

Lexical access in early stages of visual word processing: A single-trial correlational MEG study of heteronym recognition.

Olla Solomyak* and Alec Marantz*

Corresponding Author:

Olla Solomyak
olla.solomyak@nyu.edu

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

* New York University

Abstract

We present an MEG study of heteronym recognition, aiming to distinguish between two theories of lexical access: the ‘early access’ theory, which entails that lexical access occurs at early (pre 200ms) stages of processing, and the ‘late access’ theory, which interprets this early activity as orthographic word-form identification rather than genuine lexical access. A correlational analysis method was employed to examine effects of the heteronyms’ form and lexical properties on brain activity. We find support for the ‘late access’ view, in that lexical properties did not affect processing until after 300ms, while earlier activation was primarily modulated by orthographic form.

1. Introduction

Accounts of visual word recognition differ as to the nature of the neural processing involved within 200ms of visual word presentation and as to the nature of the representations accessed during this early processing. On the one hand, some theories suppose that early processing makes contact with lexical entries for words that include semantic and phonological properties, such that lexical frequency, semantic features and lexicality itself affect neural computation around 100ms post-stimulus onset in a visual word recognition task (Sereno et al., 1998; Pulvermüller et al., 2001; Hauk et al., 2006; Penolazzi et al., 2007). On the other hand, theories such as that associated with Cohen, Dehaene and colleagues suggest that early processing of visually presented words involves bottom-up visual feature detection leading to visual word-form recognition, but not necessarily access to the semantic lexicon, within the first stages of lexical access (Pykkänen & Marantz, 2003; Dehaene et al., 2005; Vinckier et al., 2007). The word form representations themselves might consist of sets of bigram, trigram or higher order combinations of (abstract) letter representations.

Heteronyms like “wind” (either ‘a strong breeze’ or ‘to move along a twisted path,’ depending on the pronunciation) — homographs that are not also homophones — present a possible testing ground for distinguishing the two hypotheses about early word recognition, the one proposing access to the semantic lexicon within 150ms or so of stimulus onset, the other proposing bottom-up recognition through access to a visual word form around 170ms post-stimulus. For the early access theories, the ambiguity of heteronyms should be relevant at the early stages of visual word recognition. In particular, the relative frequency of the different

meanings of such heteronyms might be expected to affect processing by 170ms. On the other hand, for theories that suppose early processing of visually presented words should involve first access to a visual word-form stored in terms of abstract letter patterns, followed by activation of the semantic lexicon, the special properties of the heteronyms might be relevant only after the early stages of processing, with the heteronyms behaving early like non-heteronyms with the same visual form frequency.

We report the results of an MEG lexical decision experiment with heteronyms designed to test these competing theories. In particular, we ask which brain activation from regions and time frames associated with visual word recognition correlate with the special properties of these words. Brain activity associated with access to a lexicon distinguishing words spelled the same but pronounced differently should show correlations with the relative frequency of the meanings of (words represented by) the heteronyms. Brain activity associated with processing of visual word forms independent of the semantic (and phonological) lexicon should only correlate with frequency properties of the word forms of the heteronyms.

Particularly through the work of Tarkiainen, Cornelissen, et al., a coherent picture of neural processing in early stages of visual word recognition is emerging from the ERP/MEG literature. Within 100ms of stimulus onset, visual processing areas in the occipital lobe are active, with activity correlating with the low level visual complexity of the input (the so-called Type I response of Tarkiainen et al.) (Tarkiainen et al., 1999; Tarkiainen et al., 2002; Cornelissen et al., 2003). Within 130ms or so, activity has spread to the occipital/temporal regions of the brain, with activity correlated at least to the distinction between letter strings and symbol strings

(the Type II response of Tarkiainen et al., 1999, also sometimes referred to as the letter string response). By 170ms, activity has spread along the ventral surface of the temporal lobes, including the “visual word form area” that Cohen et al.’s fMRI studies have identified (Dehaene et al., 2002; Cohen & Dehaene, 2004). By this latency, inferior temporal areas have also begun to show stimulus-related activity (see, e.g., Pammer et al. 2004). Examination of peaks in the MEG signal from MEG sensors has led also to the identification of response components to visual words at around 250 and 350ms post stimulus onset (Pylkkänen & Marantz, 2003). In dipole localization analyses of MEG data, these components localize to temporal regions, including classically defined Wernicke’s area.

For this study, we focus on the “M170” response component as well as later activity associated with the M350 component from the literature and with brain regions associated with word recognition. For each neural response identified from the grand average of MEG responses across subjects and words, following the lead of Hauk et al. (2006), we correlated the brain activity at that component with continuous stimulus variables tied to the competing hypothesis about lexical access in visual word recognition. In particular, the early lexical access hypothesis should predict an effect of the relative frequency of the meanings of the heteronyms on the M170 while the late access theory predicts that only variables associated with the visual form of the words should correlate with the M170. On the late access theory, the special properties of the heteronyms should become relevant after the M170, in areas associated with orthography to phonology conversion and in more superior temporal areas associated with the “rich” lexicon, as well as, perhaps, the inferior frontal areas reported in previous research on reading (Marinkovic, 2004; Pammer et al., 2004).

The correlational analysis method that we employ in the current study is a novel approach to MEG data analysis, although it is based in part on the approach of Hauk et al. (2006) to ERP sensor-space data analysis and relies on a standard minimum norm distributed source model for MEG source reconstruction (MNE, MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging, Charleston, MA). Rather than comparing groups of words that are split based on dimensions of interest, we focus on the effects of properties that vary continuously *within* a group of words by correlating the variables of interest directly with brain activity in brain regions and time frames of interest. This correlational analysis technique is especially advantageous in the study of linguistic properties, which are often not balanced across groups of words. In particular, trying to control simultaneously for several potentially significant variables often results in small numbers of candidate stimuli for some cells in a factorial design, for which the notion of random sampling is therefore meaningless (Baayen, 2004). The study of heteronyms presents this problem — the stimuli used in the current experiment are chosen from an extremely limited set of words, and therefore do not represent an unbiased sample. The correlational analysis method introduced here allows us to exploit the unique properties of heteronyms to address an important question about early visual word recognition without violating the assumptions of inferential statistics.

1.1. Methods

Nine subjects participated in the current study, which consisted of a lexical decision task with concurrent MEG recording. Twenty mono-morphemic heteronyms comprised the target stimuli

for the study.¹ Each heteronym had two distinct pronunciations with unrelated meanings. These were presented along with twenty mono-morphemic controls, and forty non-words. Control words were matched to the heteronyms on frequency, length and bigram frequency.² Non-words were matched to the entire group of words on length and bigram frequency. Orthographic properties of the word forms were computed from the CELEX Lexical Database (Baayen et al., 1995), including open bigram frequency (the average frequency of both adjacent and non-adjacent pairs of letters appearing in a particular order within a word) and trigram frequency (see Dehaene et al., 2005). Lexical frequencies for the individual meanings of each heteronym were also obtained from CELEX, which contains separate entries for phonologically distinct words.

The “heteronym frequency ratio,” defined as the ratio of the higher of each heteronym’s lexical frequencies to its form frequency, was computed for each heteronym. This provided a measure of the heteronyms’ ambiguity level, with a higher frequency ratio indicating lower ambiguity in the word’s meaning. In addition to this lexical property, a measure representing the heteronyms’ form was also computed. This measure was based on those of open bigram frequency, trigram frequency and word-form frequency (defined as the sum of each word’s two lexical frequencies). The form measure was defined as the first component of a principal component analysis including these three variables, thus, in theory, encompassing the maximal co-variance of all three form-related properties. The resulting component, however, was primarily correlated with open-bigram and trigram frequencies, rather than with whole-word-form frequency — its coefficients were 0.70, 0.71 and 0.04, for open-bigram, trigram and whole-word-form frequency, respectively.

¹ Heteronyms were mono-morphemic on at least one of the two possible readings.

² The group of control words was matched on its distribution over frequency to the entire group of frequency values for the target words, i.e. including the heteronyms’ word-form and individual meaning frequencies.

MEG data were analyzed using minimum-norm estimation, cortically constrained by each subject's structural MRI. Regions of interest were identified on a representative subject's cortical surface, based on heightened activation in the average minimum-norm solution across all subjects. All regions were then mapped onto individual subjects' brains, after which a trial-by-trial minimum-norm solution was computed for each region. Correlational analyses were conducted to examine the effects of the word-form and lexical properties of interest on the trial-by-trial activation, averaged across vertices within each selected ROI. In order to control for inter-subject variability, each subject's data were normalized into z-scores across trials prior to the cross-subject correlational analyses.

A positive peak in the average activation at about 190ms post stimulus onset was identified as the M170 response component. This activation was found bilaterally in the occipito-temporal fusiform gyrus area. Correlational analyses were conducted on positive activation in this region.³ The left-hemisphere region of interest identified for the analysis of the M170, along with the average time-course of positive activation in the region, is illustrated in Figure 1. M170 analyses were conducted over a 50ms window centered at each subject's M170 peak latency, with timing of activations defined relative to each individual subject's peak. A network of later, negative activation in the left hemisphere superior temporal and Sylvian Fissure areas was identified for analysis of the M350 component. This network showed heightened activation starting at the 250ms range, followed by a peak in negative activity at about 330ms. The region of interest identified by this activation, and the average time-course of negative activity in the

³ Positive activation indicates current flow outward from the cortical surface, while negative activation indicates current flow inward.

region are illustrated in Figure 2. Analyses of M350 activation were conducted over the 305-355ms time window across all subjects (a 50ms window centered at the 330ms peak latency). Analyses of both components were conducted by computing correlations for each time-point within the larger time window of interest. Temporal clusters of consecutive significant effects identified in these time-course analyses underwent a correction procedure for multiple comparisons, based on the technique introduced by Maris and Oostenveld (2007).

The multiple comparison correction was performed by 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. 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 such randomly produced wave, the $\sum r$ statistic was computed for each temporal cluster, after which the highest absolute value of $\sum r$ was taken as the statistic for that permutation. This produced a distribution of 10,000 $\sum r$ 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.

Figure 1: Left hemisphere region of M170 activation, illustrated on a representative subject's inflated cortical surface, along with the average positive activity in the region across all trials and subjects.

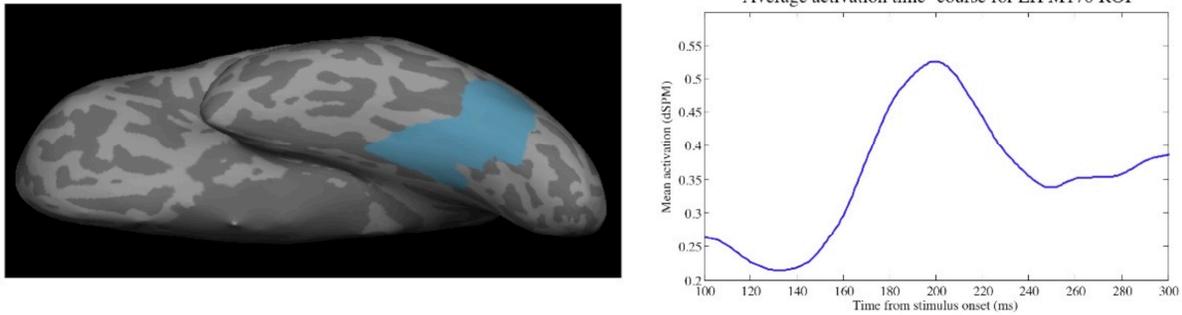
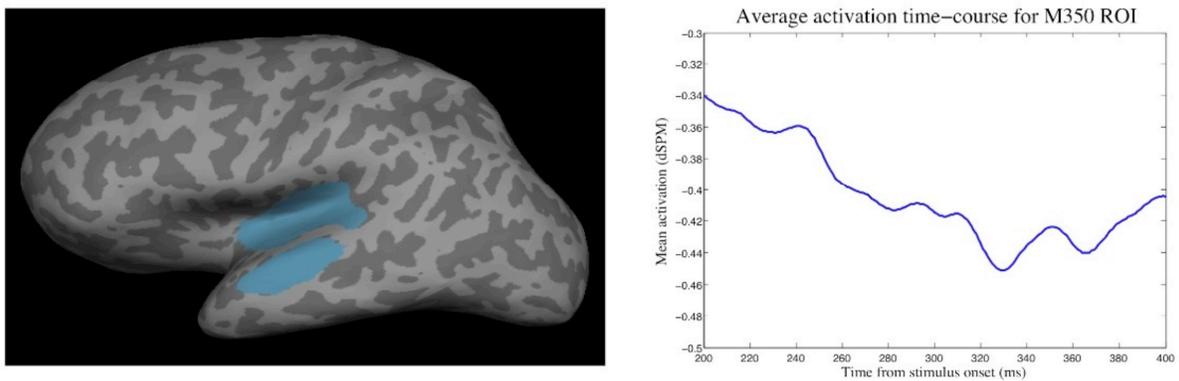


Figure 2: Region of interest identifying network of M350 activation in the left hemisphere, illustrated on a representative subject's inflated cortical surface, along with the average negative activity in the region across all trials and subjects.



2. Results

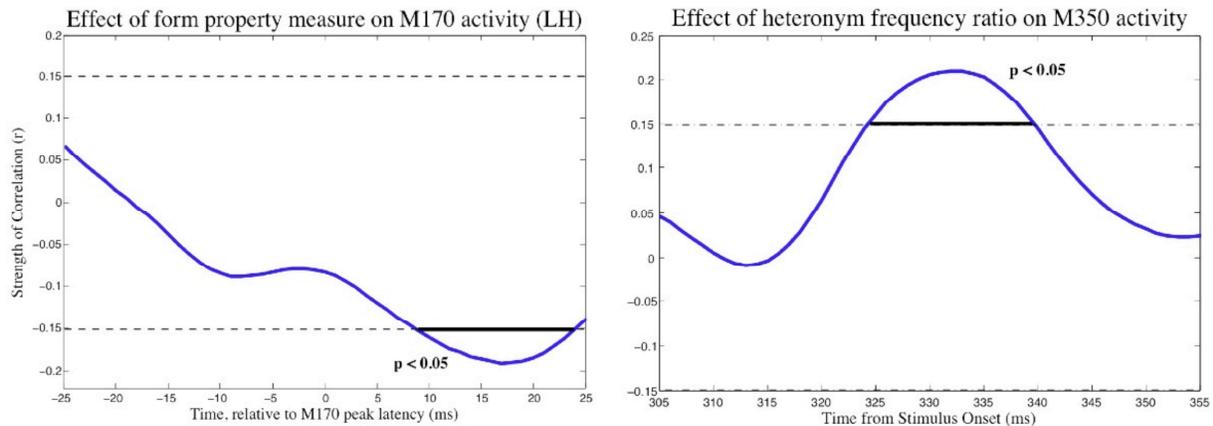
No significant differences were found between heteronyms and controls in behavioral response time or accuracy.

An analysis was conducted on M170 activation over a 50ms time window centered at each subject's individual peak latency. A direct comparison of heteronyms versus controls revealed no significant effects in the left or right hemisphere M170 activity. Correlations were then computed between brain activity and both lexical and form properties of the presented words. There were no significant effects of heteronym frequency ratio on the M170 in either the left hemisphere or the right. There were also no effects of heteronym (low or high) or control frequency. A significant effect of the principal component measure of the heteronyms' form properties was found in the left hemisphere M170 region, starting 8ms after each subject's peak latency, and continuing for 16ms ($\sum r = 2.80$ for 16 time-points, $p < 0.05$ following correction for multiple comparisons). The time-course of this effect is illustrated in the left panel of Figure 3. Right hemisphere M170 activity was not correlated with the form-property measure.

Correlations were computed between M350 activation over the 305-355ms time range and the lexical properties of interest. No significant difference was found in the M350 response to heteronyms as compared to controls. Heteronym frequency ratio was found to have a significant inhibitory effect (i.e. significantly less negative activation) on M350 activity throughout the 325-339ms range ($\sum r = 2.87$ for 15 time-points, $p < 0.05$ following multiple comparison correction). The time-course of this correlation is illustrated in the right panel of Figure 3. Heteronym (low and high) and control word frequency had no effect on the M350. The form frequency measure also had no effect on M350 activity.

Exploratory time-course analyses were also conducted on activation in the 100-150ms and 200-300ms time windows, but these analyses revealed no significant effects.

Figure 3: Left: Effect of the form frequency measure in the left hemisphere M170 region. Right: Effect of the heteronym frequency ratio in the M350 region. Each correlation is plotted over time, with the $p = 0.05$ significance level (prior to correction for multiple comparisons) indicated by the dotted line. Bold lines identify temporal clusters subject to the Monte-Carlo correction procedure. Note that M350 activation is negative, so a positive correlation indicates an inhibitory effect.



3. Discussion

The current study was designed to shed light on the time-course of processing in visual word recognition. In particular, our aim was to distinguish between abstract word-form recognition and genuine lexical access, especially in early stages of processing. Examining visual recognition of heteronyms — phonologically and semantically distinct words that share a common orthography — provides a unique opportunity to distinguish between lexical and word-form properties.

We identified two components of visual word recognition in time and brain space — the M170 and M350 — at which we examined brain activity in response to heteronyms. At the early processing stage of the M170, we found no effects of properties relating to the individual

meanings of the heteronyms, particularly of the frequency ratio of one meaning to the other. The only significant modulator of left hemisphere M170 activity was a measure of the heteronym's form, encompassing open bigram, trigram and whole-word-form frequencies. Only at the later stage of the M350 did we see any indication of higher-level lexical processing, as revealed by the effect of heteronym frequency ratio.

These results suggest that, contrary to the 'early lexical access' theory, words are not identified as lexical items in early stages of processing. Our results point to a model of word recognition in which the earlier M170 response is associated with the identification of abstract, orthographic word-form properties, while the later M350 response is concerned with the processing of higher-level lexical properties and access to the mental lexicon that connects the sound and meaning of words. The presence of a form-related effect at the M170, along with the insignificance of the heteronyms' lexical properties at this early stage in processing, suggests that the early frequency effects reported in the literature may be the reflection of abstract word-form identification rather than genuine lexical access. The heteronyms' true lexical properties do not become relevant until after 300ms post stimulus onset, which is evidence to suggest that lexical access does not occur before this point.

We should emphasize that the null results reported here for possible early lexical influence on visual word formation need to be considered together with the experiments in the literature claiming early semantic effects on lexical access (Serenó et al., 1998; Pulvermüller et al., 2001; Sereno et al., 2003; Hauk et al., 2006; Penolazzi et al., 2007). Many of these claims, however, rest at least in part on early effects of lexicality or frequency (Serenó et al., 1998; Hauk

et al., 2006; Penolazzi et al., 2007). Both of these variables are highly correlated with word-form frequency, and their effects might in fact be attributable to word-form recognition rather than true lexical access. Sereno et al. (2003) and Penolazzi et al. (2007) report early effects of semantic contextual priming, but it is both likely and completely consistent with the hypotheses supported in these papers that semantic priming is also accompanied by activation of the associated word-form. It is thus not clear that the early priming effects must be attributed to semantic rather than word-form expectations. The early contextual priming effect for ambiguous words found by Sereno et al. (2003) is particularly difficult to interpret given its direction. Unlike unambiguous low frequency words, which were facilitated by a biasing context, ambiguous words were inhibited when primed with their subordinate meanings. This discrepancy prevents an entirely straightforward interpretation of the results. In any case, the authors did not consider alternative accounts of their data that would involve activation of the lexical entry for the dominant meaning of an ambiguous word via the word form shared with the entry for the subordinate meaning when the subordinate meaning is semantically primed. The early effect of “semantic coherence” reported by Hauk et al. (2006) is also not completely straightforward. Semantic coherence in the study is based on the semantic relatedness among the members of a word’s morphological family. The semantic transparency of a derived word, like “happiness” from “happy,” has been linked to the decomposition of the word in processing and storage (see, e.g., Hay & Baayen, 2005). The semantic coherence of the morphological family of a word, then, should be related to the activation of the visual word form of the word when members of the family are encountered, with derived words that are more closely related semantically to a stem more likely to be tied to the visual word form of the stem. We might therefore expect word form frequency effects in early stages of visual word recognition to be modulated by the semantic

coherence of a word's morphological family, as reported by Hauk et al. (2006). (See Zweig and Pylkkänen (in press) for a discussion of the evidence for early morphological decomposition in visual word recognition relevant to this issue.) Pulvermüller et al. (2001) report an early effect of semantic association but this result is limited to a single subject design with only four target stimuli for each of four semantic categories. In addition, the experimental paradigm involved memorization of the words and multiply repeated presentations. Overall, these findings do not argue strongly against the late-access theory of visual word recognition, especially given the difficulty in distinguishing between word-form and semantic effects in the majority of situations. Heteronym recognition, in contrast, provides a clear case in which the theories make competing predictions.

The brain regions and patterns of activation identified in the current study are also consistent with a model of visual word recognition in which orthographic word-form properties are detected at the earliest stages of word identification, followed by semantic properties at later stages in processing. The M170 response was localized to the fusiform gyrus area, and corresponds to the “visual word form area” identified by Cohen et al. as a region that responds to abstract orthographic properties of words (Dehaene et al., 2002; Cohen & Dehaene, 2004). Brain areas classically associated with linguistic processing did not show significant activation at this early stage. We should note that nothing in this discussion depends on the existence of a visual word form area specialized for word recognition; the critical activity measured from the fusiform area in this experiment could be generated by either general or word-specific object recognition systems (Price & Devlin, 2003; Devlin et al., 2006).

In contrast to the early M170 response, the later activation of the M350 was localized to superior temporal regions more typically associated with language. This superior temporal network is likely involved in activation of lexical entries and choice among activated representations, as is suggested by the relevance of the heteronym frequency ratio. The effects of heteronym frequency ratio also make sense in the context of the time-course generally associated with word recognition. Given the consensus in the ERP and MEG literature that lexical access happens no later than 350-400ms, the choice of a match for a target heteronym from two separate lexical entries should be completed by about this time (Friederici, 2002; Hagoort, 2008).

The correlational analysis technique employed in the current study is of particular relevance to the results reported here. As the heteronyms used in the experiment were not chosen randomly, but rather comprise the full set of stimuli from English that were usable for our purposes, they do not represent an unbiased sample. Although the control words were matched to the heteronyms on various measures, the imbalance of the stimuli, with the controls randomly chosen from a large set within certain parameters and the heteronyms constituting the entire population of a particular corner of the English lexicon, prevents the direct comparison of the heteronyms versus the controls from being fully compelling. For this reason, we primarily focus on the effects of properties that vary within the group of heteronyms themselves, such as the frequency-ratio measure. The correlational analysis method is a way to test the effects of continuous variables of interest such as this without any binning of the stimuli. As discussed earlier, this technique is useful for analyzing linguistic data in general, as the properties of actual words in any particular language are not always or even usually balanced with respect to experimental variables of interest. The attempt to simultaneously control for multiple variables

often results in very limited candidate stimuli for bins in a factorial design, making random sampling virtually impossible. When some cells in a paradigm exhaust the potential stimuli from a language while others are filled by samples from a large set, standard assumptions behind inferential statistics may be violated (Baayen, 2004).

To summarize, the results of the current study elucidate the nature of visual word recognition as a process that begins with an early visual response to the orthographic properties of a word (the M170), followed by the later M350 response by which orthography is mapped to a particular lexical entry. The latter response component is localized to a region classically associated with language comprehension and seems to be involved in genuine lexical identification, as opposed to the purely abstract form identification that occurs at the stage of the M170. In particular, this model of visual word recognition would entail that, contrary to the ‘early access’ theory, true lexical access does not occur prior to 200ms, but rather at a later stage in processing. This study also presents a correlational analysis method that can provide a means for further investigation of word recognition and other forms of neurolinguistic processing.

4. Formal Methods

4.1. Design and stimuli

Twenty heteronyms comprised the target stimuli for this study. These were chosen from a complete list of heteronyms, from which the following types of words were excluded: heteronyms whose two meanings were morphologically or semantically related to one another, words with the suffix ‘er,’ and words with the past tense suffix ‘ed.’ Heteronyms that were not mono-morphemic on at least one of their two readings were also excluded. All heteronyms that survived this culling process were used in the experiment. All target words had two distinct pronunciations with unrelated meanings. Frequency statistics for the target words were obtained from the CELEX Lexical Database (Baayen et al., 1995), which contains separate entries for phonologically distinct words. (The availability of frequency counts based on pronunciation motivated using heteronyms rather than all homographs in this experiment.) Three frequency values were calculated for each stimulus: two lexical frequencies (one for each individual meaning), and one word-form frequency (for both meanings combined). The “heteronym frequency ratio,” was then computed for each heteronym, defined as the ratio of the higher of its two lexical frequencies to the summed frequency of its form. A lower frequency ratio thus signified more ambiguity in a word’s meaning. A complete list of the heteronyms used in the experiment, along with their frequency values, is provided in Table 1.

Table 1: Heteronyms and their frequency statistics

Heteronym	High Frequency (ln)	Low Frequency (ln)	Frequency-Ratio (high/summed)
bass	6.41	6.18	0.56
bow	7.92	7.68	0.56
buffet	6.27	5.36	0.71
console	6.63	5.68	0.72
content	8.76	8.05	0.67
contract	8.61	7.58	0.74
converse	6.28	4.99	0.78
desert	8.59	7.81	0.69
essay	7.95	4.76	0.96
exploit	8.30	5.86	0.92
incense	6.36	5.55	0.69
lead	10.66	8.12	0.93
resent	7.95	0*	1.00
row	8.78	7.43	0.79
sake	8.80	5.08	0.98
sow	7.41	6.49	0.71
supply	9.69	0*	1.00
tarry	4.19	4.07	0.53
tear	9.02	8.96	0.51
wind	9.67	7.69	0.88

* Raw frequency of 0 in CELEX database

Twenty control words were chosen from the English Lexicon Project (ELP) Database (Balota et al., 2002), to match the target words on length, mean bigram count and frequency. Potential controls were mono-morphemes with a lexical decision accuracy rate above 30% (as reported by the ELP). The twenty control words were chosen from this list by an automatic word-choice algorithm, such that they would match the target words on their distributions over the three variables. The distribution over frequency was based on the whole range of frequency values associated with the target words, including all three values (high, low and summed) for each stimulus. The word-choice algorithm was then employed to choose forty non-words to match the entire group of words on length and bigram frequency. Potential non-words were also obtained from the ELP, and had lexical decision accuracy rates above 30%. The 80 stimuli (40 words and 40 non-words) associated with this study were presented randomly interspersed with 646 items from another study.

In addition to the lexical properties of the heteronyms, form-related properties of the stimuli were computed from the CELEX database. These included mean open-bigram frequency (the average frequency of both adjacent and non-adjacent pairs of letters appearing in a particular order within a word) and mean trigram frequency (see Dehaene et al., 2005). A single measure was then created to represent the heteronyms' form properties. This measure was defined as the first component of a principal component analysis of three variables: open bigram frequency, trigram frequency and whole-word-form frequency. The coefficients for this principal component were 0.70, 0.71 and 0.04, for open-bigram, trigram and whole-word-form frequency, respectively.

4.2. Experimental procedures

Participants of the current study were nine right-handed native English speakers, ranging in age from 19 to 29, with a mean age of 23.3. Subjects were members of the MIT community. 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. Subjects participated in a continuous lexical decision task. The PsyScope X Experiment Control System was used to program stimulus presentation. The experiment began with 10 practice trials, consisting of 5 words and 5 non-words in randomized order, to accustom participants to the task. 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 a button with the left index finger if they recognized the stimulus as a word and the left middle finger if they did not recognize it as a word. The button press was marked by a short tone, after which there was an inter-trial interval of 500ms. Stimuli were presented in a size 48 Courier font. After every 200 trials, subjects were given a rest from the task, which lasted until they indicated by a button press that they were ready to continue.

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.

4.3. Analysis methods

4.3.1. Behavioral data

Behavioral response times and accuracy data on the lexical decision task were collected for each participant. Trials with incorrect responses, which amounted to 9% of all trials, were excluded prior to behavioral analyses. All trials with response times above 5 seconds, and one trial with a 60ms response time, were excluded as well. Response time data were then normalized for each subject across all trials (including fillers), and trials for which the response time was more than 3 standard deviations higher than the subject's mean were also excluded. Trials excluded due to extreme response time amounted to an additional 0.6% of the trials relating to the current study. After exclusion of errors and extreme response times, the mean RT across all nine subjects was 738ms. These exclusion criteria only pertained to analyses of behavioral measures and not to MEG analyses.

4.3.2. MEG data

4.3.2.1. Minimum-norm estimation

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). Raw MEG data were then imported into MNE (MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging, Charleston, MA) for processing and minimum-norm analysis. The raw data were first used in MNE to create a noise covariance matrix, which assigned weights to each MEG channel with the purpose of minimizing the effects of noise on minimum-norm calculations. The data were then baseline-corrected using 100ms of pre-stimulus data for each trial. Following baseline-correction, the data were averaged over all trials and low-pass filtered at 40Hz.

Structural MRI data 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. These reconstructions were utilized by MNE in order to constrain the minimum-norm solutions for the acquired MEG data. To make this possible, the MEG and MRI coordinate frames were aligned based on the HPI coil locations and head shape digitization, using the tools provided in the MNE software. 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 actual 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).

In order to define regions of interest for correlational analysis, the dSPM data were averaged across all subjects. For averaging purposes, each subject's cortical surface was mapped onto that of a single representative subject, by use of the morphing procedure in MNE. The average source activation was then projected onto this representative surface for visualization. Peaks in the average activity were used to define regions and time-windows of interest. The regions of interest, which were originally defined on a single cortical surface, were mapped to equivalent regions on each individual subject's brain. An inverse solution was then computed over each subject's raw MEG data, individually over each 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 that consisted of an estimate of activity per time-point at each source within the specified region. Only components of activity that were in the direction normal to the cortical surface were retained in the minimum-norm solution, while components that were in a direction tangential to the surface were ignored. These minimum-norm estimates were converted into noise-normalized dSPM values for use in the correlational analyses.

4.3.2.2. M170 region

Regions of interest were defined based on positive peaks in activity in the left and right hemisphere fusiform gyrus areas, in the M170 time range. As the directionality of the M170 component has been fairly consistent, both in the current study and in our previous experience (Pykkänen & Marantz, 2003), analyses were conducted on positive activity (i.e. current flow directed outwards from the ventral cortical surface). The negative activity was therefore zeroed, after which the activation in each region was averaged across all sources. The peak in the 150ms - 220ms range within each region's average time-course was defined as the subject's M170 peak in the specified hemisphere. The mean M170 latency across all subjects was 189.7ms in the left hemisphere and 178.9ms in the right. The range of latencies was 157 - 216ms in the left hemisphere, and 153 - 201ms in the right. Correlational analyses were conducted over a 50ms window centered at each subject's average peak latency.

4.3.2.3. M350 region

A region of interest was identified for analysis of the M350 component, based on a network of heightened negative activation in the left hemisphere in the time relevant time range. Analyses were conducted on negative activity (i.e. current flow directed inward from the cortical surface). The positive activity was zeroed prior to averaging across the sources in the region. A peak in activation at 330ms was observed in the average time-course of negative activity across all subjects. The pattern of activation in this network was not as consistent across subjects as was the M170 activity, and individual subjects' peaks could not be consistently identified. M350

component activity was therefore defined across all subjects as negative activation in the 50ms time range centered at the 330ms peak latency (i.e. the 305-355ms range).

4.3.2.4. Correlational time-course analyses

Activation at each time-point within the time-range of interest for each region 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. For the M170 analyses, time was defined as relative to each subject's individual M170 latency, while the M350 analyses were conducted with the timing of activation defined in absolute terms across all subjects. Temporal clusters showing consecutive significant effects were identified, after which a nonparametric statistical test was employed to correct for the multiple comparison problem that arises due to the large number of computations (based on the methods of Maris & Oostenveld, 2007).

The multiple comparison correction was performed by 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. 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 such randomly produced wave, the $\sum r$ statistic was computed for each temporal cluster, after which the highest

absolute value of $\sum r$ was taken as the statistic for that permutation. This produced a distribution of 10,000 $\sum r$ 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.

In addition to the peak-based analyses of the M170 and M350 components, exploratory time-course analyses were conducted on several regions and time-windows of interest. These included a region of earlier, negative activation that peaked at about 150ms, as well as the 200-300ms range in the M350 region described above. Exploratory time-course analyses were conducted in the same manner as were the M170 and M350 analyses, but were computed over larger (100ms) time-windows of interest.

4.3.2.5. MEG Outlier Exclusion

Prior to the correlational analyses, outlier rejection for each subject was performed based on the original raw MEG data. Outliers were excluded by first counting the number of data points in each trial that were not within two standard deviations of the overall mean of the data. This created a vector of ‘extreme value counts,’ with one value per trial. Trials were 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 (including fillers), of which 20 trials were relevant to the current experiment (6 heteronym trials, 1 control trial and 13 non-word trials). MEG analyses

were therefore conducted over 174 heteronym trials, 179 control trials and 347 non-word trials (across all 9 subjects).

Unlike the behavioral analyses, MEG analyses were conducted without excluding error and extreme response time trials, due to a strong correlation between frequency and accuracy. For the heteronyms, both word-form frequency and the higher of the two lexical frequencies were positively correlated with accuracy ($r = 0.35$, $p = 0$ for both). Frequency of the matched control words was also significantly correlated with accuracy ($r = 0.27$, $p < 0.0005$). These results indicate that excluding error trials from analysis would have been problematic, as it would have substantially altered the distributions of the continuous independent variables relating to frequency.

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., Piepenbrock, R. & Gulikers, L., 1995. *The CELEX Lexical Database (Release 2) [CD ROM]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Baayen, R.H., 2004. Statistics in Psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers I*.
- 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.
- Cohen, L. & Dehaene, S., 2004. Specialization within the ventral stream: the case for the visual word form area. *NeuroImage* 22(1), 466-76.
- Cornelissen, P, Tarkiainen, A, Helenius, P., Salmelin, R., 2003. Cortical effects of shifting letter position in letter strings of varying length. *Journal of Cognitive Neuroscience* 15(5), 731-46.
- 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., Cohen, L., Sigman, M., Vinckier, F., 2005. The neural code for written words: a proposal. *Trends in Cognitive Science* 9(7), 335-41.

- 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), 321-5.
- Devlin, J.T., Jamison, H.L., Gonnerman, L.M., Matthews, P.M., 2006. The role of the posterior fusiform gyrus in reading. *Journal of Cognitive Neuroscience* 18(6), 911-22.
- Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science* 6(2), 78-84.
- Hagoort, P., 2008. The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363(1493), 1055-69.
- Hauk, O., Davis, M.H., Ford, M., Pulvermüller, 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), 1383-400.
- Hay, J. & Baayen, R.H., 2005. Shifting Paradigms: gradient structure in morphology. *Trends in Cognitive Science* 9(7), 342-48.
- Marinkovic, K., 2004. Spatiotemporal dynamics of word processing in the human cortex. *Neuroscientist* 10(2), 142-52.
- Maris, E. & Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* 164, 177-90.
- Pammer, K., Hansen, P.C., Kringelbach, M.L., Holliday, I., Barnes, G., Hillebrand, A., Singh, K.D., Cornelissen, P.L., 2004. Visual word recognition: the first half second. *NeuroImage* 22(4), 1819-25.

- Penolazzi, B., Hauk, O., Pulvermüller, F., 2007. Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology* 74(3), 374-88.
- Price, C.J. & Devlin, J.T., 2003. The myth of the visual word form area. *NeuroImage* 19(3), 473-81.
- Pulvermüller, F., Assadollahi, R., Elbert, T., 2001. Neuromagnetic evidence for early semantic access in word recognition. *European Journal of Neuroscience* 13(1), 201-5.
- Pylkkänen, L. & Marantz, A., 2003. Tracking the time course of word recognition with MEG. *Trends in Cognitive Science* 7(5), 187-9.
- Sereno, S.C., Brewer, C.C., O'Donnell, P.J., 2003. Context effects in word recognition: evidence for early interactive processing. *Psychological Science* 14(4), 328-33.
- Sereno, S.C., Rayner, K., Posner, M.I., 1998. Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *NeuroReport* 9(10), 2195-200.
- Tarkiainen, A., Cornelissen, P.L., Salmelin, R., 2002. Dynamics of visual feature analysis and object-level processing in face versus letter-string perception. *Brain* 125(5), 1125-36.
- 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(11), 2119-32.
- Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., Cohen, L., 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron* 55(1), 143-56.
- Zweig, E. & Pylkkänen, L., in press. A visual M170 effect of morphological complexity. *Language and Cognitive Processes*.