does not represent contact with the mental

lexicon; semantic or other true lexical properties of words do not affect activity at the M170. However, this conclusion remains controversial, and the issue is not central to the work on decomposition described here.

Within 200ms of visual word presentation, activation begins in inferior frontal regions, and spreads through middle and superior temporal regions. The MEG literature has shown that an M250 response from medial and superior temporal areas is sensitive to phonotactics, while the M350 shows hallmarks of being an index of lexical access — contact with the mental lexicon that connects orthographic, phonological and semantic representations along with other information about words (see the discussion in Pylkkänen & Marantz, 2003). The M350, and possibly also the M250, seem to contribute to the N400 of the ERP literature, which may begin around 200ms post stimulus onset and peak around 400ms.

The majority of psycholinguistic data support a version of either a full decomposition model of lexical access or a parallel dual route model. The decomposition model most consistent with the data to be presented below is sensitive to information about the statistical relation between stems and affixes in the decomposition stage of processing, prior to lexical access (our understanding of the potentially relevant statistical variables here is due to the important work of J. Hay on morphological parsing (Hay, 2001; Hay & Baayen, 2001; Hay & Baayen, 2003; Hay & Baayen, 2005)). This means that the visual word forms of stems and affixes play a role in decomposition, independent of the lexical entries (connection to semantics and phonology) for these morphemes. The dependence of this sort of decomposition model on statistical information usually conceptualized as belonging to lexical entries makes the model harder to distinguish in

claims from a parallel dual route model, since variables correlated with surface frequency (e.g., the transition probability between stem and affix and that between affix and stem) play a role at various stages in processing. In this experiment, we do not attempt to distinguish between dual-route theories that propose that the recognition system attempts to parse every word that appears to contain an affix and the obligatory decomposition theory, which proposes that decomposition is required for recognition of every morphologically complex word.

Conscious priming, where the prime is recognized as well as the target, has yielded a complex set of data for models of lexical access. In any decomposition model of lexical access, conscious recognition of a prime might lead to primed activation of the word form representation of all or part of the target, either through identity or through spreading activation of semantically or phonologically/orthographically related lexical representations. Entries activated by the prime could also serve as competitors in any decision process or as inhibitors, on models in which there is lateral inhibition among lexical entries. The properties of masked unconscious priming, however, suggest that the masked prime (with sufficiently brief prime presentation) only begins activation of visual word forms and generally does not activate lexical entries. Thus masked priming reveals aspects of early processing stages of visual word recognition. Recent results from masked priming studies strongly suggest that affixes are stripped from stems in visual word recognition independent of the actual morphological structure of the prime. So, “brother” primes the morphologically unrelated “broth,” while “brothel” does not prime “broth,” despite the same orthographic overlap (Rastle & Davis, 2008; Rastle et al., 2004). Longer presentation of a prime is necessary to produce semantic priming, where semantic priming is an indication of the prime’s contact with the lexicon, as opposed to the storage of visual word form representations. The

MEG study of Zweig and Pylkkänen (in press) supports this interpretation of the masked priming literature. They find an effect of morphological complexity at the M170 response to visually presented words, suggesting a neural processing cost of the decomposition.

However, although Zweig and Pylkkänen show that brain activity at 170ms post-stimulus onset is modulated by morphological complexity, their complexity variable did not affect RT. In Taft's model, decomposition was essentially costless, and recent work on English and Finnish suggests that an RT cost of variables related to morphological complexity may be attributed to the recombination stage of processing in a full decomposition model (Lehtonen et al., 2007; Lehtonen et al., 2006; Taft, 2004). We must allow for the possibility that routine mental computations involve consistent variation in neural costs that do not generally translate into measurable behavioral costs, either because they are too insignificant to affect behavior or because the computation stream is timed to allow more effortful computations at one stage to catch up with easier computations before the next stage.

The experiment reported here tests the predictions of a class of full decomposition models and parallel dual route models sharing certain properties and contrasts these models with those that claim that derived words are decomposed after recognition if at all. In the first stage of processing, potential affixes are recognized by form and parsing between stem and affix is attempted based on activation of the (visual) word form of the stem. Thus we expect affix-specific variables such as affix frequency to correlate first with brain activity, followed by variables associated with parsing, such as the transition probability between stem and affix. Parsing of stem and affix leads to lexical access for the stem. At this stage, lexical variables

including lemma frequency should be relevant, even for bound roots like *toler-* in *tolerable*. Following lexical access to the stem, recombination of stem and affix occurs. Here, variables such as the transition probability between affix and stem and/or surface frequency should correlate with brain activity. While we will show evidence below that these proposed stages of word recognition do map onto time segments of neural activity that follow one another, we will not be testing claims about modularity here; we do not distinguish between strictly feed-forward processing and cascading interactive levels of computation. In addition, this study concentrates on the proposed stages of decomposition and stem access, i.e., the “early” stages of complex word recognition up to and including the MEG M350 response component.

2. Methods

2.1. Design and Stimuli

The current experiment employs a lexical decision task with simultaneous MEG recording in an attempt to investigate the time-course of linguistic processing for three categories of words. The first category of interest was that of free-stems — i.e., affixed words whose stem is also a word on its own, such as “predictable.” The second category consisted of bound-roots, such as “tolerable,” which contain a root that appears in other morphologically related words (e.g. “tolerate”). The last category consisted of words similar to those in the bound root class in that they can be parsed into a root and affix, with the affix apparently contributing syntactic category information to the word. However, the roots of these words arguably appear only in the words themselves; they contain “unique roots.” An example is the word “vulnerable,” the root of which

never appears without the particular suffix, “able.” Groups of words from each of these three classes comprised the target stimuli of the current experiment.

Nine suffixes for which there was a spread of words across the three categories of interest were chosen to be included in this study. These were: *able, al, ant, ary, ate, er, ic, ion* and *ity*. Stimuli for the experiment were chosen from the English Lexicon Project (ELP) word database (Balota et al., 2002).¹ Potential stimuli were limited to 7-11 letter words that ended in the same orthography as that of each suffix, and were either bi-morphemes (for the first two word categories) or mono-morphemes (for the unique root word category). Words were classified into the three groups of interest based on their parsing in the ELP.² All words were then coded according to their length, mean bigram frequency and HAL log-frequency, as reported by the ELP. Words in the first category were also coded for the surface frequency of the stem contained in the affixed word (e.g., the surface frequency of “predict” for the word “predictable”). For these words, the difference between their surface frequency and the surface frequency of their stems, which we will call affix-dominance, was calculated as well. An automatic word-choice algorithm was then employed, which randomly chose words from each group such that they were well and equally distributed over the above-mentioned variables. The word-choice algorithm operated with a dual purpose: to select sets of words that were well distributed over the relevant dimensions, and at the same time did not differ along these dimensions across category or affix-group. Six words were chosen from each category of words

¹ All items in the ELP for which the reported mean accuracy rate was below 30% were excluded prior to stimuli collection.

² The ELP clearly distinguishes between bi-morphemes that are free-stems and those that are bound-roots. Unique root words, on the other hand, are not clearly distinguishable from other ELP-coded mono-morphemes whose last few letters merely share the orthography of the suffixes. Part of speech was the only available coded characteristic on which to base this distinction. Words that were not of the same part of speech as were their transparently affixed counterparts were therefore excluded from the list of potential unique root words.

for each suffix, generating 162 target words (6 words x 3 categories x 9 suffixes). A detailed description of the word-choice algorithm is provided in Appendix A.

A group of 162 control words was then chosen, with 18 words matched by the frequency of their endings to each of the nine suffix frequencies.³ Suffix and ending frequencies were calculated using the CELEX database (Baayen et al., 1995). For this purpose, suffix frequency for an n -letter suffix was defined as the frequency of this letter sequence appearing at the end of a word. Potential control words were all 7-11 letter words in the ELP that both the ELP and CELEX considered to be mono-morphemes. Words with the same endings as the suffixes of the target words were excluded. For each of the nine suffixes, all words were extracted from the list for which the relevant ending frequencies were within 0.3 of the log-frequency of the suffix. The automatic word-choice routine was then used to select 18 of these words such that they would match the affixed words on their distributions over length, mean bigram frequency and surface frequency. A group of 324 non-words was then chosen to match the all the words on their overall distributions over length and mean bigram frequency. In addition to the 324 words and 324 non-words that were relevant to the current study, the experiment also contained 78 filler stimuli from another study.⁴

³ Words were matched for frequency of last bigram, trigram or tetragram, depending on the length of the suffix they were matched to.

⁴ The data collection and preliminary analysis methods described here are the same as those described in Solomyak and Marantz (in press), where the results of the study associated with the filler trials are reported.

2.1.1. Stimulus Properties

Various characteristics of the affixed words were calculated for use in correlational analyses. In order to define morphological properties of the affixed words it was necessary to identify each word's morphological family. For this purpose, the root of each affixed word (both free-stemmed and bound-root) was defined using the word's parsing in the ELP. All other words parsed as having the same roots as did the affixed words were then extracted from the ELP. This generated a preliminary list of family members for each affixed word.⁵ The CELEX database was then searched for all words that were defined to have these family members as their morphological constituents. These words were also added to each list of families. Compound words were then excluded from these lists, along with words that were clearly not morphological family members of the original stimuli.⁶

Frequency data were obtained for all family members. HAL frequencies were obtained from the ELP for those words that were included in the ELP database. Several words that were not found in the ELP were coded based on their frequencies in CELEX. Each target word's lemma frequency was defined as the summed frequency of all the words in its family. The lemma transition probability (TPL), which was defined as the ratio of each word's surface frequency to its lemma frequency, was then calculated for each affixed word.

⁵ Because of the nature of the ELP's parsing of the words "valuable" (which is parsed as "valu"+"able") and "rationality" (which is parsed as "rationale"+"ity") the words "value" and "rational," along with their family members, were manually added to their respective lists.

⁶ Clearly unrelated words were words such as "equestrian" for the target word "equable," the roots of which are orthographically, but not morphologically, the same.

As a contrast to the morphological variable TPL, another variable was calculated as a measure of the orthographic transition between the root and the affix in each affixed word. This variable, which we will call bigram transition probability or TPB, was defined as the frequency of the first two letters of the suffix given that the preceding two letters, i.e., the last two letters of the root, appear in their position relative to the end of the word. Bigram and tetragram frequencies relevant to this calculation were obtained from CELEX.

Affix frequency was also calculated in two ways, to contrast between morphological and orthographical properties related to affixation. The first measure was the same as was used for the matching of the control words — the frequency of the letters of the suffix appearing at the end of a word, regardless of the word’s morphology. The second measure of affix frequency was the morphological frequency of the affix, which only included the affix as a morpheme and not as an orthographic ending to a word. This measure included cases in which the morpheme of interest was followed by another affix, e.g., the word “predictability” was counted towards the frequencies of both *able* and *ity*. Morphological affix frequencies were calculated from the ELP database and were based on the ELP’s morphological parsing. The morphological and orthographic affix frequencies calculated for the nine suffixes in the experiment are presented in Table 1. Table 2 shows the correlation coefficients between the variables of interest for comparison.

Table 1: Orthographic and morphological affix frequency counts, presented in units of natural log frequency per million

	Orthographic Affix Frequency	Morphological Affix Frequency
able	7.48	9.11
al	9.23	10.18
ant	7.73	8.35
ary	7.30	7.99
ate	7.94	10.09
er	10.21	10.41
ic	7.96	9.18
ion	9.22	10.56
ity	8.00	9.31

Table 2: Correlation coefficients between key variables to be compared in correlational analyses, separated by word category.

	Lemma * Surface Frequency	TPL * TPB	Affix * Last n-gram Frequency
Free-Stems	0.52	0.32	0.81
Bound-Roots	0.74	0.28	0.81
Unique Root Words	N/A	N/A	0.81

2.2. Experimental Procedures

Participants were nine right-handed native English speakers, ranging in age from 19 to 29, with a mean age of 23.3. All subjects provided informed consent, and were paid for their participation. The subjects were prior participants of an MRI experiment at a collaborating lab, and therefore had structural MR images that were available for analysis. Human subjects approval for the study was obtained from the IRB at both NYU and MIT. Experimental stimuli were projected onto a screen above participants' heads while they lay in the magnetically shielded room that houses the MEG. Each trial began with a fixation point (“+”) that appeared on the screen for 500ms, after which the stimulus was presented for 300ms. Subjects then responded to the stimulus by pressing one of two buttons to indicate whether or not they recognized the stimulus as a word.

MEG data were acquired continuously throughout the duration of the task. A 157-channel axial gradiometer whole-head MEG system (Kanazawa Institute of Technology, Kanazawa, Japan), with a sampling frequency of 1 kHz, was used for the recording. Each subject's head shape was digitized prior to MEG acquisition, and HPI (head position indicator) coils were used to localize the position of the subject's head inside the MEG helmet. In addition to the MEG data recorded during the experiment, a structural MRI for each subject was acquired from a collaborating lab. The head-shape digitization and HPI locations were later used to match the MEG coordinate system to that of the MRI.

2.3. Analysis Methods

2.3.1. Minimum-norm Estimation

The MEG data processing and minimum-norm estimation procedures were the same as those previously described in Solomyak and Marantz (in press). MEG data were noise-reduced using the Continuously Adjusted Least-Squares Method (Adachi et al., 2001), in the Meg160 software (Yokogawa Electric Corporation and Eagle Technology Corporation, Tokyo, Japan). MNE (MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging, Charleston, MA) was then used to calculate a cortically constrained minimum-norm solution for each subject's MEG data. For this purpose, subjects' structural MRI's were processed in FreeSurfer (CorTechs Labs Inc., LaJolla, CA and MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging, Charleston, MA) to create a cortical reconstruction of each subject's brain. A source space of 5124 sources was then created on each subjects' cortical surface. The BEM (boundary-element model) method was used to compute the forward solution — an estimate of the resulting magnetic field at each MEG sensor, given activity at each of the 5124 sources. This forward solution was then employed in a computation of the inverse solution, which identified the spatio-temporal distribution of activity over these sources that would best account for each subject's grand-averaged MEG data across all stimuli.⁷ The resulting minimum-norm estimates were converted into a dynamic statistical parameter map (dSPM) — a noise-normalized measure of activation at each source that avoids some of the inaccuracies of standard minimum-norm calculations (see Dale et al., 2000). To identify regions of interest for correlational analysis, the

⁷ The average MEG data was baseline corrected using 100ms of pre-stimulus data and was low-pass filtered at 40Hz.

dSPM solutions were averaged across all subjects. Peaks in the average solution were used to define regions and time-windows of interest. An inverse solution was then computed over each subject's raw MEG data, individually over each identified region. As this computation was performed over the raw data rather than the grand-average, the result was a trial-by-trial minimum-norm solution, consisting of an estimate of activity per time-point at each source within the specified region.⁸ The resulting minimum-norm estimates were converted into noise-normalized dSPM values for use in subsequent correlational analyses.

2.3.1.1. Region of Interest Analyses

The averaged dSPM's exhibited a clear pattern of early activation, which consisted of an earlier (140-150ms) negative (i.e. current flowing inward) peak in a posterior occipital region, and a slightly later (180-190ms) positive (i.e. current flowing outward) peak in the occipito-temporal fusiform gyrus area. The earlier peak was defined as the M130 and the later peak as the M170. Labels were created to isolate these regions of activity in both the left and right hemisphere, both of which displayed the pattern of peaks described above. The left hemisphere activation for both the M130 and M170 is illustrated in Figure 1, while the labels identifying each of these regions can be seen in Figure 2.

As the pattern of M130 and M170 activation was fairly consistent across subjects, those who did not exhibit the typical peaks in a particular label were excluded from analysis of this region. One subject did not exemplify the typical pattern of M130 activity for the left

⁸ Only components of activation that were in the direction normal to cortical surface were retained in the minimum-norm solution.

hemisphere, and another for the right hemisphere: these subjects lacked a negative peak in the 100-190ms time range in the isolated region. Two subjects did not show the typical pattern of M170 activity for each hemisphere, as they lacked a positive peak in the 150-215ms time range in the relevant ROI.⁹ Subjects were excluded only from the analysis of the particular component and hemisphere for which their data were atypical. The mean M130 peak latency for the remaining subjects in each analysis was 146.1ms (with a range of 122 - 172ms) in the left hemisphere and 160.4ms (ranging from 142 - 177ms) in the right. The mean M170 latency was 190.9ms (ranging from 178 - 214ms) in the left hemisphere and 185.4ms (with a range of 164 - 204ms) in the right.

As the directionality of early MEG components has been fairly consistent, both in the current study and in our previous experience (Pykkänen & Marantz, 2003), M130 and M170 analyses were conducted on signed activity. As these early peaks were clearly defined, timing was defined relatively to each subject's peak, rather than in absolute terms. M130 and M170 analyses were conducted on the activation in each region, averaged over a 50ms window centered at each subject's peak latency. Analyses were also conducted to examine the rise, height and fall of the M170 peak. Three 20ms time-windows were identified for this purpose: a "rise" time-window spanning 30ms to 10ms prior to the peak, a "height" time-window centered at the peak latency, and a "fall" time-window spanning 10ms through 30ms after the peak. Trials for which the average activation over the relevant time-window was in the direction opposite to that of the associated peak (negative for the M130, positive for the M170), were excluded from

⁹ These were not the same two subjects for both hemispheres — one subject was atypical in both hemispheres, while two others were atypical in one hemisphere each.

analysis. The remaining mean activation values were normalized into z-scores within each subject, after which correlations were computed on the normalized activity across all subjects.

A network of later activation, located in the left hemisphere superior temporal and Sylvian Fissure regions, was identified as M350 component activity. The label identifying this network is illustrated in Figure 2. This region showed heightened negative activation beginning at about 250ms and lasting throughout the 300-400ms time range. M350 analyses were conducted on signed activation, averaged across all sources in the network. As individual subjects' M350 peaks could not be consistently identified, the timing of activation was defined in absolute terms, with no subjects excluded from analysis. M350 activity was examined by way of two-dimensional time-course analyses, in which correlation waves were computed to track the effects of interest over time. Correlations were computed time-point by time-point within a larger (300-400ms) range, after which the resulting correlation wave was subject to a correction procedure to account for the problem of multiple comparisons, based on the method introduced by Maris and Oostenveld (2007) and previously described in Solomyak and Marantz (in press). The procedure consisted of first defining a new statistic, $\sum r$, as the sum of all correlation coefficients within a temporal cluster of consecutive significant effects in the same direction (at the raw $p = 0.05$ significance level). A Monte-Carlo p -value was then computed, which indicated the statistical significance of the $\sum r$ statistic for each temporal cluster, given the multiple comparisons that gave rise to the effect.

Figure 1: Average left hemisphere dSPM activation across all subjects at three time-points of interest, shown on a representative subject's inflated cortical surface. Positive (red) dSPM values indicate current flowing outward from the cortical surface, while negative (blue) values indicate current flowing inward.

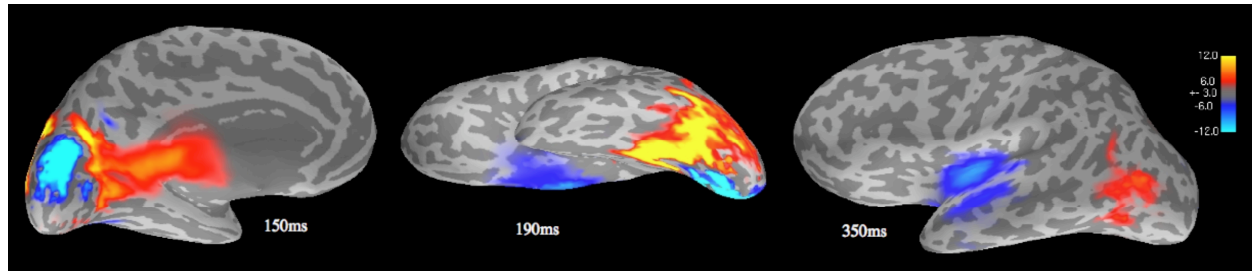
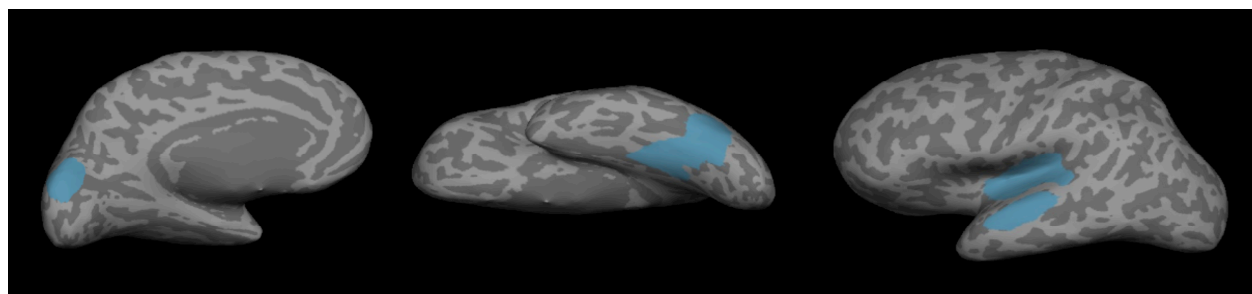


Figure 2: Left hemisphere regions of interest identified for correlational analysis, defined based on heightened activation in the isolated regions. Left: Medial view of the inflated cortical surface, showing the M130 ROI. Center: Ventral view of the cortex, showing the M170 ROI. Right: Lateral view of the cortex showing the M350 ROI.



2.3.2. Outlier Rejection

Prior to the correlational analyses, outlier rejection was performed for each subject, based on the raw MEG data. Outlier trials were defined by first counting the number of measured data points in each trial that were not within two standard deviations of the overall mean of the data. This produced a vector of 'extreme value counts,' with one value per trial. Trials were then excluded from the correlational analyses if this 'extreme value count' fell more than three standard deviations above the mean — i.e., the trial had an 'extreme number of extreme values.' This procedure excluded a total of 1.7% of all MEG trials.

For the purpose of behavioral response time analyses, outlier rejection was also performed based on behavioral measures. Trials with incorrect responses (amounting to 12% of all trials) were excluded, as were trials in which the subject did not respond within 5 seconds of stimulus onset. Remaining response times were then normalized for each subject, and trials for which the response time was more than 3 standard deviations higher than the subject's mean were excluded as well. Trials excluded due to extreme response time amounted to an additional 1.7% of the trials. Error and extreme response trials were excluded only from analyses directly involving behavioral response time, and not from other MEG analyses.

3. Results

3.1. Behavioral Results

After both MEG and behavioral outlier exclusion, the mean reaction time (RT) across all subjects was 787ms. Correlations with reaction time were computed on RT data normalized within subjects. There was no significant difference in RT between the affixed words and the controls, or between the three categories of affixed words. Surface frequency was significantly correlated with RT, both within affixed words (across all three classes, $r = -0.31$, $p = 0$) and within matched controls ($r = -0.33$, $p = 0$). Surface frequency also had a significant effect on RT within each category individually ($r = -0.34$, $p = 0$ within free-stems; $r = -0.27$, $p = 0$ within bound-roots; $r = -0.33$, $p = 0$ within unique root words). Lemma frequency was also significantly correlated with RT, across both free-stems and bound-roots ($r = -0.20$, $p = 0$), and within each of

these categories ($r = -0.15$, $p < 0.005$ for free-stems; $r = -0.25$, $p = 0$ for bound-roots) (this measure being undefined for the unique root group).

3.2. M130 Results

Correlations were computed between both left and right hemisphere M130 activity, defined as the mean over a 50ms window centered at each subject's peak, and each continuous independent variable of interest. A significant effect of morphological affix frequency (hereafter, simply 'affix frequency') on left hemisphere M130 activity was found across all three categories of affixed words ($r = 0.07$, $p < 0.05$). Orthographic affix frequency (hereafter, 'last n-gram frequency') was also significantly correlated with the left hemisphere M130 across all three categories ($r = 0.07$, $p < 0.05$). When these effects were examined in each affix category individually, both affix frequency and last n-gram frequency were only significant in the class of bound-roots ($r = 0.12$, $p < 0.05$ for affix frequency; $r = 0.15$, $p < 0.01$ for last-n-gram frequency). Note that as M130 analyses were conducted on signed, negative activity, these positive correlations are indicative of higher affix frequency corresponding to *less* negative activation. The right hemisphere M130 was not significantly correlated with either measure of affix frequency.

3.3. M170 Results

Correlations were computed on M170 activity averaged over a 50ms window centered at each subject's peak, as well as on activity averaged over 20ms time-windows corresponding to the

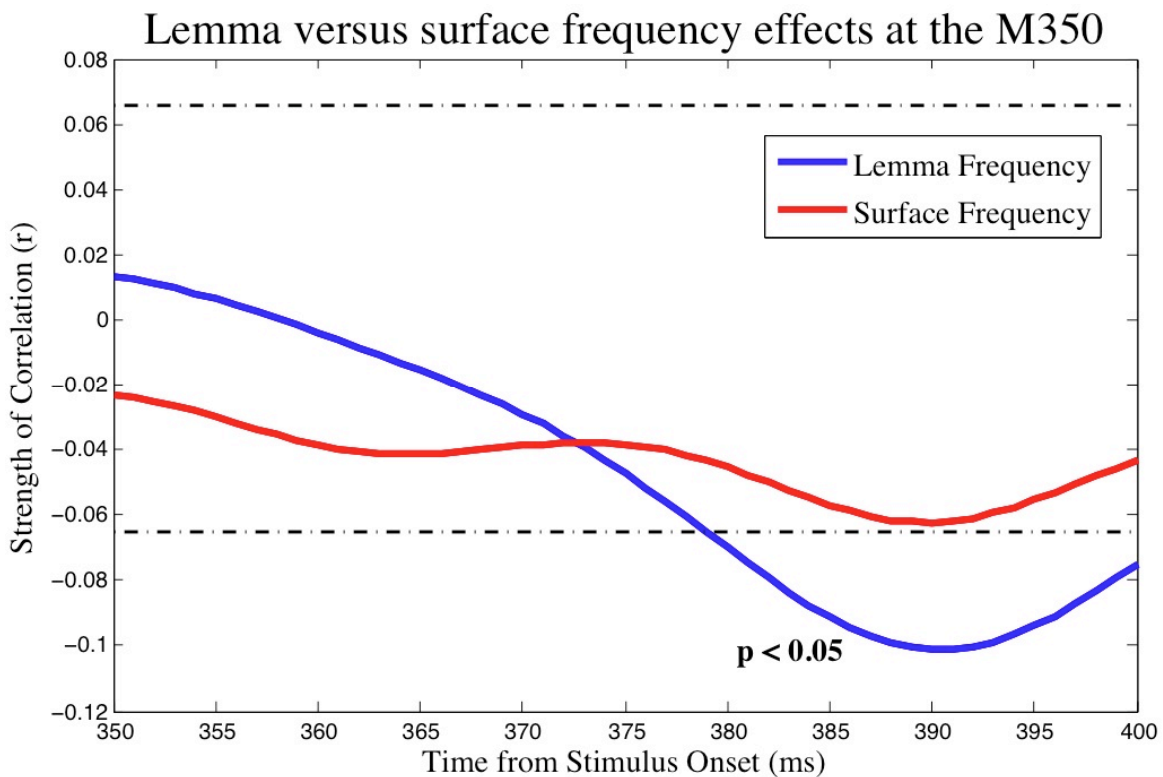
rise, height and fall of the peak. Affix frequency was found to have a significant effect on left hemisphere activity in the 20ms window centered at the M170 peak, across all three categories of affixed words ($r = -0.10, p < 0.005$). Last n-gram frequency, however, had no effect on this activity, for affixed or control words ($p > 0.3$ for both affixed and control). There was also a significant correlation between lemma transition probability (TPL) and left hemisphere M170 activity in the 20ms time window corresponding to the fall of the peak. This effect was significant across the free-stems and bound-roots ($r = 0.10, p < 0.05$). The TPL effect was also present in the free-stem group alone ($r = 0.16, p < 0.01$), but only neared significance in the bound-roots ($r = 0.10, p = 0.09$). There was no significant effect of bigram transition probability (TPB) on the fall of the M170 across these two groups, in either group individually, or in the controls ($p > 0.4$ across and within the two affixed groups; $p > 0.2$ for controls). No effects of affix frequency, last n-gram frequency, TPL or TPB were found in the right hemisphere, or in other left hemisphere time windows.

3.4. M350 Results

The M350 region of activation was examined by way of two-dimensional time-course analyses exploring possible effects of lemma and surface frequency. Analyses were conducted on activation in the 300-400ms time range. The M350 analysis revealed an effect of lemma frequency across both free-stems and bound-roots. The correlation with lemma frequency was significant throughout the 379-400ms range ($\sum r = 1.95$ for 22 time-points, $p < 0.05$ following correction for multiple comparisons). This effect was positive, in the sense that higher values of lemma frequency were associated with stronger negative activation in the region. No effect of

surface frequency was found for the affixed or control word groups. The effect of lemma frequency, contrasted with that of surface frequency, can be seen in Figure 3.

Figure 3: Effects of lemma and surface frequency on M350 activation, across both free-stems and bound-roots. Correlations are plotted over time, with the uncorrected $p = 0.05$ significance level indicated by the dotted lines. The cluster that surpasses this line retained significance at the $p = 0.05$ level following correction for multiple comparisons. As the M350 activation was primarily negative, the negative correlation with lemma frequency implies that higher frequency values were associated with *stronger* negative activation.



4. Discussion

The current study was designed to test for morphological decomposition in the visual recognition of three types of complex words: free-stems, bound-roots and unique root words. Three brain

response components — the M130, M170 and M350 — were isolated to test for the effects of morphological properties throughout the recognition process. For the earlier stages in processing, the words' morphological characteristics were contrasted with corresponding statistical properties of letter strings, in an attempt to distinguish between genuine morphological decomposition, in which morphemes are treated as qualitatively unique elements, and the detection of merely quantitative statistical regularities of letter strings.

We found the left hemisphere M130 to be sensitive to affix properties of the complex words. Both a morphological and an orthographic measure of affix frequency exerted significant effects on M130 activity, across all three groups of words. While this result does suggest that activity at the M130 is associated with the detection of affix properties, it appears that the relevant properties at this stage in processing may be orthographic and not morphological in nature. As there was no distinction between affix and last n-gram frequency at this stage, it is likely that these effects reflect the detection of statistical regularities rather than that of the words' morphology. This conclusion also makes sense in light of the localization of the M130 to a posterior region in the visual cortex, which typically responds to lower-level visual stimulus characteristics rather than abstract properties (Tarkiainen et al., 1999).

In contrast to our findings at the M130, we found what appear to be effects of genuine morphological decomposition at the slightly later stage of the M170. There was a significant effect of morphological affix frequency, but not last n-gram frequency, on the left hemisphere M170 across all three classes of words. Although these two measures are highly correlated with one another and both were found to exert roughly the same influence on the M130, their

divergence at the M170 was substantial — there was a significant negative correlation between M170 activity and morphological affix frequency, while the correlation with last n-gram frequency did not even near significance. The M170 was also significantly affected by TPL (the transition probability from the lemma to the affix), across both free-stems and bound-roots (the only two categories for which this variable has a non-trivial definition). The TPL measure is contrasted with TPB (bigram transition probability), which was not found to have an effect at the M170.

These findings suggest that activity at the M170 is concerned with genuine morphological decomposition, rather than the detection of statistical properties of letter strings that occurs at earlier stages in processing. Though the morphological characteristics and their corresponding orthographic measures are highly correlated, it appears that the former cannot be reduced to the latter — the morphological properties explain variations in brain activity that cannot be explained by orthographic regularities. This stage in word recognition has been previously associated with morphological processing by Zweig and Pykkänen (in press), who report an M170 effect of morphological complexity. The current findings are also informative with respect to the function of the occipito-temporal fusiform gyrus area, to which the M170 response component was localized. This area has been previously implicated in the recognition of visual word-forms (Dehaene, Le Clec, Poline, Le Bihan, & Cohen, 2002). Our results suggest that these word-forms may not be purely orthographic, but rather might assume a more abstract level of representation, at which morphemes are recognized and parsed as unique items.

Evidence for morphological decomposition of the complex words was also present at later stages in processing. There was a significant effect of lemma frequency on activation in the superior temporal network in the 379-400ms time range (across both free-stems and bound-roots), while surface frequency did not have this effect. The effect of lemma frequency is not completely straightforward, as its direction is opposite to that which would be expected from previous N400 findings — N400 amplitude is typically reduced, rather than heightened, in response to higher frequency words (Marinkovic, 2004). The current finding also cannot be compared directly to previous literature on the M350, as M350 effects have primarily involved peak latency rather than amplitude (Pylkkänen & Marantz, 2003). Nevertheless, the relevance of lemma frequency at this stage in processing suggests that stem identification does play a key role in lexical access. Overall, this finding further supports a decompositional model of word recognition, which entails that (at least some) words are accessed via their morphological constituents.

While it is clear that our results support a model of complex word recognition that involves some sort of morphological decomposition, it is difficult to distinguish between a full decomposition model, which holds that complex words are always decomposed, and a dual-route model, in which decomposition is only utilized for some complex words but not others. The latter model claims that while decomposition must always be attempted, in some cases it may not be the winning route to recognition. The relevance of the morphological characteristics at the M170 stage in processing suggests that the brain is attempting to parse the presented word, but does not necessarily imply that the parsing is successful and will lead to decompositional, rather than whole word, lexical access. Our results at the M350 do suggest that (at least) some words

are accessed via their parts, given the relevance of lemma frequency rather than surface frequency at this stage in processing. Nevertheless, further research is necessary to shed light on the distinction between the dual-route and full decomposition models of recognition.

Our general findings at the M130 and M170 stages of processing suggest that all three classes of complex words undergo the same attempts at decomposition. Early effects of affix frequency were present across all three classes, while TPL and lemma frequency effects were present across the two more obviously affixed word groups (both measures being undefined for the unique root group). In many cases, the cross-category effects were not found in some or all of the groups individually, but this is likely due to the loss in statistical power from the limited number of stimuli in each group, especially given that there was no distinguishable pattern in the differences in significant effects between categories. Overall, it appears that parsing attempts are aimed at all three classes of complex words in early stages of processing.

Results at the M350 stage also do not point to any differences between the free-stem and bound-root classes. The lemma frequency effect was present across these two categories, which suggests that some words of each class are in fact accessed via their decomposition. As lemma frequency was indistinguishable from surface frequency for unique root words, our results do not provide enough evidence to determine whether or not these words are ever accessed via their parts. While it appears that parsing of the unique root words is attempted at earlier stages in processing, it is unclear whether or not these words are ultimately decomposed as are those of the genuinely affixed classes.

It is worth noting that the correlational analysis technique employed in the current study is particularly advantageous in the investigation of linguistic variables such as those examined here. Traditional experimental designs, in which stimuli are binned along certain dimension of interest, assume that stimuli are sampled from a population that is balanced along these variables. In the study of language, however, this assumption is often violated — the words of a language are naturally imbalanced over many linguistic variables, and variables of interest for particular research questions are often highly correlated with one another. This makes it difficult, if not impossible, to create unbiased bins of stimuli for a traditional analysis. This was the case for many of the factors relevant to the current study. For example, an attempt to bin the target words by the two factors of morphological and orthographic affix frequency would have been nearly impossible. As the two variables are highly correlated, filling a bin with words that were high in one measure and low in the other would have required us to sample from an extremely limited, and likely strange, set of items. Furthermore, as the words were not evenly distributed along the affix frequency measures themselves, they did not naturally lend themselves to be divided into distinct bins of *high* and *low* frequency. Artificial grouping of stimuli in cases such as these may violate the assumptions behind traditional statistical designs (Baayen, 2004). The correlational analysis method employed in the current study allows for the exploration of continuous variables of interest, without the need for such artificial grouping. This technique may be useful in the investigation of a variety of issues in psycho- and neurolinguistics.

5. Conclusion

Our findings throughout the stages of word recognition provide evidence that morphological decomposition is, at the very least, attempted, in the processing of all three classes of complex words: free-stems, bound-roots, and unique root words. We suggest that while the early visual M130 response reflects the detection of orthographic statistical properties of affixed words, the later M170 response reflects genuine morphological parsing of the affix from the stem. At this later stage, the brain makes use of properties such as affix frequency and transitional probability in its attempt to decompose a word into its morphological constituents. The word forms that are recognized through the decomposition process then activate lexical entries for lemmas at the M350 stage, at which lemma, rather than surface frequency is the relevant factor. While more evidence is needed to establish which particular words are ultimately accessed through the decompositional route, or if there are differences between various classes of derived words, our results clearly point to a model of word recognition in which parsing is attempted at the early stage for all complex words, after which at least some are accessed via their morphological constituents.

References

- Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & Ochiai, M. (2001). Reduction of Nonperiodical Extramural Magnetic Noise in MEG measurement by Continuously Adjusted Least Squares Method, *IEEE Transactions on Applied Superconductivity*, Proceedings of 2000 Applied Superconductivity Conference. Virginia Beach, Virginia.
- Baayen, R. H. (2004). Statistics in Psycholinguistics: A critique of some current gold standards. *The Mental Lexicon Working Papers I*.
- Baayen, R. H., Piepenbrock, R. & Gulikers, L. (1995). The CELEX Lexical Database (Release 2) [CD ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Baayen, R. H., Wurm, L. H., & Aycok, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, 2(3), 419-463.
- 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://ellexicon.wustl.edu/>, Washington University.
- Dale, A. M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., & Halgren, E. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26(1), 55-67.

- Dehaene, S., Le Clec, H.G., Poline, J.B., Le Bihan, D., & Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, *13*(3), 321-325.
- Embick, D., & Marantz, A. (2005). Cognitive Neuroscience and the English Past Tense: Comments on the Paper by Ullman et al. *Brain and Language*, *93*, 243-247.
- Hauk, O., Davis, M. H., Ford, M., Pulvermuller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, *30*(4), 1383-1400.
- Hay, J. (2001). Lexical frequency in morphology: Is everything relative. *Linguistics*, *39*(6), 1041-1070.
- Hay, J., & Baayen, H. (2001). Parsing and Productivity. In G. Booij, Jaap van Marle (Ed.), *Yearbook of Morphology: 2001* (pp. 203-236): Springer.
- Hay, J., & Baayen, H. (2003). Phonotactics, parsing and productivity. *Italian Journal of Linguistics*, *1*, 99–130.
- Hay, J., & Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, *9*(7), 342-348.
- Lehtonen, M., Cunillera, T., Rodríguez-Fornells, A., Hultén, A., Tuomainen, J., & Laine, M. (2007). Recognition of morphologically complex words in Finnish: Evidence from event-related potentials. *Brain Research*, *1148*, 123-137.
- Lehtonen, M., Vorobyev, V. A., Hugdahl, K., Tuokkola, T., & Laine, M. (2006). Neural correlates of morphological decomposition in a morphologically rich language: An fMRI study. *Brain and Language*, *98*(2), 182-193.

- Marinkovic, K. (2004). Spatiotemporal Dynamics of Word Processing in the Human Cortex. *Neuroscientist, 10*(2), 142-152.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*, 177-190.
- Marslen-Wilson, W. D., Bozic, M., & Randall, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes, 23*(3), 394-421.
- McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7*(7), 293-299.
- Pinker, S. (2000). *Words and Rules: the ingredients of language*. New York: Harper Perennial.
- Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*(11), 456-463.
- Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences, 7*(5), 187-189.
- Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language, 81*(1-3), 666-678.
- Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*.
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review, 11*(6), 1090-1098.

- Solomyak, O., & Marantz, A. (in press). Lexical access in early stages of visual word processing: A single-trial correlational MEG study of heteronym recognition. *Brain and Language*.
- Stockall, L., & Marantz, A. (2008). A single route, full decomposition model of morphological complexity. *The Mental Lexicon*, 1(1), 85-123.
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology, Section A*, 57(4), 745-765.
- Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 638–647.
- Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, 122 (Pt 11), 2119-2132.
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92(1-2), 231-270.
- Zweig, E., & Pykkänen, L. (in press). A visual M170 effect of morphological complexity. *Language and Cognitive Processes*.

Appendix A: Word-Choice Algorithm

The target words were selected to be well and equally distributed over length, bigram frequency and surface frequency for all three word categories, along with affix dominance for the free-stemmed words. The word-choice routine first normalized each variable, with the exception of length. Frequency and bigram frequency were normalized across all words in the potential word list, while affix-dominance was normalized across all potential words of the first category. For each variable, words were then binned into six groups based on their normalized z-scores. Length, the only variable that was not normalized as described above, was binned into five groups corresponding to its five possible values within the range of 7-11. Each variable's bins were then given weights that determined their desired proportions in the selected word list. All six bins were given equal weights for frequency, bigram frequency and affix-dominance, in order to create a slightly flattened normal distribution over these variables. For the variable of length, shorter words (7-9 letters) were given slightly stronger weights than longer words (10-11 letters), as very few of the longer words were available for each suffix and category. Bins and weight distributions can be found in Tables A1 and A2.

Table A1: Bins for desired distributions over bigram frequency, surface frequency and affix dominance

Normalized Z-Score	Desired Weight
$-0.33 < z \leq 0.33$	1
$-1 < z \leq -0.33$	1
$0.33 < z \leq 1$	1
$1 < z \leq 2$	1
$-2 < z \leq -1$	1
$2 < z$ OR $z \leq -2$	1

Table A2: Bins for desired distribution over length

Length in Letters	Desired Weight
7	1.5
8	1.5
9	1.5
10	1
11	0.5

The word-choice routine selected words for each suffix and category independently. The program first made a random selection of six words from each potential list. It then calculated the difference between the number of words from each bin in the selected list and the weight of each bin in the desired distribution. The sum of the squares of these error terms for all of the relevant variables provided a measure of how close the distribution of the selected words was to

the desired distribution. The program then replaced a random number of the selected words with other, randomly chosen, words from the potential list. The new list's error value was then calculated and compared to that of the last list. The word list with the lower error value was kept as the selected list. The program cycled through one million iterations of this routine, after which the last selected list was chosen to be included in the experiment.

For the selection of the control words, the routine worked as it did for the target words, the only difference being that the words were binned based on their values relative to the mean and standard deviation of the chosen affixed words. The weight values for the bins were also based on the actual distributions of the target words, rather than the original desired distributions. The word-choice routine selected lists for each of the nine suffixes independently, but this was done in order of least to greatest number of potential words to choose from. As some words were potential matches for more than one suffix, all words that were chosen as matches for a particular suffix were subsequently excluded from the remaining lists. For the selection of the non-words, potential non-words were binned based on their values relative to the entire group of selected words. The weights of the bins for each variable were again determined by the actual distribution of the selected words. The routine was used to select the entire list of non-words at once.