Brain & Language 129 (2014) 14–23

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

The role of morphology in phoneme prediction: Evidence from MEG

Allyson Ettinger [a,b,d,*], Tal Linzen [a], Alec Marantz [a,b,c]

[a] Department of Linguistics, New York University, United States
[b] Department of Psychology, New York University, United States
[c] NYUAD Institute, New York University Abu Dhabi, United Arab Emirates
[d] Department of Linguistics, University of Maryland, United States

ARTICLE INFO

Article history:
Accepted 24 November 2013

Keywords:
MEG
Spoken word recognition
Prediction
Morphology
Surprisal
Entropy

ABSTRACT

There is substantial neural evidence for the role of morphology (word-internal structure) in visual word recognition. We extend this work to auditory word recognition, drawing on recent evidence that phoneme prediction is central to this process. In a magnetoencephalography (MEG) study, we crossed morphological complexity (*bruis-er* vs. *bourbon*) with the predictability of the word ending (*bourbon* vs. *burble*). High prediction error (surprisal) led to increased auditory cortex activity. This effect was enhanced for morphologically complex words. Additionally, we calculated for each timepoint the surprisal corresponding to the phoneme perceived at that timepoint, as well as the cohort entropy, which quantifies the competition among words compatible with the string prefix up to that timepoint. Higher surprisal increased neural activity at the end of the word, and higher entropy decreased neural activity shortly after word onset. These results reinforce the role of morphology and phoneme prediction in spoken word recognition.

1. Introduction

Recent work has illuminated the role of morphology in visual word recognition. Evidence from both behavioral and brain-based studies strongly indicates that visually presented words are decomposed into morphemes based on their visual forms, and that this visual decomposition feeds lexical access for the lexical information associated with morphemes (Fiorentino & Poeppel 2007a, 2007b; Rastle, Davis, & New, 2004; Solomyak & Marantz, 2010, inter alia). Recognition of visually complex words, then, follows the decomposition, look-up, and recomposition model championed by Taft and others based on behavioral data (Taft, 2004; Taft & Forster, 1975).

The role of morphological structure in auditory word recognition is less studied and less well understood, though it has been known for some time that morphological structure plays a role in auditory word recognition as well (Marslen-Wilson, Tyler, Waksler, & Older, 1994). Recent studies by Baayen, Wurm, and colleagues point to a predictive role for morpheme recognition during auditory word recognition (Balling & Baayen, 2008, 2012; Wurm, 1997; Wurm, Ernestus, Schreuder, & Baayen, 2006). In particular, Balling and Baayen (2008 and 2012) contrast two general models of recognition. On one, all of the full words consistent with the auditory input, the full word cohort of the word eventually recognized, are activated during recognition, with their competing representations compared against the incoming acoustic signal (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). On this class of models, the morphological decomposition of members of the cohort would not be relevant to cohort competition and would not interact with prediction of upcoming phonemes from the cohort consistent with the already processed phonemes. A second model would leave a role for recognition of component morphemes of the word being processed (Balling & Baayen, 2008, 2012). Recognition of a morpheme, for example a morphological prefix or a stem, would yield predictions for upcoming morphemes, and upcoming phonemes as part of these morphemes. We test the hypothesis that this additional prediction associated with morphological structure might enhance the prediction of upcoming phonemes based on the cohort of full words consistent with the input.

In addition to questions associated with the role of morphology in auditory word recognition, the actual mechanisms, both cognitive and neural, whereby a cohort of possible words influences prediction and processing of the incoming speech stream has also been addressed in the recent literature. On the one hand, one might endorse a competition model in which all members of a cohort are activated to an extent proportional to their frequency of occurrence, with lateral competition between activated cohort members (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). On this view, the more members of a cohort and the more evenly distributed their frequency, the more cognitive and neural activity associated with activation and active inhibition. In a recent paper, Gagnepain, Henson, and Davis (2012) suggest that cohort

* Corresponding author at: Department of Linguistics, University of Maryland, 1401 Marie Mount, College Park, MD 20742, United States.
E-mail address: aetting@umd.edu (A. Ettinger).

0093-934X/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bandl.2013.11.004

competition might not be a main driver of brain activity associated with auditory word recognition. Rather, they suggest, the unexpectedness of the incoming input given the probability distribution over possible continuations from what has already been processed is the factor that drives neural activity and affects response times. They suggest that activity from areas around auditory cortex reflect the mismatch between predicted and incoming phonemic material—a prediction error signal. Their results are also consistent with the activity reflecting surprisal, which quantifies the change in the probability distribution over the members of the active cohort based on the incoming stimulus (Hale, 2001).

For the role of morphology in auditory word recognition, Balling and Baayen (2012) propose that online morphological decomposition leads to prediction, for stems from morphological prefixes and for suffixes from stems, that is observable in reaction times in lexical decision. All other things being equal, being morphologically complex aids in auditory word recognition over suitably matched monomorphemic words (Balling & Baayen 2008; Ji, Gagné, & Spalding, 2011). From Gagnepain et al. (2012), we derive the hypothesis that cohorts of words consistent with auditory input show their face neurally not through competition, where higher entropy among members of a cohort would lead to more activity, but through surprisal, where changes in the probability distribution over cohort members, with new input, might drive activity—or through prediction error, where observed lower probability continuations conflict with their higher probability cohort members. From neural and behavioral work, we might be led to an opposite conclusion about cohort competition at the beginning of the recognition of auditorily presented words: the higher the entropy over the cohort, the less activation we will observe (Baayen, Wurm, & Aycock, 2007; Linzen, Marantz, & Pylkkänen, 2013; Wurm et al., 2006). On this view, which we will refer to as the Low-Entropy Dependent Prediction model (LEDP), higher entropy prevents commitment to prediction for upcoming input, leading to less predictive work and less neural activity, while low entropy allows more commitment to continuations, thus more work and more activity.

The present study is a preliminary look at whether contemporary magnetoencephalography (MEG) measurement and analysis techniques are suitable for investigating the role of morphology and of cohort entropy online, as participants listen to spoken language. In particular, we ask whether morphological complexity enhances the prediction of upcoming phonemic material such that we see more neural activity associated with surprisal during auditory word processing for morphologically complex as opposed to monomorphemic words.

To approach these questions we manipulated morphological complexity and surprisal in a 2-by-2 factorial design. The morphological complexity manipulation consisted of two morpheme conditions: monomorphemic and bimorphemic words. The surprisal manipulation consisted of two continuation surprisal conditions: high- and low-surprisal continuations of a shared string prefix. For bimorphemic pairs (e.g., *bruises*/*bruiser*), this string prefix consisted of a morphological stem (*bruis-*), and the continuations consisted of differing suffixes (*-es*/*-er*). For monomorphemic pairs (e.g., *bourbon*/*burble*), this shared string prefix consisted of phonological material that did not constitute a morpheme (*[bɹb]*), and the continuations were the final phonemes of the words, which similarly were not morphemes (e.g., *-ən*/*-əl*).

Each participant heard both words in each pair. This introduced a third variable: order of presentation within pair. Given evidence of long-distance morphological priming in auditory word recognition (Kouider & Dupoux, 2009), this variable is of particular note for bimorphemic pairs. As such, the ordering position of a stimulus relative to its pair counterpart was also taken into account in the analysis.

1.1. Continuous variables

In addition to the factorial design described above, we also investigate whether we can detect neural activity related to individual phoneme prediction on a trial-by-trial basis (Gagnepain et al., 2012). We used a frequency database to calculate the probability distribution of all English words compatible with the string prefix at each time point in each trial. This enabled us to derive millisecond-by-millisecond estimates of two information-theoretic quantities: phoneme surprisal and cohort entropy.

The surprisal of the phoneme that is currently being heard is the inverse of the log of its conditional probability given the phonemes that preceded it. This probability can be calculated by dividing the total frequency of the present cohort by the total frequency of the cohort that was "alive" prior to hearing the current phoneme. Formally, if $PrefFreq(x)$ is the summed frequency of all words that start with the phoneme sequence x, then the surprisal of the third phoneme u in *bruiser* ([bruzr]) would be given by

$$-\log_2(PrefFreq(bru)/PrefFreq(br)).$$

The cohort entropy is the entropy of the probability distribution over all words that are compatible with the string prefix heard thus far. If C is the cohort, $f(w)$ is the frequency of a word w, and F_C is the total frequency of the cohort, then the cohort entropy is given by:

$$\sum_{w \in C} \frac{f(w)}{F_C} \log_2 \frac{f(w)}{F_C}.$$

We predict that higher phoneme surprisal should lead to increased neural activity. We also investigate whether we can observe an effect of cohort entropy, and, if so, whether the effect is such that neural activity increases with increased entropy, as predicted by competition models, or decreases with increased entropy, as predicted by the LEDP.

A third variable that we calculated was cohort *frequency*, which is the summed frequency of all the words in the cohort. The prior literature does not afford a specific hypothesis about the potential significance of this variable. However, in light of the pervasiveness of frequency effects in language processing, we examine the effect of this variable informally, and leave an in-depth investigation of its significance for future research.

1.2. Predictions

In light of prior work suggesting a key role of error detection/surprisal in auditory processing, as well as work suggesting a predictive role for morpheme recognition in auditory processing, we expect to observe an effect of surprisal which is enhanced in morphologically complex words, relative to simple words. This interaction with morphological complexity should apply to the categorical variable of continuation surprisal, as well as to any effects of phoneme-by-phoneme surprisal.

We furthermore test for facilitatory effects of entropy early in the stimulus, as predicted by LEDP models. While cohort competition models predict an inhibitory effect of entropy during word comprehension, LEDP models would predict facilitation near word onset, when entropy is high.

In accordance with the results of Gagnepain et al. (2012), we expect to see effects of surprisal in auditory regions such as the superior temporal gyrus (STG) and transverse temporal gyrus (TTG). Given that morphological decomposition involves accessing lexical entries, we may also expect to see evidence of the predicted interaction with morphological complexity in the middle temporal gyrus (MTG). These will serve as our three regions of interest (ROIs) for the analysis.

2. Methods

2.1. Design and stimuli

The experiment consisted of an auditory lexical decision task, with simultaneous MEG recording of the magnetic fields induced by electrical activity in the brain. The factorial design included two two-level stimulus variables of interest: morphological complexity (bimorphemic and monomorphemic) and continuation surprisal (high- and low-surprisal continuations).

Stimulus words were chosen in pairs from the English Lexicon Project (ELP) (Balota et al., 2007). For bimorphemic words, continuations of various stems were identified within the ELP, and from these the highest and lowest-frequency disyllabic items were selected. For monomorphemic words, a search was performed on the ELP to determine uniqueness points (UP) of disyllabic monomorphemic words. Inflected forms were excluded as competitors in this search, to avoid the UP falling always at word offset. Once the UP was identified, the point in the transcription just prior to that UP (the final point at which competitors remained) was then selected as the end of what served as the shared string prefix for our monomorphemic stimuli. As with the bimorphemic words, disyllabic competitors sharing this phonological material were identified, and the highest and lowest-frequency items among these were selected. In all but one monomorphemic pair (*guava*/*guano*), the point at which the pair members became phonologically distinct occurred either at or after the syllable break. The average number of shared phonemes was 3.8 (average number of shared phonemes in bimorphemic words was 4.3).

The bimorphemic and monomorphemic stimulus lists were then balanced for surface frequency. For the purposes of balancing, in some cases the next-highest-/next-lowest-frequency disyllabic competitor was substituted for the highest-/lowest-frequency pair member.

This process yielded 356 disyllabic words: 178 bimorphemic words paired according to shared initial morpheme (*bruises*/*bruiser*), and 178 monomorphemic words paired according to shared phonological material (*bourbon*/*burble*). Each stimulus pair consisted of a high-surprisal and a low-surprisal continuation of the shared material.

In order to determine whether stimulus conditions were matched, continuation surprisal values were calculated for all stimuli (see Table 1). There was a main effect of morphological complexity ($p = .005$), such that monomorphemic words had on average higher continuation surprisal than bimorphemic words. Crucially, there was no significant interaction between morphological complexity and continuation surprisal ($p = .774$), which is of most interest in this study.

Stimulus recordings were drawn from the online pronouncing dictionary Howjsay (http://www.howjsay.com/). In addition to the 356 stimuli of interest, we selected 356 nonwords. Since we used a pronouncing dictionary, our nonwords consisted of extremely low-frequency disyllabic words available in the dictionary (e.g., *awacs*, *blatnoy*, *edfu, flagyl*, *hemsut*, *judder*, *lobar*, *quassin*, *wejack*, *zenzic*).[1]

[1] A referee points out that the dictionary we used is a British pronunciation dictionary, yet our participants were generally speakers of American English. We used this dictionary because unlike comparable websites it provides recordings of inflected forms, and contains a large number of low frequency items that could serve as nonwords. In cases of a clear divergence between the British and American pronunciation (e.g. "tomato", pronounced [təmɑːtəʊ] in the UK and [təmeɪroʊ] in the US), the site provides both pronunciations; we used the American pronunciations for these words. Even the American pronunciations were recorded by a speaker who had a British-sounding accent, however. We informally debriefed the participants about the materials. While a few of them reported that the task was made somewhat more difficult by the speaker's accent, this never resulted in the participants being unable to comprehend the stimuli.

Table 1
Continuation surprisal means for all conditions.

	High surprisal	Low surprisal
Monomorphemic	4.00	1.21
Bimorphemic	3.58	0.68

The nonwords were almost certainly unfamiliar to our participants. When debriefed post-experiment about their impression of the word-to-nonword ratio, most subjects reported impressions of a 50–50 split. A few subjects reported an impression of more nonwords than words, and only a single subject suspected the presence of low-frequency words being used in place of nonwords—suggesting that the extremely low-frequency words indeed functioned as nonwords, as intended.

Stimulus editing was performed in Audacity (http://audacity.sourceforge.net/). Silence at the beginning and end of each recording was trimmed, and peak volume was normalized across recordings. Each recording was then modified to a constant duration of 750 ms. This duration modification was performed using the Audacity function "Change tempo", a functionality allowing a change in duration without a change in pitch. Duration modifications were monitored to ensure that final products sounded as natural as possible.

While the stimulus recordings were edited to durations of 750 ms, the final .wav files contained a 12 ms buffer prior to the onset of the acoustic signal, as well as a 10 ms buffer after the offset. As a result, the acoustic signal occurred approximately between 12 and 740 ms into the sound file.

The participants heard both items of each pair. In order to counterbalance the order of presentation within each pair, stimuli were divided into two super-blocks, each containing one member of every stimulus pair, with morphological complexity and continuation surprisal balanced between blocks. The relative order of these super-blocks was then varied between subjects (see details in Procedure section).

2.2. Procedure

Thirteen right-handed native English speakers (6 female) participated in the experiment. Handedness was assessed using the Edinburgh Handedness Inventory (Oldfield, 1971). All subjects provided written informed consent before participation.

Prior to recording, the head shape of each participant was digitized to allow source localization and coregistration with structural MRIs (Fastscan; Polhemus, VT). We also digitized three fiducial points (the nasion and the left and right pre-auricular points) and the position of five coils, placed around the participant's face. Once the participant was situated in the magnetically shielded room for the experiment, the position of these coils was localized with respect to the MEG sensors, allowing us to assess the position of the participant's head for source reconstruction. Data were recorded continuously at the KIT/NYU facility with a 157-channel axial gradiometer (Kanazawa Institute of Technology, Kanazawa, Japan) in a dimly lit magnetically shielded room. Data were low-pass filtered at 200 Hz, with a notch filter at 60 Hz. Stimuli were presented using Matlab PsychToolbox (Brainard, 1997; Pelli, 1997). Each trial of the experiment consisted of a fixation cross-presented for 500 ms, followed by the onset of the auditory stimulus. The fixation cross remained on the screen for the duration of the stimulus and disappeared only after a response was given. Subjects responded to the stimulus by pressing one of two buttons with their left hand to indicate whether they recognized the stimulus as a word of English. The inter-trial interval was randomly selected between 750 ms and 1250 ms. Trials were randomized

within super-blocks (halves) of the experiment, and each super-block was further divided into two sub-blocks in order to allow subjects additional breaks. In total, subjects heard four blocks of 178 trials each, with breaks between each block. Participants were allowed to choose the length of their breaks and proceed to the following block when ready.

2.3. Data processing

The preprocessing and analysis of the MEG data closely followed the procedures of Solomyak and Marantz (2009 and 2010). Environmental noise was removed from the data by regressing signals recorded from three orthogonally oriented magnetometers approximately 20 cm away from the recording array against the recorded data using the continuously adjusted least squares method (CALM; Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 2001).

Outlier trials were excluded based on an absolute threshold of ±3.0 pT, enforced over the time window [0 ms, +1000 ms] for the noise-reduced MEG data (the number of rejected trials ranged from 55 to 144, mean 90, median 79). All nonword trials were excluded.

2.4. Source space analysis

MNE software (Martinos center MGH, Boston) was used to estimate neuroelectric current strength based on the recorded magnetic field strengths using minimum l_2 norm estimation (Dale & Sereno, 1993; Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). Current sources were modeled as three orthogonal dipoles spaced approximately 5 mm apart across the cortical surface (Dale et al., 2000), yielding 2562 potential electrical sources per hemisphere. For all subjects, structural MRIs were available from previous experiments, and each subject's cortical surface was reconstructed based on this structural MRI using Freesurfer (Martinos center). The neuromagnetic data were co-registered with the structural MRI using MNE by first aligning the fiducial points, and then using an Iterative Closest Point algorithm to minimize the difference between the points defining the head shape of each participant, and the scalp.

The forward solution was calculated for each source using a single-layer boundary element model (BEM) based on the inner-skull boundary. The estimated activation was normalized by dividing the estimated activation by the predicted standard error of the estimate, yielding Dynamic Statistical Parametric Maps (Dale et al., 2000).

Regions of interest were defined anatomically, using the cortical parcellation performed by FreeSurfer based on the Desikan–Killiany gyral atlas (Desikan et al., 2006). Signed activity was summed across each ROI.

3. Analysis methods

3.1. Regions of interest analysis

Analysis was conducted on three ROIs in the temporal lobe: the transverse temporal gyrus (auditory cortex), superior temporal gyrus, and middle temporal gyrus. The anatomical FreeSurfer labels (Desikan et al., 2006) corresponding to these regions served as the ROIs for the analysis. The inverse solution over all trials was calculated within the target label of each individual subject. The transverse temporal, superior temporal, and middle temporal ROIs, and grand average activation (at around 500 ms) are pictured in Fig. 1A. Fig. 1B shows grand average activation in all ROIs over time.

3.2. Factorial analysis

The continuation surprisal manipulation had its critical effect at the end of each stimulus (the beginning of the stimulus was identical between the two words in each pair). As such, the 200 ms time window following the offset of the word was selected for the factorial analysis. We averaged the neural activity in the 200 ms time window following the offset of the word to yield a single value per trial, and submitted the single-trial averaged activity values to a series of linear mixed-effects models using the *lme4* package (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2012). Each model included a by-item intercept, a by-subject intercept, and by-subject slopes for all of the independent variables (a maximal random effect structure, following Barr, Levy, Scheepers, & Tily, 2013). We obtained *p*-values using the chi-squared approximation for the likelihood ratio test. In this test, the difference between the deviance (twice the log likelihood) of a model without the fixed effect of interest and a model with the fixed effect (but still with a maximal random effects structure) is assumed to be χ^2-distributed with one degree of freedom (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000). This was done sequentially: the two predictors (morphological complexity and continuation surprisal) were added one by one, followed by their interaction. This procedure is a generalization of a sequential ("Type I") ANOVA. In a balanced design such as ours, the *p*-values derived from this procedure are an accurate measure of the contribution of each factor. The same procedure was employed for the analysis of the behavioral measures, with the exception that the accuracy data were analyzed by means of logistic regression. Trials for which reaction times were more than 2.5 standard deviations above the (log-transformed) mean reaction time for all subjects were excluded for reaction time analyses.

3.3. Continuous analysis

Phoneme-by-phoneme surprisal and entropy values were calculated based on the procedure described in the introduction. Specifically, we used the SUBTLEX-US word frequency database (Brysbaert & New, 2009) to calculate the probability distribution over all string-prefix cohorts in English, based on the phonemic transcriptions provided in the English Lexicon Project (Balota et al., 2007). Segment boundaries were then located in the acoustic signal of each stimulus through combined use of the Penn Forced Aligner (Yuan & Liberman, 2008) and visual inspection. In this manner, millisecond-by-millisecond surprisal and entropy values were obtained for use in mixed-effects model analysis (see Fig. 2).

Analysis of the phoneme surprisal variable involved the following variation on the correlation wave method used in Solomyak and Marantz (2009). We fit linear mixed-effects models at each millisecond within our time window of interest. For each millisecond of stimulus duration (750 ms), the corresponding surprisal values were correlated with the brain activity at one millisecond within a 750 ms time window following (not necessarily directly; see below) the onset of the stimulus.

Multiple comparisons correction was performed over the same time window, using a cluster-based permutation test (Maris & Oostenveld, 2007), as adapted by Solomyak and Marantz (2009). Specifically, the largest cluster of significant *t*-values in the same direction (with significance threshold set at $t > 1.96$, $p < 0.05$ uncorrected) was identified and compared to the largest cluster of significant values in each of 1000 random permutations of the independent variable (cluster size here refers to the sum of the *t*-values within the cluster in question). We fit a total of 750,000 mixed-effects models for each analysis (750 timepoints times 1000 permutations), which made it computationally impractical to use a maximal random effect structure in the linear

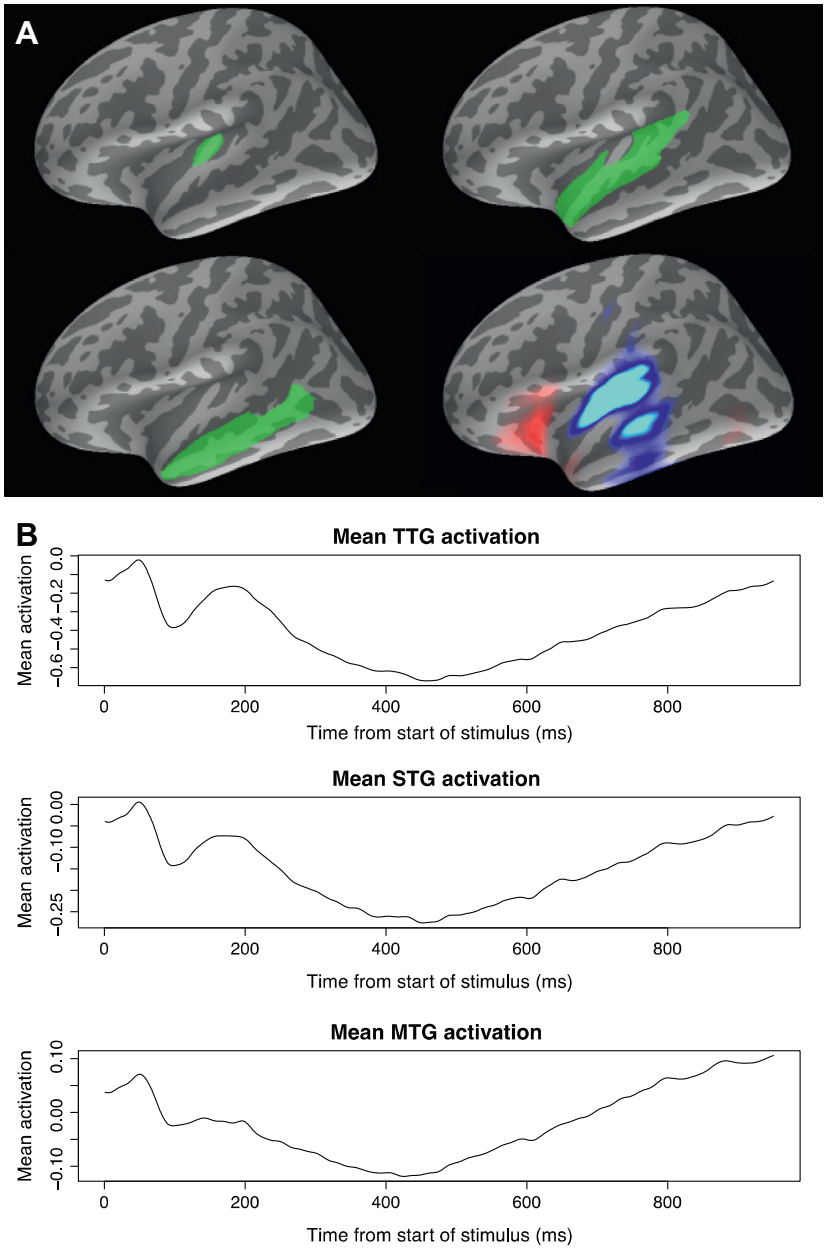

Fig. 1. (A) TTG (top left), STG (top right) and MTG (bottom left) ROIs, as well as grand average activation at approximately 500 ms (bottom right). Blue coloring signifies negative activation; red signifies positive activation. (B) Mean activation over time for TTG (top), STG (middle), and MTG (bottom) ROIs.

mixed-effects models. We used a limited random effect structure instead, with only random intercepts for items and subjects. Note that the suboptimal random effect structure does not result in an inflation of the Type I error, since our *p*-values are derived from the Monte Carle procedure outlined above.

The location of the largest significant cluster of *t*-values for the main effect of phoneme surprisal was used as the time window of interest for assessing the significance of the interaction between phoneme surprisal and morphological complexity. This interaction was calculated through model comparison between a model with only the two fixed effects and a model with the two fixed effects and their interaction. The log-likelihood ratio test produced a chi-squared statistic at every millisecond within the time window of interest. Multiple comparisons correction was then performed over that time window, comparing the largest cluster of significant chi-squared values in the same direction (with significance threshold set at $\chi^2 > 3.84$, $p < 0.05$ uncorrected for a chi-squared distribution with 1 degree of freedom) with the largest cluster in each of 1000 random permutations of the independent variable.

The millisecond-by-millisecond analysis may overstate the temporal resolution of MEG in terms of independent observations. In addition, the length of a typical phoneme is much longer than a millisecond (typically on the order of magnitude of tens of milliseconds). Of course, the beginning of each phoneme is not aligned across trials, which makes any significant temporal binning or smoothing likely to distort the results somewhat. Nevertheless, we replicated our major analyses at a much lower temporal resolution, by binning the data into 50 ms bins (1–50 ms, 51–100 ms, and so on), averaging the neural activity within each bin, and repeating the correlational analysis in these longer time units.

Performing a correlational analysis with a vector of stimulus variables aligned to temporal events (phonemes) necessitated determining the appropriate lag between the point at which a segment is presented and the point at which the effects of the

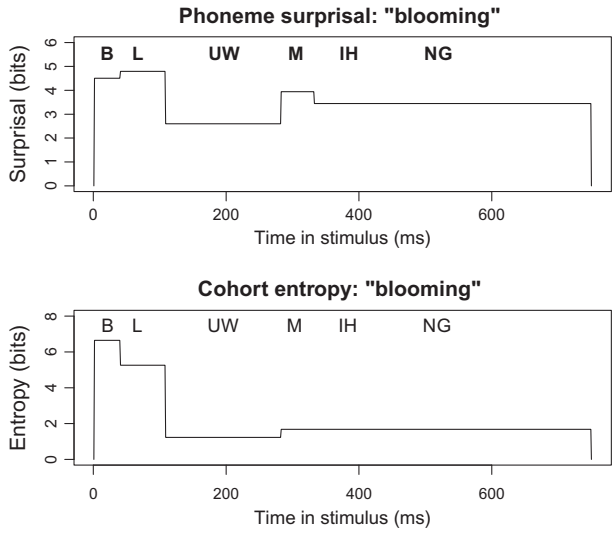

Fig. 2. Phoneme-by-phoneme surprisal and entropy values for the word "blooming".

surprisal or cohort entropy value appropriate for that segment are reflected in the neural activity. The prior literature does not afford a clear prediction for the precise latency of a surprisal effect. In the absence of a direct precedent, we decided to test three lags within the intermediate stage of auditory processing (100–200 ms) observed by Todorovic and de Lange (2012) to be facilitated by stimulus expectation. We tested the main effect of surprisal at lags of 100, 150 and 200 ms, and assessed the appropriateness of each lag based on the size of the largest cluster of consecutive significant *t*-values in the same direction (significance threshold set at $t > 1.96$, as before). For all three temporal ROIs, the 200 ms lag yielded the largest cluster of significant *t*-values (TTG: 100 ms, $\Sigma t = 676.4$, 150 ms, $\Sigma t = 799.1$, 200 ms, $\Sigma t = 822.3$; STG: 100 ms, $\Sigma t = 503.7$, 150 ms, $\Sigma t = 774.9$, 200 ms, $\Sigma t = 799.3$; MTG: 100 ms, $\Sigma t = 485.7$, 150 ms, $\Sigma t = 762.0$, 200 ms, $\Sigma t = 852.9$). We therefore used a lag of 200 ms for subsequent analyses. Note that for all lags and in all ROIs, the cluster of significant effects was around the end of the word.

4. Results

4.1. Behavioral

Both accuracy and reaction time (RT) showed significant effects of morphological complexity (RT, $p = .002$; accuracy, $p < .001$) and continuation surprisal (RT, $p < .001$; accuracy, $p < .001$), such that monomorphemic words were responded to more slowly and less accurately than bimorphemic words, and high surprisal items were responded to more slowly and less accurately than low surprisal items. No significant interactions emerged between morphological complexity and continuation surprisal (RT, $p = .124$; accuracy, $p = .539$): see Fig. 3 and Table 2.

There was no significant main effect of order of presentation (first or second, relative to pair counterpart) (RT, $p = .717$; accuracy, $p = .590$), nor was there a significant interaction between continuation surprisal and order (RT, $p = .587$; accuracy, $p = .898$). Reaction time did show a significant interaction between morphological complexity and order ($p = .013$), such that monomorphemic words were responded to more slowly when they were the second member of their pair to be presented, while bimorphemic words were responded to more quickly when they were presented second: see Fig. 4A. To explore the long-distance priming trend for bimorphemic words, we checked whether the distance within the experiment between two words that shared a stem affected the amount of facilitation, calculated as the difference between the RT to the second word and the RT to the first word. We fit a linear mixed-effects model with RT reduction as the response variable, distance as a fixed effect, and by-subject and by-stem random intercepts. The effect of distance on RT reduction was far from being significant ($t = 0.06$). In addition, there was no significant interaction between morphological complexity and order in the accuracy measure ($p = .351$). In summary, there was no evidence for an effect of distance on the amount of priming. However, our design was not optimized to detect such an effect, and the priming effect was small, so we would hesitate to conclude that distance does not affect priming.

A marginally significant three-way interaction emerged between morphological complexity, continuation surprisal, and order in RT ($p = .056$). Closer inspection reveals a significant interaction between morphological complexity and continuation surprisal on trials in which the stimulus is the first member of its pair to be presented ($p = .024$) while this interaction is not significant on second-presentation trials ($p = 1$): see Fig. 4B.

Table 2
Reaction time means and accuracy data for all conditions.

	High surprisal	Low surprisal
Reaction times		
Bimorphemic	1.19	1.11
Monomorphemic	1.24	1.19
Accuracy		
Bimorphemic	78.5%	89.5%
Monomorphemic	63.1%	77.7%

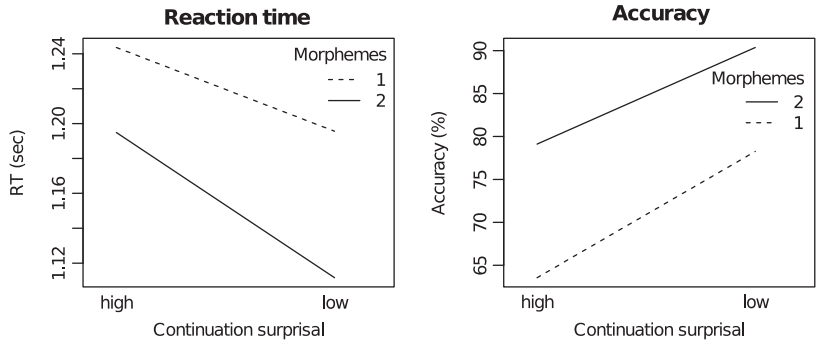

Fig. 3. Behavioral results. For both reaction time and accuracy, there was a main effect of both continuation surprisal and morphological complexity, and the interaction was not significant.

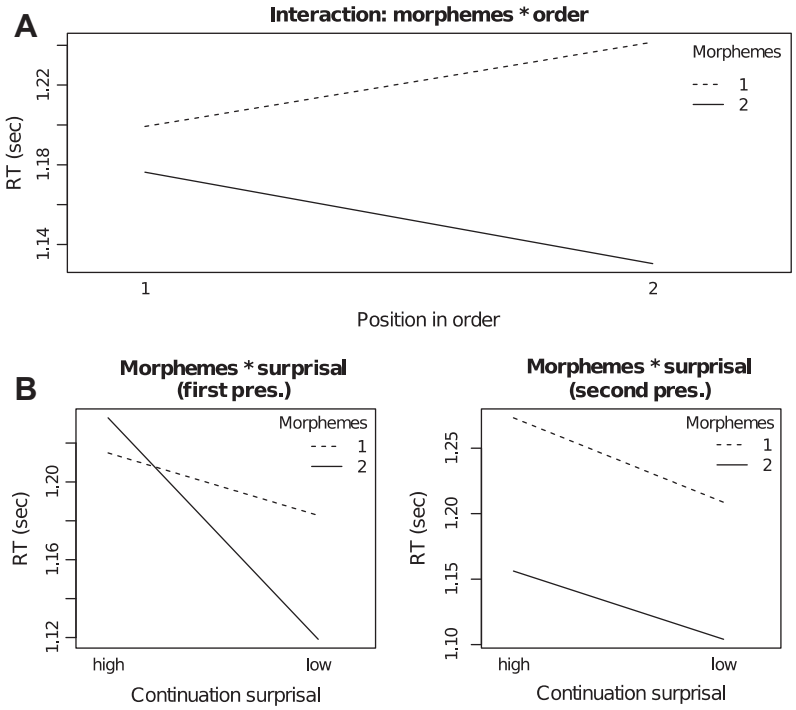

Fig. 4. (A) Interaction between morphological complexity and order. (B) Decomposition of the marginally significant three-way interaction between morphological complexity, continuation surprisal and order.

4.2. Neural: factorial

All three ROIs showed significant main effects of morphological complexity (TTG: $p = .001$; STG: $p = .006$; MTG: $p < .001$) and continuation surprisal (TTG: $p < .001$; STG: $p = .002$; MTG: $p = .004$). In the transverse and superior temporal gyri, where activation is in the negative direction, bimorphemic stimuli elicited less (negative) activation than did monomorphemic stimuli (facilitation), and high surprisal items elicited more negative activation than did low surprisal items (inhibition). In the middle temporal gyrus, activity was in the positive direction. In this region the pattern of effects was, in terms of absolute amplitude, reversed: bimorphemic stimuli elicited less positive activity, as did low surprisal stimuli. However, if we disregard zero as a fixed threshold, we can consider the pattern to be consistent in that monomorphemic and high surprisal items are eliciting more negative-going activity than bimorphemic and low surprisal items for all ROIs.

There was no significant main effect of order in any ROI (TTG: $p = .491$; STG: $p = 1$; MTG: $p = .784$). A significant interaction emerged between morphological complexity and continuation surprisal in all three ROIs (TTG: $p = .033$; STG: $p = .018$; MTG: $p = .006$), such that the increase in neural activity for high-surprisal continuations was greater for bimorphemic words than for monomorphemic words.

A significant interaction also emerged between continuation surprisal and order in transverse temporal ($p = .016$) and superior temporal ($p = .013$) ROIs, with a marginally significant interaction between these variables in the middle temporal ROI ($p = .095$), such that the effect of continuation surprisal was greater for items that were presented first relative to their pair counterpart. None of the ROIs showed a significant interaction between morphological complexity and order (TTG: $p = .401$; STG: $p = .433$; MTG: $p = .413$), and in none of the three ROIs was there a significant three-way interaction between morphological complexity, continuation surprisal, and order (TTG: $p = .447$; STG: $p = .263$; MTG: $p = 1$).

4.3. Neural: continuous

In the continuous analysis (employing, as discussed above, a lag of 200 ms), a highly significant main effect of phoneme surprisal emerged at the end of the stimulus in all three ROIs (TTG: 547–742 ms, $p < .001$: STG: 545–740 ms, $p < .001$; MTG: 542–740 ms, $p < .001$). Note that time windows listed here represent the time within the stimulus variable; adding 200 ms will yield the time-point in the brain activity, starting from the presentation of the word onset, with which the phoneme-specific variable is being correlated. For all three ROIs the direction of this correlation was negative, indicating that higher surprisal elicited more negative-going activity.

The interaction between phoneme surprisal and morphological complexity was also significant in all three ROIs (TTG: $p = .03$; STG: $p = .05$; MTG: $p < .001$), with the effect of phoneme surprisal being greater for bimorphemic than for monomorphemic words.

Fig. 5 shows correlation waves for the transverse temporal ROI.

A significant main effect of cohort entropy also appeared in the beginning of the word (TTG: 135–177 ms, $p = .004$ corrected over 1–300 ms time window): see Fig. 6A.

As mentioned in the Methods section, we repeated the major continuous analyses after averaging the neural activity in non-overlapping 50 ms bins (plots are available as part of the Online Supplementary materials). The pattern of results was similar across regions; we only report the result from the TTG for reasons of space. The qualitative shape of the binned plots was very similar to the millisecond-by-millisecond ones. The effect of surprisal differed between the monomorphemic and bimorphemic conditions: the bimorphemic words showed a strong surprisal effect from 550 ms onwards (peaking in $t = -5.3$ between 700 ms and 750 ms), whereas the monomorphemic words showed a much smaller surprisal effect (peaking in $t = -2.94$ between 650 ms and 700 ms). The late surprisal effect for bimorphemic words was very similar between the first and second presentation of the same stem. Somewhat counterintuitively, the smaller surprisal

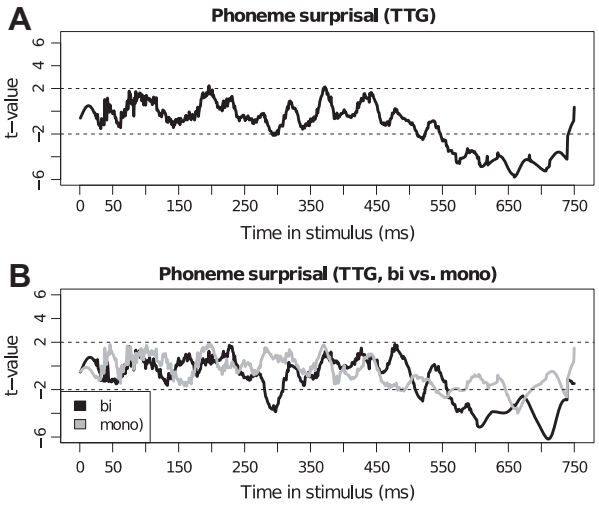

Fig. 5. Timepoint-by-timepoint correlations between phoneme surprisal and activity in transverse temporal gyrus, calculated by fitting linear mixed-effects models to neural activity at each timepoint, with subjects and items as random effects. A 200 ms lag is employed; the timepoints given on the *x*-axis above represent time within the stimulus variable—surprisal values in this time window have been correlated with neural activity in a time window of 200–950 ms (starting 200 ms after presentation of the stimulus onset).

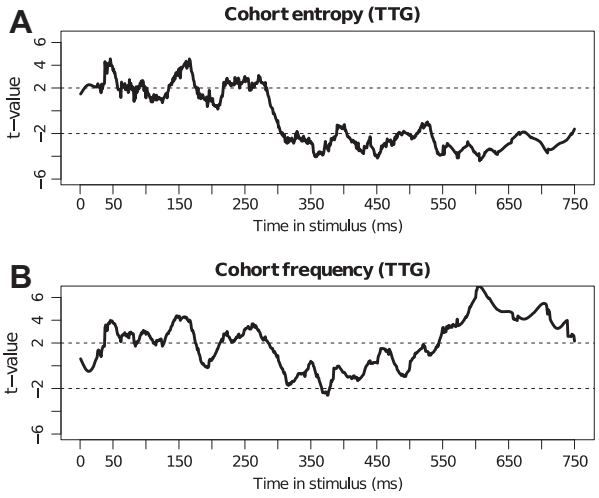

Fig. 6. Cohort entropy and cohort frequency correlations. See Fig. 5 for details.

effect at the end of the monomorphemic words seems to be due to the first presentation of the shared string prefix. Finally, the effect of entropy was positive in the first 250 ms (ranging between $t = 2.12$ and $t = 3.02$), and became negative from 350 ms through the end of the word (ranging between $t = -2.6$ and $t = -4.1$). The later negative effect was largely driven by the bimorphemic words.

Finally, we examine the effect of cohort frequency. Visual inspection suggests that this variable had a facilitatory effect in both early and late time windows; see Fig. 6B. The early effect of this variable is similar to the effect of cohort entropy early in the stimulus. We are unable at this point to distinguish between the effects of this variable and those of cohort entropy. The late effect is similar to the surprisal effect observed late in the stimulus, albeit in the opposite direction. This late effect is to be expected, due to the fact that typically only a single candidate remains in the cohort at word end, such that cohort frequency amounts to the frequency of that candidate. Low frequency words are likely to have high surprisal because they have higher frequency competitors. This explains why surprisal and cohort frequency have mirroring effects at word end.

5. Discussion

This study attempted to address a number of questions, both theoretical and methodological. On the methodological level, the study sought to explore the effectiveness of MEG for investigating the role of morphology in phoneme prediction during auditory word recognition, as well as the effectiveness of millisecond-by-millisecond correlation of MEG data with information-theoretic variables time-locked to phoneme boundaries within stimuli.

Our results suggest positive answers to both of these methodological questions. Significant and interpretable effects emerged in the analysis of the neural data, both in the results of the factorial design and in the results of the millisecond-by-millisecond analysis of continuous stimulus variables. This outcome supports the viability of these approaches in addressing theoretical questions. The variables we used to predict neural activity quantify the information-theoretic properties of the spoken word recognition task. Our results are therefore strictly at the computational level, in the sense of Marr (1982). In future work, our timepoint-by-timepoint correlational analysis could be used to predict brain activity from particular mechanistic models of speech recognition, such as TRACE (McClelland & Elman, 1986) or Shortlist B (Norris & McQueen, 2008).

On the theoretical level, the study addressed two main questions. First, based on findings supporting surprisal as a strong predictor of neural activity during processing of incoming phonemes in auditory word recognition (Gagnepain et al., 2012), we sought to investigate whether morphological structure would serve to enhance the phoneme prediction process suggested by these results (Balling & Baayen, 2012). Second, we sought to shed light on the nature of cohort competition in auditory word recognition—what effect, if any, would emerge in a millisecond-by-millisecond correlation of cohort entropy with neural activity in and around the auditory cortex? Would high entropy increase neural activity, as predicted by the cohort model (Marslen-Wilson, 1987); decrease neural activity, as predicted by the LEDP model we propose; or have no effect at all, as suggested by Gagnepain et al. (2012)?

For both categorical and continuous measures of surprisal, significant correlations emerged with activity in transverse and superior temporal gyri, adding to existing evidence in favor of a role of prediction in auditory word recognition. The direction of this effect accords with intuition: higher surprisal led to greater neural activation.

Significant interactions of morphological complexity with both measures of surprisal furthermore provide support for the hypothesis that morphological structure can enhance the capacity for phoneme prediction: in both measures, the strong main effect of surprisal was significantly stronger for bimorphemic words than for monomorphemic words. This suggests that the presence of internal structure does strengthen the capacity for prediction of upcoming phonemes in recognition of spoken words. It should be noted that this interaction did not emerge in reaction time or accuracy data, a discrepancy that suggests that the use of MEG techniques in investigating these questions may indeed reveal effects not apparent from behavioral results alone, at least in studies with fewer participants.

For the continuous measure of phoneme-by-phoneme cohort entropy, a significant facilitatory effect emerged in auditory cortex at the beginning of the stimulus—a result consistent with the predictions of LEDP models, on which high entropy leads to delay of prediction processes and therefore to less neural work. If we accept as a working hypothesis the idea that prediction is delayed under conditions of high entropy, this suggests a potential explanation

for the emergence of our surprisal effect only at the end of the stimulus: if in recognizing a spoken word, the brain suspends predictive processes until entropy reduces to a certain level, effects of error detection may not emerge until later in the stimulus, when cohort entropy is lower.

A clear main effect emerged for morphological complexity in both behavioral measures. This result could be due to processing advantage for complex words relative to simple words (Balling & Baayen, 2008; Ji et al., 2011), or due to the fact that continuation surprisal was not matched between monomorphemic and bimorphemic stimuli included in the study.

In the reaction time data, there was a significant interaction between order and morphological complexity, such that bimorphemic words were responded to more quickly when presented second within their pair, while monomorphemic pair members were responded to more slowly when presented second. This result suggests that there was indeed long-distance priming of shared morphological stems (cf. Kouider & Dupoux, 2009), whereas inhibition occurred between shared non-morphological string prefixes. This finding lends further support for morphological decomposition in auditory word recognition. A three-way interaction also emerged in the behavioral data between morphological complexity, continuation surprisal, and order. We have no ready explanation for why the two-way interaction observed between morphological complexity and continuation surprisal should emerge only on first-presentation trials.

The larger effect of prediction in bimorphemic words than in monomorphemic words shows that when predicting upcoming segments in complex words, participants use information that goes beyond the probability distribution of full word forms compatible with the initial string prefix. This information could only be derived from the stem, which must therefore be accessed before the end of the word. This argues against full listing models, in which morphologically complex words are representationally indistinguishable from monomorphemic words (Butterworth, 1983; Norris & McQueen, 2008), and in favor of models in which complex words are obligatorily decomposed into their subparts (Rastle, Davis, & New, 2004; Solomyak & Marantz, 2010; Taft & Forster, 1975). Future work should investigate the nature of the additional information provided by stems such as *kill-* (as in *killing* and *killer*) but not by meaningless string prefixes such as *brb-* (in *bourbon* and *burble*). For instance, participants could be making more reliable probabilistic predictions for stems with which they have limited past experience by using the overall frequencies of suffixes in the language. The prediction of the affix could also be informed by semantic properties of the stem. For example, stems associated with properties (*nice*, *long*) are much more likely to be followed by the superlative affix *-est* than stems associated with actions (*kill*, *eat*). These additional sources of information could be fruitfully modeled in a realistic generative model of English morphology.

6. Conclusion

Recent work in auditory word recognition has provided evidence for phoneme-level prediction occurring during the processing of a spoken word. Additional work points to an interaction of these predictive processes with a word's morphological structure. Drawing on these two lines of research, we investigated the role of morphological structure in phoneme prediction, while additionally exploring the precise manner in which the cohort of words consistent with the input affects the neural processes of auditory word comprehension. We found that morphological structure does indeed enhance phoneme prediction, leading to a larger effect of surprisal in morphologically complex words. We additionally found that higher cohort entropy leads to decreased neural activity.

The results of this study suggest that MEG is a viable tool for the investigation of morphological structure and phoneme prediction, and furthermore that it allows for successful millisecond-by-millisecond analysis of information-theoretic variables at the single trial level.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. BCS-0843969 and by the NYU Abu Dhabi Research Council under Grant G1001 from the NYUAD Institute, New York University Abu Dhabi. We thank Jeff Walker for technical assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bandl.2013.11.004.

References

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & Ochiai, M. (2001). Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. *IEEE Transactions on Applied Superconductivity, 11*, 669–672.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

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.

Balling, L. W., & Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes, 23*, 1159–1190.

Balling, L. W., & Baayen, R. H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes (Version 0.999999-0).

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.). *Language production* (Vol. 2). 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.

Dale, A. M., & Sereno, M. I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience, 5*, 162–176.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage, 31*, 968–980.

Fiorentino, R., & Poeppel, D. (2007a). Compound words and structure in the lexicon. *Language and Cognitive Processes, 22*, 953–1000.

Fiorentino, R., & Poeppel, D. (2007b). Processing of compound words: An MEG study. *Brain and Language, 103*, 18–19.

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology, 22*, 615–621.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Paper presented at the proceedings of the second meeting of the North American chapter of the association for computational linguistics on language technologies*.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics, 65*, 413.

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language, 65*, 406–430.

Kouider, S., & Dupoux, E. (2009). Episodic accessibility and morphological processing: Evidence from long-term auditory priming. *Acta Psychologica, 130*, 38–47.
Linzen, T., Marantz, A., & Pylkkänen, L. (2013). Syntactic context effects in visual word recognition: An MEG study. *The Mental Lexicon, 8*(2).
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*, 177–190.
Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71–102.
Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101*, 3–33.
Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29–63.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.
Norris, D., & McQueen, J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357–395.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.
Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
R Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
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.
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.
Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience, 22*, 2042–2057.
Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology, 57A*, 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.
Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *The Journal of Neuroscience, 32*, 13389–13395.
Wurm, L. H. (1997). Auditory processing of prefixed English words is both continuous and decompositional. *Journal of Memory and Language, 37*, 438–461.
Wurm, L. H., Ernestus, M. T., Schreuder, R., & Baayen, R. H. (2006). Dynamics of the auditory comprehension of prefixed words: Cohort entropies and conditional root uniqueness points. *The Mental Lexicon, 1*, 125–146.
Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America*, 123.