

Evidence for Early Morphological Decomposition in Visual Word Recognition

Olla Solomyak and Alec Marantz

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 100–200 msec poststimulus onset in the previously identified letter string and visual word-form areas. MEG data were analyzed using cortically constrained minimum-norm estimation. Correlations were computed between activity at functionally defined ROIs 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 evi-

dence of decomposition for both free stems and bound roots 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 (~130 msec). We find that effects of decomposition at the M170 can, in fact, be attributed to morphological properties of complex words, rather than to 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 containing bound roots as well as free word-stems. ■

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 (2000) "words and rules" model and the issue of single- versus 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; Ullman, 2004; Pinker & Ullman, 2002). In single-route models (e.g., Butterworth, 1983), all morphologically complex words are recognized via the same mechanism, either as whole words for models in which there is no decomposition prior to lexical access, or via their parts for models in which obligatory decomposition precedes lexical access. In dual-route models (e.g., Schreuder & Baayen, 1997), either some words are recognized via decomposition whereas

others are recognized via their parts, or all words are subject to parallel whole word and decompositional routes, with success in lexical access coming from one or the other route depending on the word and other factors. Although the psycholinguistic and computational modeling literature (Marslen-Wilson, Bozic, & Randall, 2008; Hay & Baayen, 2005; Taft & Forster, 1975) has examined sophisticated variations on these models of lexical access that attempt to come to terms with a wide range of conflicting data, the neurolinguistic literature has tended to adopt a more simplified view. Here frequency effects on behavior (higher-frequency items yield faster responses) are associated with the declarative (associative) memory system and complexity effects (structurally more complex items yield slower responses) 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 a decomposition route, with 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).

The logic of equating surface frequency effects on RT without base frequency effects with a whole word access route depends on assumptions about the relationship between neural computation and RT, on the one hand, and the relative importance of item frequency as opposed to conditional probabilities in the processing of complex stimuli on the other. On the first assumption, one might interpret as support for the hypothesis that lexical access does not involve decomposition the fact that surface frequency accounts for most of the variance in lexical access RT for some morphologically complex words but the variable of being complex versus being morphologically simple does not. However, 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 as decomposition effects are evident for morphologically complex words for which complexity itself does not directly modulate RT (see, e.g., Zweig & Pykkänen, 2008, and below). MEG techniques allow us to follow (putative) stages of visual word recognition in the brain to evaluate models of complex word recognition, providing the means to uncover stimulus modulation of neural activity that does not correlate directly with RT in the experimental situation. 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 a speed up in brain responses correlates with a slow down in RT, reflecting processing costs at later stages (Pykkänen, Stringfellow, & Marantz, 2002). On the second assumption about frequency effects, we find support in this experiment for the proposal that what counts most for modulating the neural computations associated with recognizing morphologically complex words is not the absolute frequency of the stem of such words but rather the conditional probability of the affixed word given the stem.

This study provides additional evidence from MEG for obligatory morphological decomposition of derived words prior to lexical access in visual word recognition, independent of any effects of decomposition on behavioral 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). Strong evidence from masked priming and MEG for decomposition in English has so far been limited to cases of affixation to a word, although evidence from Hebrew and Arabic suggests decomposition to the (bound) roots in these languages (Boudelaa & Marslen-Wilson, 2001; Frost, Forster, & Deutsch, 1997). Taft and Forster (1975) followed the linguistic literature in proposing decomposition for words containing bound as well as free stems. The inclusion of unique stem words allows us to explore the possible limits of decomposition. Does decomposition depend on the identification of a stem that occurs in more than one word, or can the internal properties of a single word

themselves motivate decomposition? For example, finding the *-able* in unique stem *tenable* predicts syntactic and semantic properties of the word—it is an adjective describing a dispositional property—and predicts that the nominal form of the word could be *tenability* (as opposed to *tenableness*; see Embick & Marantz, 2008). 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 or her structural MRI data. This approach allows us to evaluate the effect of two continuous variables implicated in decomposition and compare these to strictly orthographic but related (and correlated) variables. The morphological variables are the transition probability between the stem and the affix (the conditional probability of encountering the whole word given the stem) and affix frequency. The corresponding orthographic (nonmorphological) variables are the transition probability between the last letters of the stem and the first letters of the affix and the positional letter frequency of the affix (the frequency of the sequence of letters making up the affix at the end of words).

Taft and Forster (1975) published the most influential work on morphological decomposition in word recognition starting in the 1970s. Their obligatory decomposition 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 impact of Taft and Forster's experiments, the behavioral literature since the 1970s paints a mixed picture of morphological decomposition, sometimes supporting and sometimes contradicting predictions drawn from the obligatory decomposition model. In conscious priming experiments, the apparent lack of significant behavioral priming in cross-modal presentation for pairs like "darkness" and "darkly" (Marslen-Wilson & Zhou, 1999), 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 at a stage between lexical activation and, for example, lexical decision (see the discussion in Stockall & Marantz, 2008). Although any morphologically complex word should activate the lexical entries of all words containing a shared stem, a behavioral task such as lexical decision involves a processing stage beyond lexical activation at which some choice or decision is made, and competition among activated lexical entries might slow down this stage. In languages such as English, there seems to be no simple effect of morphological complexity on RT in naming or lexical decision, whereas in languages such as Finnish, in some cases there is a robust inhibitory effect of complexity associated with inflectional morphology, that is, a slower response to inflected words

than to mono-morphemic control words matched on length and surface frequency (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 stages.

The evoked response literature has elucidated the early stages of visual word recognition, with MEG studies providing spatial and temporal information about early processing consistent with the results from ERP and fMRI. Although there is much agreement in the literature on the spreading brain activation stimulated by visually presented words (the results of the present experiment are generally consistent with the description in Marinkovic, 2004), 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 150 msec 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 170 msec poststimulus onset, the M170 (the MEG counterpart to the N170 evoked by visually presented objects) from 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 and Marantz (2009), 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 200 msec of visual word presentation, after the activity from posterior regions described above is well underway, activation also begins in inferior frontal regions, as well as middle and superior temporal regions. The MEG literature has suggested that the M350 from these middle and superior temporal regions 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 contributes to the N400 of the ERP literature, which may begin around 200 msec poststimulus onset and peak around 400 msec.

Although the evoked response literature provides a guide to the neural activity examined in the present study, the current experiment is not designed around testing hypotheses about evoked response components such as the M170. As described in the Methods section, we defined the brain regions and time windows of interest in this experiment according to the averaged activity across trials and subjects, looking for regions and windows in which brain activity evoked by the visually presented letter strings was significantly above the prestimulus baseline. In particular, our regions and time windows are determined in source space and not in terms of peaks in the data from the MEG sensors, as are the standard evoked response components. We use these regions and windows to test hypotheses about the neural computations behind complex word recognition derived from the linguistic and psycholinguistic literature.

The majority of psycholinguistic data support a version of either a full decomposition model of lexical access or a parallel dual-route model in which decomposition is attempted for each word prior to lexical access. We use the MEG data to test a particular claim about the statistical variables relevant to decomposition from the important work of Hay and Baayen (2001, 2003, 2005) and Hay (2001) on morphological parsing. Hay claims that it is not the absolute frequency of the stem that primarily affects morphological parsing, but rather the statistical relationships between stem and affix. The importance of these statistical variables for processing 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. 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. Rather, we look for evidence for morphological decomposition in the correlation between brain activity (at our regions and time windows of interest) and statistical properties of words hypothesized to be relevant for decomposition.

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 the lexical entries of the primes (see, e.g., Rastle et al., 2004).

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. Thus, “brother” primes the morphologically unrelated “broth,” whereas “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 Pykkänen (2008) 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. The masked priming experiments use primes that exhaustively parse into affixes plus stems that are existing words, and Zweig and Pykkänen found their effect by contrasting words that consist of a word and a suffix, such as “farmer,” with words that consist of a nonword and a suffix, such as “winter.” Thus, although these studies support a notion of decomposition independent of lexical entries (the decomposition of “brother” into “broth” and “er” is insensitive to lexical information associated with “broth” or “brother”), they implicate a mechanism that involves recognition of the visual word forms of morphemes, one that depends on the status of both stem and suffix, not just the suffix.

A recent ERP study of masked priming (Morris, Frank, Grainger, & Holcomb, 2007) questions whether the stage of processing associated with the influence of the prime on the target really involves stripping of an affix from a stem independent of the lexical entry for the morphologically complex prime. The authors claim that semantic relatedness between prime and target for the unrelated (“opaque”) items (e.g., between “whisker” and “whisk”) might account for the contrasts among their stimulus groups, undermining the lexicon-free affix-stripping account. However, their results conflict with those of Lavric, Clapp, and Rastle (2007) in a similarly designed experiment; the latter study found the same magnitude effect across “opaque” and “transparent” (morphologically related) classes. More importantly, although Morris et al. controlled for prime (complex word) and target (stem) frequency across their sets of opaque and transparent stimuli, they did not control for the statistical relation between complex word and stem frequency, which correlates with the transition probability between stem and suffix in the present experiment. Investigation of their materials reveals a significant difference between their opaque and transparent stimuli on this measure. We show below, in agreement with the work of Hay and Baayen, that the relative frequency of stem and complex word is a crucial factor in decomposition. For this reason, the Morris et al. result cannot be interpreted to support a semantic influence on early decomposition (the authors, in fact,

suggest that their results are consistent with an interpretation in which semantic factors influence the visual word-form lexicon, without exerting a direct effect on on-line decomposition).

Although Zweig and Pykkänen show that brain activity at 170 msec poststimulus onset is modulated by morphological complexity, with complex words such as “farmer” yielding a larger M170 response than simple words such as “winter,” their complexity variable did not affect RT. The behavioral result is consistent with Taft’s original full decomposition model, where decomposition (affix stripping) was essentially costless, and recent work on English and Finnish suggests that any 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., 2006, 2007; Taft, 2004) rather than to affix stripping. In the present experiment, we test for the neurological correlates of decomposition not by comparing morphologically complex with morphologically simple words but by correlating brain responses to morphologically complex words with variables associated with the ease or difficulty of decomposition within this class of words.

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 (see, e.g., Grainger & Giraudo, 2000). 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 such as *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. Although we will show evidence below that proposed early 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 feedforward processing and cascading interactive levels of computation. In addition, this study concentrates on the proposed stages of decomposition and stem access, that is, the “early” stages of complex word recognition up to and including the MEG M350 response component.

The present study represents an advance on the literature on morphological decomposition in several respects. First, we sample equally from a wide range of English derivational suffixes rather than from a small set that might exhibit idiosyncratic properties. Second, we extend the

range of morphologically complex words from those containing free stems, used in recent masked priming and MEG studies on English, to bound roots. Recent linguistic work on morphology, particularly in the framework of Distributed Morphology (see Embick & Marantz, 2008 for discussion relevant to the words in this experiment), has argued for the uniformity between affixation to bound and (many) free stems, cutting up categories of derived words along different dimensions. Third, we explore the effect of a statistical relation between stem and suffix, the transition probability, highlighted by recent work in morphology but not investigated in the masked priming and evoked response literature (and, as mentioned above, not controlled for in Morris et al., 2007). Finally, we further develop the single-trial correlational approach to MEG source activation analyses that we introduced in Solomyak and Marantz (2009) as particularly appropriate for neuro-linguistic investigations.

METHODS

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, that is, 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 an 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). Potential stimuli were limited to 7–11 letter words, for which the mean accuracy rate (as reported by ELP) was not below 30%. To find potential target words, all 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), were first extracted from the ELP. Words were then classified into the three groups of interest based on their morphological parsing, as defined by the ELP.¹ An automatic word-choice algorithm was then employed, which randomly chose words

from each group such that they were well and equally distributed over length, mean bigram frequency, and log surface frequency. The free stems were also selected to be well distributed over affix dominance, which was defined as the difference between each word’s surface frequency and the surface frequency of its stem. 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 for each suffix, generating 162 target words (6 words × 3 categories × 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. Words were matched for frequency of last bigram, trigram, or tetragram, depending on the length of the relevant suffix. Suffix and ending frequencies were calculated using the CELEX database (Baayen, Piepenbrock, & Gulikers, 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 nonwords was then chosen to match the words on their overall distributions over length and mean bigram frequency. In addition to the 324 words and 324 nonwords that were relevant to the current study, the experiment also contained 78 filler stimuli from another study, all of which were mono-morphemes.² A complete list of the words used in the current study can be found in Appendix B. A list of summary statistics for the words is provided in Appendix C.

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.³ The CELEX database was then searched for all words that were defined to have these family members as their morphological constituents. These

Table 1. Orthographic and Morphological Affix Frequency Counts, Presented in Units of Natural Log Frequency per Million

	<i>Orthographic Affix Frequency</i>	<i>Morphological Affix Frequency</i>
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

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.⁴

Frequency data were obtained for all family members. Hyperspace analogue to language (HAL) frequencies were obtained from the ELP for those words that were included in the ELP database.⁵ 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.

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 orthographic properties related to affixation. The first measure was the same as was used for the matching of the control words—the fre-

quency 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, for example, the word “predictability” was counted toward 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.

To test for decomposition of the affixed words at early stages of brain processing, the effects of the morphological variables TPL and affix frequency were compared to those of their orthographic counterparts, that is, TPB and orthographic affix frequency. To explore the role of morphology at later stages in processing, the effects of lemma frequency were compared to those of surface frequency. Table 2 shows the correlation coefficients between the variables of interest for comparison.

Experimental Procedures

Participants were nine right-handed native English speakers, ranging in age from 19 to 29 years, with a mean age of 23.3 years. 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 participants’ approval for the study was obtained from the Institutional Review Boards 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 500 msec, after which the stimulus was presented for 300 msec.⁶ 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 head position indicator coils were used to localize the position

Table 2. Correlation Coefficients between Key Variables to Be Compared in Correlational Analyses, Separated by Word Category

	<i>Lemma × Surface Frequency</i>	<i>TPL × TPB</i>	<i>Morphological × Orthographic Affix Frequency</i>
Free stems	.52	.32	.81
Bound roots	.74	.28	.81
Unique roots	N/A	N/A	.81

of the subject's head inside the MEG helmet. The head-shape digitization and head position indicator locations were later used to match the MEG coordinate system to that of each subject's MRI.

Analysis Methods

Minimum-norm Estimation

The MEG data processing and minimum-norm estimation procedures were the same as those previously described in Solomyak and Marantz (2009). MEG data were noise-reduced using the Continuously Adjusted Least-Squares Method (Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 2001), in the MEG160 software (Yokogawa Electric Corporation and Eagle Technology Corporation, Tokyo, Japan). Each subject's MEG data were then averaged across all trials. Average data were low-pass filtered at 40 Hz and baseline corrected using 100 msec of prestimulus data. 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 MRIs were processed in FreeSurfer (CorTechs Labs, La Jolla, 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 subject's cortical surface. The 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. 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 (ROIs) for correlational analysis, the 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. Only components of activation that were in the direction normal to the cortical surface were retained in the minimum-norm solution. The resulting minimum-norm estimates were converted into noise-normalized dSPM values for use in subsequent correlational analyses.

Region-of-interest analyses. The averaged dSPMs exhibited a clear pattern of early activation, which consisted of an earlier (140–150 msec) negative (i.e., current flowing inward) peak in a posterior occipital region, and a slightly later (180–190 msec) positive (i.e., current flowing outward) peak in the occipito-temporal fusiform gyrus area. The earlier activation was defined as the M130 and the later activation as the M170. ROIs were created to isolate these regions of activity in the left and right hemispheres, both of which displayed the pattern of peaks described above. The averaged dSPMs also displayed a network of later activation, located in the left hemisphere superior temporal and sylvian fissure regions. This area showed heightened negative activation beginning at about 250 msec and lasting throughout the 300–400 msec time range. This activation was identified as the M350 component in processing. An additional ROI was created to isolate this region of activity. The left hemisphere activation in all three regions identified for analysis is illustrated in Figure 1, whereas the ROIs isolating 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 ROI were excluded from analysis of this region. One subject did not exemplify the typical pattern of M130 activity for the left hemisphere, and another for the right hemisphere: These subjects lacked a negative peak in the 100–190 msec time range in the isolated region. Two subjects did not show the typical

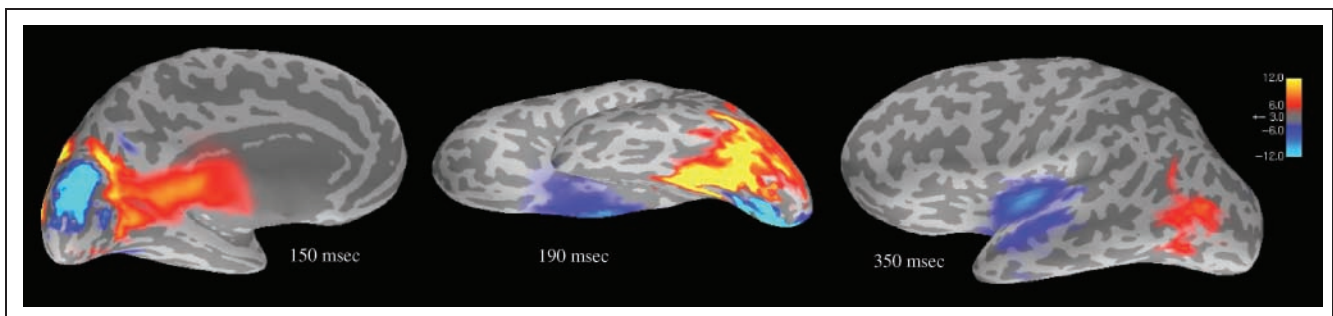


Figure 1. Average left hemisphere dSPM activation across all subjects at three time points of interest, shown on a representative subject's inflated cortical surface. Positive (red) dSPM values indicate current flowing outward from the cortical surface, whereas negative (blue) values indicate current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

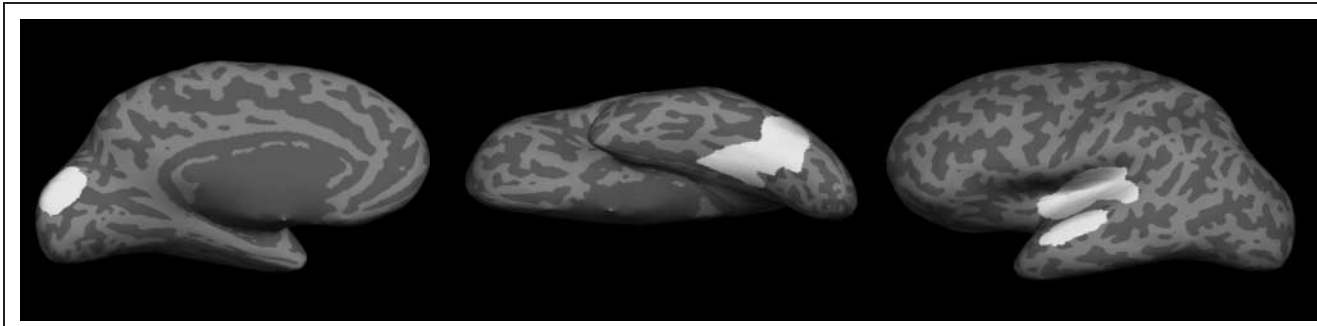


Figure 2. Left hemisphere ROIs identified for correlational analysis, defined based on heightened activation in the isolated regions. Left: Medial view of the inflated cortical surface, showing the M130 ROI. Center: Ventral view of the cortex, showing the M170 ROI. Right: Lateral view of the cortex showing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

pattern of M170 activity for each hemisphere, as they lacked a positive peak in the 150–215 msec time range in the relevant ROI. These were not the same two subjects for both hemispheres—one subject was atypical in both hemispheres, whereas two others were atypical in one hemisphere each. 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 146.1 msec (with a range of 122–172 msec) in the left hemisphere and 160.4 msec (ranging from 142–177 msec) in the right. The mean M170 latency was 190.9 msec (ranging from 178–214 msec) in the left hemisphere and 185.4 msec (with a range of 164–204 msec) in the right.

As the directionality of early MEG components has been fairly consistent, both in the current study and in our previous experience (Pylkkänen & 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 50-msec window centered at each subject’s peak latency. Analyses were also conducted to examine the rise, height, and fall of each peak. Three 20-msec time windows were identified for this purpose: a “rise” time window spanning 30 to 10 msec prior to the peak, a “height” time window centered at the peak latency, and a “fall” time window spanning 10 through 30 msec 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 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.

M350 analyses were conducted on signed activation in the isolated 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 corre-

lations were computed to track the effects of interest over time. For this purpose, the average activation at each time point was first normalized within subjects, after which all subjects’ data were concatenated into a single matrix (time point by trial) of activation values. Correlations were then computed time point by time point across all subjects’ normalized values, creating a correlation wave tracking the effects of each independent variable of interest over time.

The resulting correlation waves were then 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 (2009). The procedure consisted of first defining a new statistic, Σr , as the sum of all correlation coefficients within a temporal cluster of consecutive significant effects in the same direction (at the raw $p = .05$ significance level). A Monte Carlo p value was then computed, which indicated the statistical significance of the Σr statistic for each temporal cluster, given the multiple comparisons that gave rise to the effect. This p value was computed as follows: The correlation wave was computed over time, as it had been in the original analysis, for 10,000 random permutations of the independent variable of interest. For each randomly produced wave, the Σr statistic was computed for each of its temporal clusters, after which the highest absolute value of Σr was taken as the statistic for that permutation. This produced a distribution of 10,000 Σr values, to which the original value could be compared. The percentage of values higher in absolute value than the original statistic was taken to be the Monte Carlo p value for the original cluster.

Outlier Rejection

Prior to the correlational analyses, outlier rejection was performed for each subject on the basis of 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, that is, 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 sec 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. Error trials were not excluded from MEG data analyses due to the strong correlation between error rate and surface frequency ($r = .69, p < .001$). Excluding these trials would have made it impossible to maintain the wide range of frequency values that is required for the correlational analysis technique.

RESULTS

Behavioral Results

After both MEG and behavioral outlier exclusion, the mean RT across all subjects was 787 msec. Correlations with RT 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, $r = -.31, p < .001$) and within matched controls ($r = -.33, p < .001$). Surface frequency also had a significant effect on RT within each category individually ($r = -.34, p < .001$ within free stems; $r = -.27, p < .001$ within bound roots; $r = -.33, p < .001$ within unique root words). Lemma frequency was also significantly correlated with RT, across both free stems and bound roots ($r = -.20, p < .001$), and within each of these categories ($r = -.15, p < .005$ for free stems; $r = -.25, p < .001$ for bound roots), this measure being undefined for the unique root group.

M130 Results

Correlations were computed between both left and right hemisphere M130 activity and the continuous independent variables of interest. Correlations were computed with activation averaged over a 50-msec window centered at each subject’s peak, as well as with activation averaged over 20-msec time windows corresponding to the rise, height, and fall of the peak. A significant effect of morphological affix frequency was found on left hemisphere M130 activity in the 50-msec time window, across all three categories of affixed words ($r = .07, p < .05$). Orthographic

affix frequency was also significantly correlated with this activation, across all three categories ($r = .07, p < .05$). Effects of both morphological and orthographic affix frequency were also found in the time window corresponding to the fall of the M130 peak, across all three categories of words ($r = .08, p < .05$ for morphological affix frequency; $r = .09, p < .01$ for orthographic). When each affix category was examined individually, the bound roots showed a significant effect of both morphological and orthographic affix frequency in the 50-msec M130 time window ($r = .12, p < .05$ for morphological; $r = .15, p < .01$ for orthographic). The free stems exhibited effects only in the time window corresponding to the fall of the M130—morphological affix frequency showed a significant effect on this activation ($r = .12, p < .05$) and the effect of orthographic affix frequency neared significance ($r = .11, p = .07$). No effects of either measure of affix frequency were found within the unique root words alone. The right hemisphere M130 was not significantly correlated with either measure of affix frequency for any of the three word categories. It is important to note that as M130 analyses were conducted on signed, negative activity, the positive correlations reported here are indicative of negative effects, that is, higher affix frequency corresponding to *less* negative activation.

M170 Results

Correlations were computed on left and right hemisphere M170 activation averaged over a 50-msec window centered at each subject’s peak, and on activation averaged over 20-msec time windows corresponding to the rise, height, and fall of the peak. Morphological affix frequency was found to have a significant effect on left hemisphere activity in the 20-msec window centered at the M170 peak, across all three categories of affixed words ($r = -.10, p < .005$). Orthographic affix frequency, however, had no effect on this activity, for affixed or control words ($p > .3$ for both affixed and control). There was also a significant correlation between TPL and left hemisphere M170 activity in the 20-msec time window corresponding to the fall of the peak. This effect was significant across the free stems and bound roots ($r = .10, p < .05$). The TPL effect was also present in the free-stem group alone ($r = .16, p < .01$), but only neared significance in the bound roots ($r = .10, p = .09$). There was no significant effect of TPB on the fall of the M170 across these two groups or in either group individually ($p > .4$ across and within the two affixed groups). No effects of morphological or orthographic affix frequency, TPL or TPB, were found in the right hemisphere, or in other left hemisphere time windows.

To further isolate the effects of morphology from those of orthography, each orthographic measure was partialled out from the corresponding morphological measure; that is, TPB was partialled out from TPL and orthographic from morphological affix frequency. The effects of the residuals of these regressions—the parts of the morphological

measures unaccounted for by orthographic properties—were then also tested for their effects on the M170 activation. The effect of TPL on the fall of the M170 retained significance even after TPB was partialled out ($r = .10, p < .05$, across the free stems and bound roots). However, the effect of morphological affix frequency on the peak of the M170 was no longer significant after orthographic affix frequency had been partialled out ($p > .1$).

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 300–400 msec 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 379–400 msec range ($\Sigma r = 1.95$ for 22 time points, $p < .05$ 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.

DISCUSSION

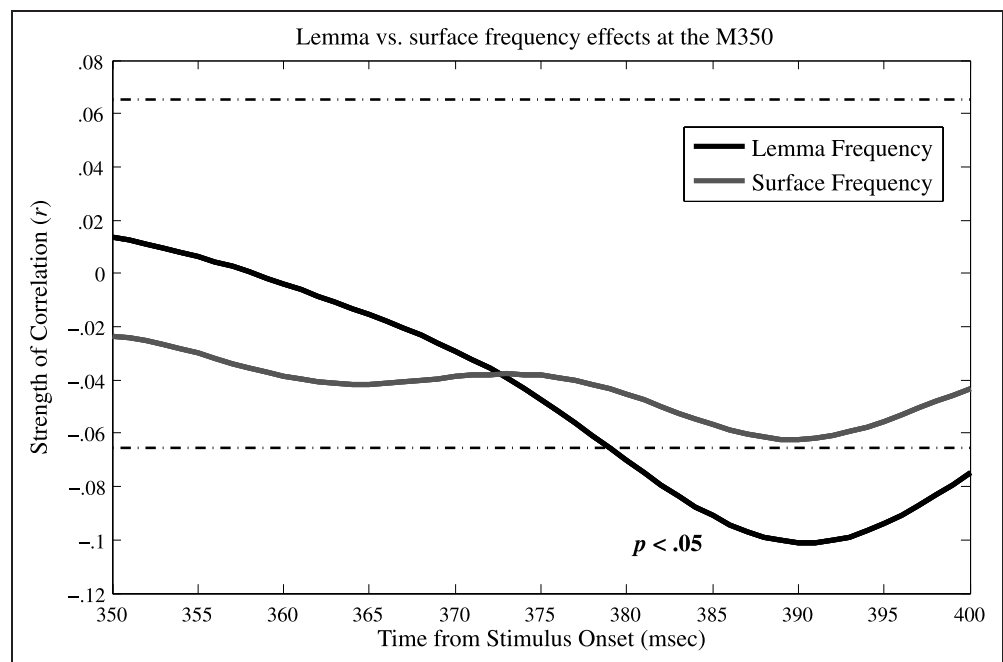
The current study was designed to test for morphological decomposition in the visual recognition of three types of complex words: free stems, bound roots, and unique root

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. Although 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 morphological and orthographic affix 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 visual cortex, which typically responds to lower-level visual stimulus characteristics rather than to 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. Left hemisphere M170 activation was significantly correlated with TPL (the transition probability from the lemma to the affix), across both free stems and bound roots (the only two

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 $p = .05$ significance level indicated by the dotted lines. The cluster that surpasses this line retained significance at the $p = .05$ 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.



categories for which this variable has a nontrivial definition). The TPL measure is contrasted with TPB, which was not found to have any effect at the M170. Furthermore, the effect of TPL on M170 activity retained its significance even after the TPB measure had been partialled out. There was also a significant effect of morphological affix frequency, but not orthographic affix frequency, on the 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, whereas the correlation with orthographic affix frequency did not even near significance. Despite this divergence, the morphological affix frequency effect was no longer significant after the orthographic measure was partialled out. This suggests that the strength of the correlation between these two variables made it impossible to fully isolate the two effects from one another given the small number of subjects and items in the current experiment. Nevertheless, it is clear that there is some distinction between the two measures, as is evident from their individual effects on the M170.

Overall, 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. Although 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 (2008), 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 the forms of 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 379–400 msec time range (across both free stems and bound roots), whereas surface frequency did not have this effect. Interestingly, the directionality of the lemma frequency effect was 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). It appears that rather than facilitating performance, high lemma frequency exerts

inhibitory effects on processing. This inhibition might be due to competition arising from members of a word's morphological family—as lemma frequency is the summed frequency of all members of a morphological family, higher lemma frequency is indicative of high family size, which means more possible competitors in the recognition process. This result is consistent with the previous findings of Pykkänen, Feintuch, Hopkins, and Marantz (2004), who also found inhibitory effects of family frequency at the M350 stage. The presence of competition effects among morphological family members suggests that morpheme identification does, in fact, play a key role in lexical access. Our results at the M350 therefore further support a decompositional model of word recognition, which entails that (at least some) words are accessed via their morphological constituents.

Although 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 although 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 suggest that this parsing is, in fact, successful for at least some words, as the effect of lemma frequency indicates that the words' morphological families have been identified by this stage in processing. Nevertheless, our results give us no means of determining which particular words are ultimately recognized by way of their morphological constituents. Further research is necessary to shed light on the distinction between the dual-route and full decomposition models of recognition.

Our general findings at all three stages of processing suggest that both free stems and bound roots undergo the same attempts at decomposition. However, our results are inconclusive as to whether the unique root words undergo attempted decomposition as well. Early effects of affix frequency were present across all three classes, but when each category was examined individually, only the free stems and the bound roots were found to have significant effects. TPL effects were present across the free stems and the bound roots, and were significant or near-significant within each group individually. The lemma frequency effect at the M350 was also present across both the free stems and bound roots. In some cases, cross-category effects were not present in each group 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 were no differences in significant effects between the two categories. Overall, it appears that decomposition

is equally attempted for both free stems and bound roots. The fact that morphological affix frequency effects were significant across all three classes suggests that decomposition might be attempted for unique roots as well, but our results do not provide enough evidence to determine this conclusively.

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.

Conclusion

Our findings throughout the stages of word recognition provide evidence that morphological decomposition is, at the very least, attempted in the processing of both the free-stemmed and bound-root classes of complex words. We suggest that the early visual M130 response reflects the detection of orthographic statistical properties of affixed words, whereas 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. The effect of lemma

frequency is suggestive of competition between morphological family members, which arises only on the decompositional route to word recognition. Although more evidence is needed to establish which particular words are ultimately accessed through the decompositional route, our results clearly point to a model of word recognition in which parsing is attempted at early stages of complex word processing, after which at least some complex words are accessed via their morphological constituents, including bound stems that do not appear as independent words.

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 categories of words, 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, whereas 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.

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

Table A1. Bins for Desired Distributions over Bigram Frequency, Surface Frequency, and Affix Dominance

<i>Normalized z-Score</i>	<i>Desired Weight</i>
$-0.33 < z \leq .33$	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

<i>Length in Letters</i>	<i>Desired Weight</i>
7	1.5
8	1.5
9	1.5
10	1
11	0.5

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 nonwords, potential nonwords 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 nonwords at once.

APPENDIX B: LIST OF STIMULI

<i>Free Stems</i>		<i>Bound Roots</i>		<i>Unique Roots</i>	
bookable	pulsate	durable	levitate	affable	investigate
perishable	ulcerate	equable	stipulate	amenable	isolate
predictable	urate	hospitable	striate	capable	syncopate
printable	contender	numerable	adulterer	formidable	brazier
taxable	destroyer	tolerable	gondolier	malleable	carpenter
variable	discoverer	valuable	idolater	vulnerable	chandelier
autumnal	grounder	criminal	mariner	horizontal	character
figural	pillager	funereal	quarter	lateral	cucumber
parental	voyager	general	sorcerer	magisterial	monster
political	bombastic	literal	aerobatic	parochial	catatonic
proverbial	classic	reciprocal	kinetic	pectoral	cinematic
traditional	metallic	umbilical	phlegmatic	rational	empiric
combatant	symbolic	discrepant	phonetic	adamant	galactic
exultant	talismanic	distant	psychic	blatant	phantasmic
ignorant	titanic	flippant	specific	itinerant	romantic
reliant	destination	important	emulsion	piquant	bullion
trenchant	elusion	tolerant	illusion	valiant	centurion
triumphant	flotation	vibrant	pretension	verdant	champion
budgetary	rebellion	hereditary	purgation	contrary	complexion
dietary	renovation	necessary	section	culinary	ganglion
legionary	vexation	ordinary	station	mammary	mention

APPENDIX B *(continued)*

<i>Free Stems</i>		<i>Bound Roots</i>		<i>Unique Roots</i>	
missionary	avidity	plenary	annuity	papillary	amenity
primary	banality	salivary	fraternity	secondary	animosity
sedimentary	community	voluntary	heredity	veterinary	debility
commentate	continuity	expectorate	precocity	asphyxiate	nativity
motivate	density	facilitate	quantity	escalate	propensity
originate	rationality	generate	vacuity	frustrate	vicinity

<i>Matched Control Words</i>					
address	contact	firmament	lightning	pertain	stalagmite
apprentice	content	flabbergast	liniment	pinafore	starling
architect	contrast	fortress	macaroon	pittance	sterling
armistice	costume	furlough	maintain	plantain	strenuous
artillery	countenance	furnace	mandible	pompadour	substance
balance	cummerbund	furniture	massacre	porcelain	suggest
battery	dachshund	generous	mediocre	precinct	supersede
because	deleterious	ginseng	menopause	previous	support
boisterous	derriere	gizzard	minestrone	principle	syndrome
brassiere	dessert	glycerine	miniature	profound	tenement
bratwurst	determine	griddle	mortgage	provenance	tentative
bronchus	detriment	herring	mountain	pudding	termite
cahoots	disgruntle	holocaust	nicknack	quatrain	terrain
caricature	distinct	hypotenuse	nicotine	querulous	thought
cashmere	distraught	impetuous	nightingale	rationale	torment
casserole	distress	incentive	nonplus	redding	torture
certain	document	ingredient	nuisance	reserve	tournament
chamberlain	dysentery	inkling	obsidian	respond	trampoline
chaperone	earnest	instinct	onslaught	rudiment	tremendous
chieftain	effervesce	instrument	overture	sentence	troubadour
commodore	emollient	intellect	palindrome	shilling	turpentine
compartment	enterprise	interest	palisade	sincere	understand
complain	epicure	isosceles	parsimony	smorgasbord	vamoose
concert	establish	jamboree	participle	soluble	vindictive
condiment	evening	language	pediment	sophomore	voluptuous
connive	example	lascivious	penance	squabble	wildebeest
consort	figment	lattice	persevere	squiggle	witness

APPENDIX C: SUMMARY STATISTICS FOR EXPERIMENTAL STIMULI

	<i>Free Stems</i>	<i>Bound Roots</i>	<i>Unique Roots</i>	<i>Control Words</i>
Length in letters	Mean = 8.61 SD = 1.28	Mean = 8.28 SD = 1.11	Mean = 8.39 SD = 1.19	Mean = 8.43 SD = 1.17
Length in syllables	Mean = 3.26 SD = 0.71	Mean = 3.19 SD = 0.73	Mean = 3.19 SD = 0.83	Mean = 2.59 SD = 0.62
Orthographic neighborhood density	Mean = 0.22 SD = 0.46	Mean = 0.22 SD = 0.46	Mean = 0.17 SD = 0.38	Mean = 0.31 SD = 0.96
Number of morphemes	Mean = 2.00 SD = 0.00	Mean = 2.00 SD = 0.00	Mean = 1.00 SD = 0.00	Mean = 1.00 SD = 0.00
Log surface frequency	Mean = 6.19 SD = 2.62	Mean = 6.56 SD = 2.92	Mean = 6.38 SD = 2.32	Mean = 6.63 SD = 2.81
Log lemma frequency	Mean = 9.61 SD = 1.95	Mean = 8.59 SD = 2.64	N/A	N/A
Lemma transition probability (TPL)	Mean = 0.15 SD = 0.21	Mean = 0.35 SD = 0.33	N/A	N/A
Bigram transition probability (TPB)	Mean = 0.10 SD = 0.14	Mean = 0.10 SD = 0.15	Mean = 0.07 SD = 0.13	Mean = 0.12 SD = 0.20
Morphological affix frequency	Mean = 9.49 SD = 0.91	Mean = 9.49 SD = 0.91	Mean = 9.49 SD = 0.91	N/A
Orthographic affix frequency	Mean = 8.34 SD = 0.93	Mean = 8.34 SD = 0.93	Mean = 8.34 SD = 0.93	Mean = 8.29 SD = 0.93

Reprint requests should be sent to Olla Solomyak, Department of Philosophy, New York University, 5 Washington Place, New York, NY 10003, or via e-mail: olla.solomyak@nyu.edu.

Notes

1. 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.
2. The data collection and preliminary analysis methods described here are the same as those described in Solomyak and Marantz (2009), where the results of the study associated with the filler trials are reported.
3. 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.
4. Clearly unrelated words were words such as "equestrian" for the target word "equable," the roots of which are orthographically, but not morphologically, the same.

5. The HAL frequencies reported in the ELP are calculated from the HAL corpus, which consists of approximately 131 million words gathered from 3000 Usenet newsgroups in February of 1995 (Balota et al., 2007).

6. Using a photodetector, we determined that our stimuli appeared on the screen in front of the subject, consistently, 30 msec after the "trigger" that set time zero for our analyses. Therefore, one may subtract 30 msec from all MEG latencies reported here to obtain latencies from true stimulus onset.

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, VA.
- 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., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression

- study across tasks and modalities. *The Mental Lexicon*, 2, 419–463.
- Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., et al. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. Retrieved from <http://elexicon.wustl.edu/>, Washington University. Accessed in February 2007.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459.
- Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, 81, 65–92.
- Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production* (Vol. 2, pp. 257–294). London: Academic Press.
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26, 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, 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.
- Embick, D., & Marantz, A. (2008). Architecture and blocking. *Linguistic Inquiry*, 39, 1–53.
- Frost, R., Forster, K., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 829–856.
- Giraudo, H., & Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes*, 15, 421–444.
- 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, 1383–1400.
- Hay, J. (2001). Lexical frequency in morphology: Is everything relative. *Linguistics*, 39, 1041–1070.
- Hay, J., & Baayen, H. (2001). Parsing and productivity. In G. Booij & J. van Marle (Eds.), *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, 342–348.
- Lavric, A., Clapp, A., & Rastle, K. (2007). ERP evidence of morphological analysis from orthography: A masked priming study. *Journal of Cognitive Neuroscience*, 19, 866–877.
- 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, 182–193.
- Marinkovic, K. (2004). Spatiotemporal dynamics of word processing in the human cortex. *Neuroscientist*, 10, 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, 394–421.
- Marslen-Wilson, W. D., & Zhou, X. (1999). Abstractness, allomorphy, and lexical architecture. *Language and Cognitive Processes*, 14, 321–352.
- 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, 293–299.
- Morris, J., Frank, T., Grainger, J., & Holcomb, P. J. (2007). Semantic transparency and masked morphological priming: An ERP investigation. *Psychophysiology*, 44, 506–521.
- 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, 456–463.
- Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: An MEG study. *Cognition*, 91, B35–B45.
- Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, 7, 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, 666–678.
- Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23, 942–971.
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11, 1090–1098.
- Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, 37, 118–139.
- Solomyak, O., & Marantz, A. (2009). Lexical access in early stages of visual word processing: A single-trial correlational MEG study of heteronym recognition. *Brain and Language*, 108, 191–196.
- Stockall, L., & Marantz, A. (2008). A single route, full decomposition model of morphological complexity. *The Mental Lexicon*, 1, 85–123.
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology, Section A*, 57, 745–765.
- Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14, 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, 2119–2132.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92, 231–270.
- Zweig, E., & Pylkkänen, L. (2008). A visual M170 effect of morphological complexity. *Language and Cognitive Processes*, 24, 412–439.