Syntactic context effects in visual word recognition

An MEG study

Tal Linzen[1], Alec Marantz[1, 2, 3], and Liina Pylkkänen[1, 2, 3]
[1] Department of Linguistics, New York University /
[2] Department of Psychology, New York University /
[3] NYUAD Institute, New York University Abu Dhabi

Words are typically encountered in the context of a sentence. Recent studies suggest that the contexts in which a word typically appears can affect the way it is recognized in isolation. We distinguish two types of context: collocational, involving specific lexical items, and syntactic, involving abstract syntactic structures. We investigate the effects of syntactic context using the distribution that verbs induce over the syntactic category of their complements (subcategorization frames). Magnetoencephalography (MEG) data was recorded while participants performed a lexical decision task. Verbs with low-entropy subcategorization distributions, in which most of the probability mass is concentrated on a handful of syntactic categories, elicited increased activity in the left anterior temporal lobe, a brain region associated with combinatory processing. Collocational context did not modulate neural activity, but had an effect on reaction times. These results indicate that both collocational and syntactic contextual factors affect word recognition, even in isolation.

Keywords: lexical processing, argument structure, subcategorization frames, syntax, information theory, magnetoencephalography, anterior temporal lobe

Research on isolated word recognition has uncovered an array of lexical, orthographic and semantic factors that affect the word recognition process (see Balota, Yap, & Cortese, 2006 for a review). Many of these factors are properties of the specific form being recognized, such as its frequency or its length. Some, however, are properties of the environments in which the recognized form is typically embedded. The role of context has been most thoroughly investigated in the case of word-internal structure (morphology). For instance, it has been shown

The Mental Lexicon 8:2 (2013), **117–139.** DOI 10.1075/ml.8.2.01lin
ISSN 1871–1340 / E-ISSN 1871–1375

that the speed of recognition of a monomorphemic word, such as *look*, is modulated by the number and frequency of words that are derived from it, in this case words like *looked* or *looking* (Baayen, Feldman, & Schreuder, 2006; Schreuder & Baayen, 1997).

Research in theoretical linguistics indicates that word-internal structure and word-external structure have much in common (Halle & Marantz, 1993, 1994). It is therefore natural to ask whether word-internal morphological context effects extend to word-external context: the words and structures that typically surround the recognized word in texts. This indeed turns out to be case. McDonald and Shillcock (2001b) found that words were responded to more slowly in isolation if they occurred in an unusual set of sentential contexts compared to the typical contexts in the language. More recently, contextual effects in lexical decision have been reported for the distribution of prepositions and adjectives preceding a noun: nouns that co-occur with an unusual set of prepositions take longer to recognize (Baayen, Milin, Djurdjević, Hendrix, & Marelli, 2011).

Two types of context could potentially have an effect on word recognition: collocational context and syntactic context. We define the collocational context of a word as comprising the specific lexical items (*blue*, *dog*) that tend to co-occur with the recognized word, regardless of the syntactic structure of the sentence. By contrast, the syntactic context of the word abstracts away from particular lexical items, focusing instead on the syntactic representation of the phrases that the word appears in: Is it usually modified by an adjective? Does it tend to be followed by a verb?

Previous studies of contextual effects in word recognition either have not attempted to dissociate collocational and syntactic context (McDonald & Shillcock, 2001b), or explicitly controlled for the syntactic environment in order to isolate the collocational context (Baayen, 2010; Baayen et al., 2011). Baayen et al. (2011), for example, limited their definition of context to the prepositions that preceded the noun being recognized, ignoring verbs, adjectives and other syntactic categories. It is therefore unknown whether syntactic context affects word recognition in isolation. Furthermore, contextual effects have only been reported for reaction times in behavioral experiments, and it is not known whether or how they modulate neural activity. This study fills both of these gaps, by measuring the effects of both syntactic and collocational context in a lexical decision task, while recording neural activity with magnetoencephalography (MEG).

To examine the effects of syntactic context, we exploit the fact that verbs vary in the types of syntactic phrases that they can take as their complements (their *subcategorization frames,* Chomsky, 1965). The verb *devour,* for example, is always followed by a noun phrase, *dine* is never followed by a noun phrase, and *eat* can appear either with a noun phrase or without one:

(1) Idiosyncratic subcategorization restrictions
  a. We ate.
  b. We ate the turkey.
  c. *We devoured.
  d. We devoured the turkey.
  e. We dined.
  f. *We dined the turkey.

Verbs also differ in the statistical distribution of their subcategorization frames (we abbreviate "subcategorization frames" as SCF in what follows). Both *accept* and *prove,* for example, can occur with either a noun phrase (NP) or a subordinate clause (SC), yet they differ in the relative frequencies of these two frames (Garnsey, Pearlmutter, Myers, & Lotocky, 1997):

(2) Verbs differ in their subcategorization frequencies
  a. P(NP | *accept*) = 0.98: He accepted the proposal.
  b. P(SC | *accept*) = 0.01: He accepted that he was wrong.
  c. P(NP | *prove*) = 0.23: He proved the claim.
  d. P(SC | *prove*) = 0.61: He proved that I was wrong.

Language comprehenders are sensitive to the distribution of verbs' subcategorization frames, and can use this information to make predictions about the syntactic category of a verb's complement during sentence processing (Arai & Keller, 2013; Garnsey et al., 1997; Wilson & Garnsey, 2009).

Subcategorization distributions must be represented as vectors, with each component of the vector corresponding to the probability of a given subcategorization frame. It is not immediately obvious how to relate these vectors to dependent measures such as reaction times or neural activity. We explore two different ways to summarize a verb's SCF distribution as a single quantity: first, the entropy of the distribution, and second, its relative entropy compared to the overall distribution of SCFs in English. The entropy of the SCF distribution is a combined measure of the number of possible frames and the extent to which their distribution is balanced, reflecting the degree of uncertainty about the syntactic category of the verb's complement (Moscoso del Prado Martín, Kostić, & Baayen, 2004). In the case of a verb that only allows one type of syntactic complement, there is no uncertainty at all as to the category of its complement, so the verb's SCF entropy is equal to 0. Among verbs with two possible SCFs, entropy will be higher when the two are equally likely. Conversely, when one of the frames is much more likely than the others, the entropy will be close to 0. Finally, a verb with three equally distributed frames will have higher entropy than a verb with only two equally distributed frames. Mathematically, if a verb X has n possible frames, and the probability of the i-th frame is p_i, its SCF entropy will be as follows:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Higher entropy has been associated with shorter lexical decision reaction times in the morphological domain. A word's morphological family is defined as the set of complex words in which the word appears as a constituent (Schreuder & Baayen, 1997). This set can be divided into an inflectional family (*thinks* or *thinking* for the base word *think*) and a derivational family (*thinker*, *rethink*). Both derivational family entropy and inflectional family entropy have facilitatory effects in visual lexical decision (Baayen et al., 2006). Likewise, high entropy over a word's morphological continuations facilitates reaction times in auditory lexical decision (Baayen, Wurm, & Aycock, 2007; Wurm, Ernestus, Schreuder, & Baayen, 2006).

In other domains, higher entropy may result in a processing slowdown. Studies of lexical ambiguity, for example, have shown that words that have two unrelated meanings take longer to respond to in a lexical decision task than do unambiguous words (Rodd, Gaskell, & Marslen-Wilson, 2002). Moreover, ambiguous words take longer to process when both of the meanings of the word are equally frequent (high meaning entropy) than when one of the meanings is dominant (low meaning entropy) (Duffy, Morris, & Rayner, 1988). These results have been attributed to competition between the two meanings of the ambiguous word: since the meanings inhibit each other, the semantic activation associated with the word does not reach the threshold required to make a lexical decision. Similar results have been reported in fMRI, where ambiguous words elicited increased BOLD signal in language areas (Rodd, Davis, & Johnsrude, 2005). Finally, in MEG, ambiguous words with high meaning entropy lead to increased neural activity compared to low meaning entropy words (Simon, Lewis, & Marantz, 2012). This suggests that competition leads to increased MEG signal, though in some studies competition modulated the latency rather than the amplitude of the neural response (Beretta, Fiorentino, & Poeppel, 2005; Pylkkänen, Feintuch, Hopkins, & Marantz, 2004).

In summary, higher SCF entropy may have either a facilitatory or inhibitory effect: facilitatory if subcategorization frames behave like morphological continuations, and inhibitory if they behave like meanings competing for selection.

SCF entropy is a property of the SCF distribution of a single verb. By contrast, relative SCF entropy quantifies the divergence between the verb's specific distribution and the distribution of the "average" English verb, obtained by collapsing together the SCF distributions of all verbs in the language. Assume, for example, that English only had three subcategorization frames: prepositional phrase (*talk about something*), noun phrase (*break something*) and the intransitive frame (*snore*), and that their overall probabilities in the language were 0.2, 0.4 and 0.4, respectively. If *break* has the SCF distribution (0.2, 0.45, 0.35), which is similar to the overall distribution, then its relative SCF entropy will be low. On the other hand, if *snore*, which is only compatible with the intransitive frame, has the distribution (0, 0, 1), then its relative entropy will be fairly high.

Formally, if the overall probability of frame *i* in the language is q_i, and its probability given the verb *X* is p_i, then the verb's SCF relative entropy is given by:

$$\sum_{i=1}^{n} p_i \log_2 \frac{p_i}{q_i}$$

In the morphological domain, nouns with high relative inflectional entropy are responded to more slowly in lexical decision (Milin, Djurdjević, & Moscoso del Prado Martín, 2009). High divergence between the collocational context of the word being recognized and the average collocational context has a similar inhibitory effect (Baayen et al., 2011; McDonald & Shillcock, 2001b). The prediction is therefore that higher SCF relative entropy will result in higher processing load.

The experiment described in this paper investigated the effect of SCF entropy and relative entropy on participants' neural activity while they were performing a lexical decision task on a set of verbs. Source localization techniques were used to determine which brain regions generated the observed MEG signal at each timepoint (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). We used a region of interest analysis to reduce the dimensionality of the data and incorporate the results of prior research. Our primary region of interest was the left lateral anterior temporal lobe (ATL), which we define as the parts of the superior temporal gyrus (STG) and middle temporal gyrus (MTG) that lie anterior to the auditory cortex (see Figure 1). ATL lesions are associated with impaired performance on basic morphosyntactic tasks (Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004). This region shows greater neural activity on two-word phrases compared to individual words (Bemis & Pylkkänen, 2011), and is consistently more active when processing sentences than when processing unstructured word lists, across techniques and modalities (Brennan & Pylkkänen, 2012; Humphries, Binder, Medler, & Liebenthal, 2006; Mazoyer et al., 1993). In MEG studies, combinatory effects in the ATL typically appear between 200 ms and 300 ms (Bemis & Pylkkänen, 2011, 2013; Brennan & Pylkkänen, 2012). Finally, multi-voxel pattern analysis of fMRI data has shown differential ATL activity patterns corresponding to different argument structure realizations of the same verb (Allen, Pereira, Botvinick, & Goldberg, 2012). The ATL therefore emerges as the region most likely to be sensitive to the properties of the immediate context of a word, and specifically to subcategorization frames.

We also report results from two other left-hemisphere language regions: the posterior temporal lobe (PTL), which includes the parts of the STG and MTG

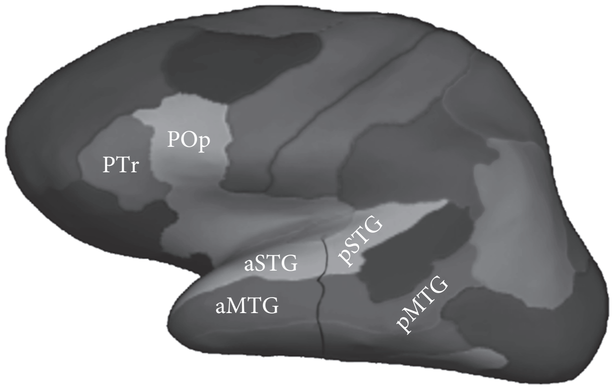

Figure 1. Language-related regions of interest in the left hemisphere. 1. Anterior temporal lobe (ATL): aMTG – anterior middle temporal gyrus; aSTG – anterior superior temporal gyrus. 2. Posterior temporal lobe (PTL): pSTG – posterior superior temporal gyrus; pMTG – posterior middle temporal gyrus. 3. Broca's area: PTr – inferior frontal gyrus, pars triangularis (Brodmann area 45); POp – inferior frontal gyrus, pars opercularis (Brodmann area 44).

posterior to the auditory cortex; and the posterior inferior frontal gyrus (Broca's area), which includes Brodmann areas 44 and 45. PTL lesions are associated with word-level impairments, and damage to Broca's area is associated with impaired processing of syntactically complex sentences (Dronkers et al., 2004). The functional dissociation of the three regions is not clear-cut, however. In addition to its involvement in composition of words into phrases, the ATL is recruited in lexical semantic processing (Rogers et al., 2004; Patterson, Nestor, & Rogers, 2007; Bi, Wei, Wu, Han, Jiang, & Caramazza, 2011). At the same time, PTL activity in fMRI is modulated by factors related to verb argument structure (Shetreet et al., 2007; Thompson et al., 2007) and by the size of syntactic constituents (Pallier et al., 2011). In light of this uncertainty as to the role of each region, we report results from all three regions, while focusing on the ATL as our primary region of interest.

Previous MEG studies indicate that the time window most likely to show combinatory effects in the ATL is 200–300 ms after the presentation of the verb (Bemis & Pylkkänen, 2011, 2013; Brennan & Pylkkänen, 2012). While some studies have demonstrated somewhat earlier sensitivity to syntactic properties of the stimulus (Hahne & Friederici, 1999; Pulvermüller, Shtyrov, Hasting, & Carlyon, 2008), these effects are likely to be related to form prediction (Dikker, Rabagliati, & Pylkkänen, 2009); there is no reason to assume that a verb's subcategorization distribution is reflected in its orthographic form. Both ATL and PTL show differential activity in sentences compared to word lists, both in the 200–300 ms and in the 300–400 ms time windows (Brennan & Pylkkänen, 2012). Lexical modulations are likely to be reflected in the M350/N400m component, which is evident in the 300–500 ms time region and shows sensitivity to frequency and lexical semantics (Halgren et al., 2002; Pylkkänen & Marantz, 2003; Pylkkänen, Llinas, & Murphy, 2006). This component encompasses most of the temporal lobe around 350 ms, and later spreads to the prefrontal cortex, including Broca's region (Halgren et al., 2002). Since there is some uncertainty as to the expected time window for each region, we analyze the activity in all three regions in three time windows: 200–300 ms, 300–400 ms and 400–500 ms.

Methods

Participants

18 participants (13 female) from New York City participated in the experiment. All subjects provided informed consent and were paid for their participation. Participants ranged in age from 20 to 44 (median 27). Two participants were excluded from analysis because of equipment failures. All subjects were right-handed (assessed using the Edinburgh Handedness Inventory; Oldfield, 1971) and were native speakers of English with normal vision.

Stimuli

The verbs analyzed in this paper were part of a larger lexical decision study. All of the words presented in the experiment were monomorphemic and monosyllabic 4-letter English words. Many 4-letter words can function either as a noun or as a verb (e.g. *lock*). For the purposes of the present study, we define a verb as a word that is used as a verb at least twice as frequently as it is used as a noun, based on the CELEX corpus (Baayen & Piepenbrock, 1995). None of the words presented in the experiment can be used in any other part of speech (adjective, adverb, etc.). In total, there were 189 verbs, out of a total of 750 words presented in the experiment.

Nonwords were selected from the ARC nonword database (Rastle, Harrington, & Coltheart, 2002) such that there was no significant difference in mean bigram letter frequency between the word and nonword stimuli. Participants performed a short practice block, which consisted of 7 trials, during which they received feedback. The stimuli were subsequently presented in 15 blocks of 100 trials, in pseudorandom order, such that in each sequence of ten trials, five were nonwords and five were words. Participants did not receive feedback on their answers.

Procedure

The experiment was conduced in the KIT/NYU facility at New York University. Prior to recording, the head shape of each participant was digitized to allow source localization and co-registration with structural MRIs (Fastscan; Polhemus, VT). We also digitized three fiducial points (the nasion and the left and right preauricular points) and the position of five coils, placed around the participants' 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 with a 157-channel axial gradiometer (Kanazawa Institute of Technology, Kanazawa, Japan).

Stimuli were presented using PsychToolBox (Brainard, 1997; Pelli, 1997) and projected onto a screen approximately 50 cm away from the participant. The stimuli were presented in white 30-point Courier font, on a gray background. Each trial began with a fixation cross (+) that appeared on the screen for 300 ms, followed by a blank screen for 300 ms, after which the stimulus was presented for 300 ms. Subjects then responded to the stimulus by pressing one of two buttons with the left hand to indicate whether they recognized the stimulus as a word. If the subject did not respond within 2 seconds, the next word was presented (timeouts of this sort only happened in 5 trials in one of the subjects). The inter-trial interval was randomly selected from values between 300 ms and 600 ms (in 50 ms increments).

Data Processing

The preprocessing and analysis of the MEG data closely followed the procedures of Solomyak and Marantz (2009, 2010). Environmental noise was removed from the data by regressing signals recorded from three orthogonally oriented magnetometers, placed 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). The data were then low-pass filtered at 40 Hz, resampled to 250 Hz to facilitate analysis, and high-pass filtered at 0.1 Hz. MEG channels in which there was no signal or excessive amounts of noise were interpolated from neighboring channels or rejected (at most 3 per subject). Trials in which at least one channel showed a peak-to-peak amplitude exceeding 3000 fT were rejected, as these amplitude values are likely to reflect blinks and noise artifacts (the number of rejected trials ranged from 74 to 497, mean 150, median 106).

Source Space Analysis

The MNE software package (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 et al., 1993). Current sources were modeled as three orthogonal dipoles spaced approximately 5 mm apart across the cortical surface (Dale et al., 2000), yielding approximately 2500 potential electrical sources per hemisphere. For nine of the 16 subjects, structural MRIs were available from previous experiments, and their cortical surfaces were reconstructed based on their structural MRIs using Freesurfer (Martinos center). For the 7 remaining subjects, a cortical surface based on an averaged brain provided by Freesurfer was used. The neuromagnetic data were co-registered with the structural MRI (9 subjects) or the averaged cortex (7 subjects) using MNE by first aligning the fiducial points, and then using an Iterative Closest Point algorithm to minimize the difference between the scalp and the points defining the head shape of each participant.

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). The middle and superior temporal gyri were manually divided into anterior and posterior portions, using the anterior edge of the transverse temporal gyrus as a dividing landmark (following Brennan & Pylkkänen, 2012). Signed activity was summed across each region of interest.

Lexical Variables

Subcategorization entropy and relative entropy were calculated according to the definitions given above. SCF frequencies were obtained from the automatically acquired subcategorization lexicon VALEX (Korhonen, Krymolowski, & Briscoe, 2006). We used the filtered and smoothed version of the lexicon (for details, see Korhonen et al., 2006). We used the ANLT subcategorization frame typology (Boguraev & Briscoe, 1987), which distinguishes 28 different frames in total.

As a control variable, we also replicated the contextual distinctiveness (CD) variable proposed by McDonald and Shillcock (2001a, 2001b). This variable measures the extent to which the collocational context of a word diverges from the average collocational context in the language. Consider, for example, the words *customer* and *lane*, which have identical frequency, yet differ in their CD: *lane* has a CD of approximately 1 bit, whereas *customer* has CD of approximately 0.5 bit. This reflects the fact that *lane* occurs in several common collocations (*fast lane, bike lane*), and therefore diverges more than *customer* from the average collocational context (McDonald & Shillcock, 2001a). Disfluencies such as *ah* and *erm* receive the lowest CD values (very close to 0 bits), because they can occur in any context. At the opposite end of the CD spectrum are words such as *amok*, which only occur in a specific collocation (*run amok* in this case).

Collocational context was calculated from the full British National Corpus. We removed punctuation, capitalization and sentence boundary information. The corpus was lemmatized using the WordNet lemmatizer included in the Python Natural Language Toolkit (NLTK: Bird, Klein, & Loper, 2009), taking into account the part-of-speech tagging provided with the corpus. Following McDonald and Shillcock (2001a), we removed frequent function words such as pronouns and common prepositions, based on NLTK's "stop word" list.[1] CD has two free parameters: the number of context words (i.e., the size of the vector representation), and the size of the window around each target word. We selected the values reported by McDonald and Shillcock (2001b) to be optimal for predicting lexical decision latencies: 500 content words (chosen to be the most frequent words in the corpus), and a window of five words on either side of the target word.

More formally, the prior distribution $P(c_i)$ is defined as the overall distribution of context words in the corpus, independent of the target word. The posterior distribution $P(c_i \mid w)$ is the distribution of context words around the target word w. The CD of a target word w is then defined as the relative entropy between the prior distribution $P(c_i)$ and the posterior distribution $P(c_i \mid w)$:

$$CD(w) = \sum_{i=1}^{n} P(c_i \mid w) \log_2 \frac{P(c_i \mid w)}{P(c_i)}$$

In addition to contextual distinctiveness, we controlled for the verbs' number of senses, as listed in WordNet (Miller, 1995). This was done to address the concern that a larger number of senses may be correlated with higher SCF entropy, if the different senses of the verb select different frames (Hare, McRae, & Elman, 2003; Roland & Jurafsky, 2002).

The final control variable was log-transformed frequency, as listed in the SUBTLEX database (Brysbaert & New, 2009). Log-transformed frequency was correlated with both number of senses ($r = 0.53$) and contextual distinctiveness ($r = -0.52$). To reduce collinearity, we regressed frequency out of both variables. The residualized variables were highly correlated with the original variables

Table 1. Pearson correlations between lexical predictors (after regressing log frequency out of contextual distinctiveness and number of senses).

	Entropy	Relative entropy	Frequency	CD	Number of senses
Entropy		−0.12	0.24	−0.07	−0.08
Relative entropy			0.15	−0.14	−0.15
Frequency				0	0
CD					−0.18

(number of senses: $r = 0.84$, contextual distinctiveness: $r = 0.85$), suggesting that they can be interpreted in the same way. Following this residualization step, correlations between variables were all mild ($|r| < 0.25$): see Table 1.

Results

Behavioral

Accuracy of responses ranged from 83.6% to 98.4% (mean 92.8%, median 93.5%). Mean reaction times ranged from 498 ms to 984 ms (mean 671 ms, median 644 ms).

Reaction times were log transformed and submitted to a linear mixed effects model (Bates, Maechler, & Bolker, 2012). We included a by-item intercept, a by-subject intercept, and a by-subject slope for the two subcategorization variables that were of main interest in this paper. We did not include by-subject slopes for the control variables, because models with more elaborate random effect structures often did not converge.

Table 2 shows the model fitted to the verb trials. The p-values for fixed effects here and in what follows are derived using model comparison: the full model is compared to a model with the same random effect structure but without the predictor for which the p-value is being calculated. The difference in log likelihood between the partial and full model is then evaluated using the χ^2 approximation: $-2LL \sim \chi^2\,(1)$.

Only frequency had a significant effect on reaction times in verb trials. The effect went in the expected direction: frequent verbs were responded to faster. SCF entropy did not affect reaction times. SCF relative entropy and contextual distinctiveness, which measure the degree to which the word's context deviates from the average context in the language (syntactic and collocational context respectively), both showed non-significant trends in the expected direction: verbs with unusual SCF distributions or unusual collocational contexts elicited numerically longer reaction times ($p_{\chi^2} = 0.12$ and $p_{\chi^2} = 0.11$, respectively).

Table 2. Reaction times, verbs trials only (linear mixed-effects model).

Predictor	Estimate	Std. error	t-value	p-value (χ^2)
SCF entropy	0.005	0.018	0.03	0.92
SCF relative entropy	0.018	0.017	1.11	0.12
Frequency	–0.018	0.003	–5.6	< 0.001
Contextual distinctiveness	0.04	0.03	1.38	0.11
Number of senses	0.001	0.0008	1.25	0.18

Table 3. Reaction times, all words (linear mixed-effects model).

Predictor	Estimate	Std. error	t-value	p-value (χ^2)
Frequency	–0.022	0.002	–12.4	< 0.001
Contextual distinctiveness	0.028	0.01	2.55	0.004
Number of senses	0.006	0.0006	0.89	0.247

An additional model was fitted to the entire data set, including nouns (Table 3). The subcategorization variables were excluded from the analysis, since they were not applicable to nouns. By-subject random slopes for frequency and contextual distinctiveness were added to the model. The effect of frequency was again highly significant ($p_{\chi^2} < 0.001$). Contextual distinctiveness also reached significance in the larger data set ($p_{\chi^2} = 0.004$): Words with more unusual contexts were responded to more slowly, replicating McDonald and Shillcock (2001b).

MEG

We analyzed the total neural activity in each of the three left-hemisphere regions of interest — anterior temporal lobe (ATL), posterior temporal lobe (PTL) and Broca's area. Activity was averaged in three 100 ms time windows: 200–300 ms, 300–400 ms and 400–500 ms after stimulus onset, based on the time course of effects in previous MEG studies (Bemis & Pylkkänen, 2011; Brennan & Pylkkänen, 2012). A linear mixed-effects model was fitted to each time window in each region of interest.

Verbs

SCF entropy was negatively correlated with ATL activity between 200 ms and 300 ms ($\beta = -0.063$, $p_{\chi^2} = 0.009$): higher entropy verbs elicited less ATL activity. This correlation was weaker and no longer significant between 300 ms and 400 ms ($\beta = -0.045$, $p_{\chi^2} = 0.11$). Also between 300 ms and 400 ms, there was a marginal positive correlation with subcategorization relative entropy such that higher

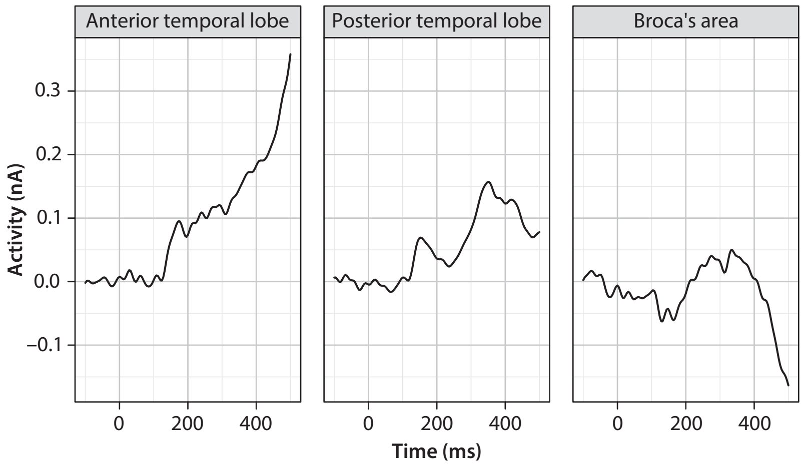

Figure 2. Verb trials: Grand mean of neural activity, across subjects.

relative entropy resulted in increased activity ($\beta = 0.052$, $p_{\chi^2} = 0.06$). The two subdivisions of the ATL, the aMTG and the aSTG, did not differ in the qualitative pattern of results.

PTL activity showed no effect of SCF entropy, and a marginal effect of SCF relative entropy between 300 ms and 400 ms ($\beta = -0.035$, $p_{\chi^2} = 0.08$). This marginal effect went in the opposite direction from the ATL effect: higher relative entropy resulted in less PTL activity. There was additionally a marginal effect of frequency, such that higher frequency words elicited more PTL activity (300–400 ms: $\beta = 0.008$, $p_{\chi^2} = 0.06$, 400–500 ms: $\beta = 0.009$, $p_{\chi^2} = 0.08$). An inspection of the two subdivisions of the PTL showed that the relative entropy effect was primarily in the pMTG (300–400 ms: $\beta = -0.052$, $p_{\chi^2} = 0.029$), and the frequency effect was stronger in the pSTG (300–400 ms: $\beta = 0.01$, $p_{\chi^2} = 0.03$; 400–500 ms: $\beta = 0.012$, $p_{\chi^2} = 0.03$).

Neither of the SCF variables had a significant effect in Broca's area. Overall, the SCF entropy effect was specific to the ATL: none of the other regions showed a significant effect of this variable, and an ANOVA revealed a significant interaction between SCF entropy and region ($p_{\chi^2} = 0.012$).

All Words

We also analyzed the entire set of words, including nouns. Due to model convergence issues, we fitted separate models for each of the three relevant variables — frequency, contextual distinctiveness and number of senses — each with the relevant by-subject slope. Since the variables were decorrelated from each other, the results should be very similar to a model containing all of the variables.

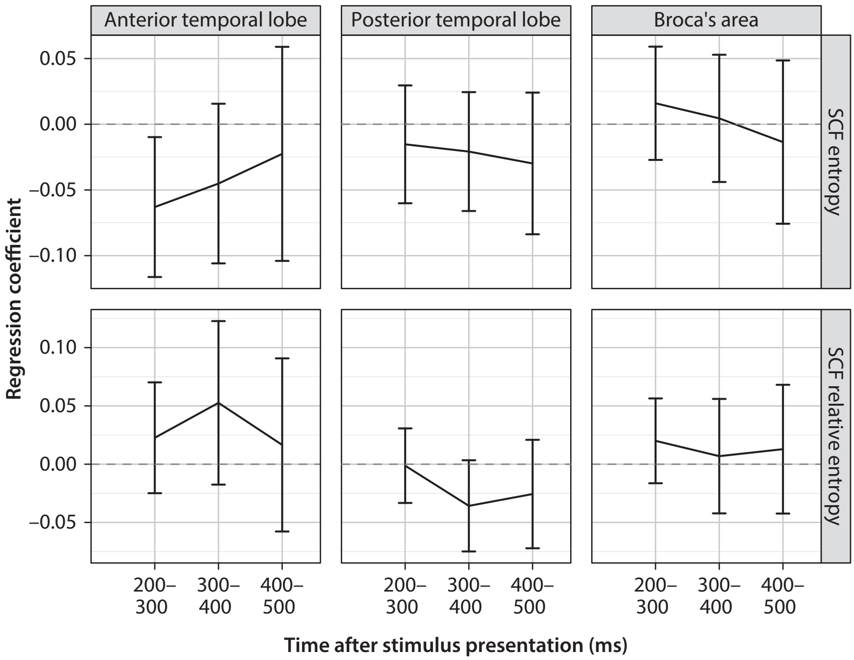

Figure 3. Verb trials: Regression coefficients for subcategorization variables. Approximate confidence intervals extend to twice the standard error of the regression coefficients in each direction.

There were no neural effects of contextual distinctiveness. Number of senses had a marginal negative effect in Broca's area between 300 ms and 400 ms (β = −0.002, p_{χ^2} = 0.05). Frequency, on the other hand, had a significant effect between 300 ms and 400 ms after stimulus presentation, both in the PTL (β = 0.008, p_{χ^2} = 0.0007) and in Broca's area (β = 0.009, p_{χ^2} = 0.0005). In both areas, more frequent words led to increased activity. The effect of frequency between 300 ms and 400 ms in the ATL did not reach significance. However, it trended in the same direction as the frequency effect in other regions (β = 0.005, p_{χ^2} = 0.08), and a direct comparison between the effects of frequency in the ATL and the PTL did not reveal a significant frequency by region interaction (p_{χ^2} = 0.4).

Discussion

The present study showed that the typical syntactic context of a word — specifically, the distribution of a verb's subcategorization frames — affects activity in the left anterior temporal lobe during the recognition of the word. Verbs with high

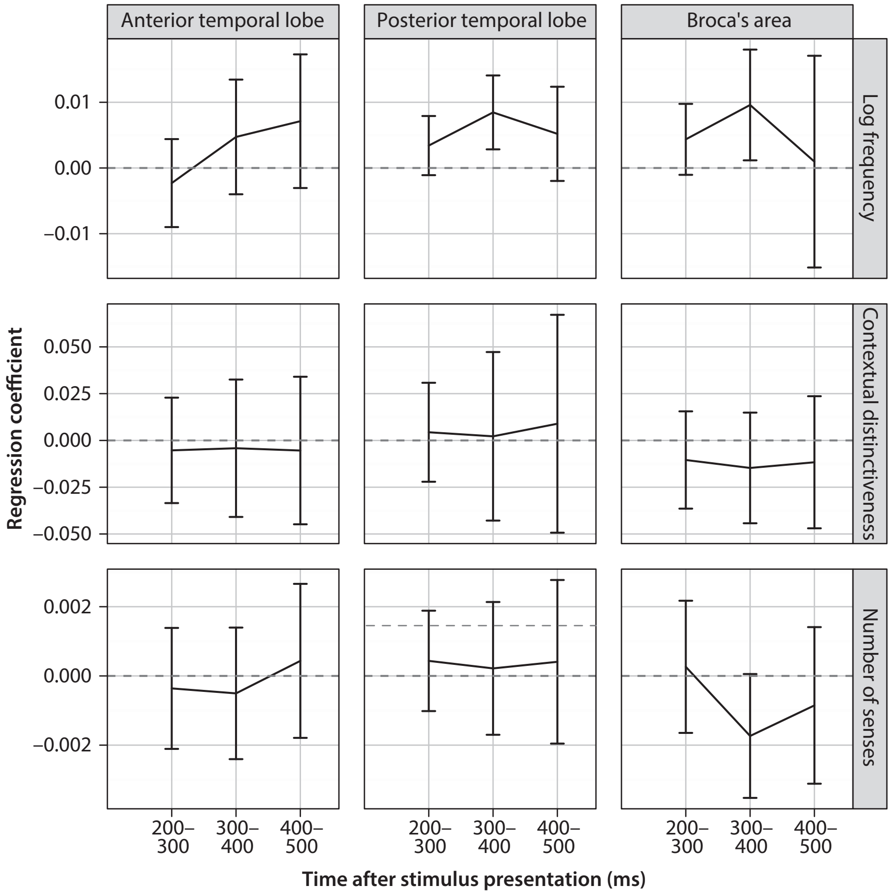

Figure 4. All trials: Regression coefficients for control variables. Approximate confidence intervals extend to twice the standard error of the regression coefficients in each direction.

subcategorization frame entropy elicited less ATL activity than low entropy verbs between 200 ms and 300 ms, the time window associated with composition of phrases from individual words in the same region (Bemis & Pylkkänen, 2011, 2013). This effect could not be attributed to any of the control variables we investigated: word frequency, collocational context or number of senses.

The reduction in neural activity in response to increased entropy over continuations is in line with previous behavioral findings in the morphological domain. Morphological family entropy has a facilitatory effect in visual lexical decision (Baayen et al., 2006). Likewise, high morphological continuation entropy leads to shorter reaction times in auditory lexical decision (Baayen et al., 2007; Wurm et al., 2006). These parallels between word-internal morphological continuations and word-external syntactic continuations are expected in light of the similarities between word-internal and word-external context, as suggested by Halle and Marantz (1993) and Baayen et al. (2011) (for a different view, see Cappelle, Shtyrov, & Pulvermüller, 2010; Pulvermüller, Cappelle, & Shtyrov, 2013). If competition entails more activity, as in the case of competition between the two meanings of an ambiguous word (Simon et al., 2012), the present pattern of results is at odds with an account whereby all possible continuations are activated and compete for selection: a competition account would predict higher activity for a larger number of continuations.

In addition to the entropy effect, there was a marginal effect of relative entropy in the ATL between 300 ms and 400 ms, such that higher relative SCF entropy caused an increase in activity. SCF relative entropy measures the divergence between the SCF distribution of the recognized verb and the average SCF distribution in the language. This result mirrors the effect of relative inflectional entropy: Serbian masculine words whose distribution over inflected forms diverges from that of the average masculine word take longer to recognize (Milin et al., 2009; Baayen et al., 2011). The average SCF distribution is likely to serve as the reader's prior distribution, in advance of seeing the specific verb; the verb's specific distribution is more surprising the larger the divergence between this specific distribution and the prior distribution. Increased neural activity in response to stimuli with unexpected properties ties in with theories of predictive coding (Rao & Ballard, 1999; Friston, 2005), according to which neural responses reflect the degree to which the incoming stimulus forces an update in the reader's probabilistic expectations. Specifically in language comprehension, numerous studies have shown that words with unexpected properties elicit a stronger N400 component than expected words (see Kutas & Federmeier, 2011, for a recent review). A potential interpretation of the later latency of the marginal relative SCF entropy effect (300–400 ms) compared to the SCF entropy effect (200–300 ms) may be that the relative entropy effect is part of the N400 surprise response, whereas the entropy effect reflects an earlier combinatorial component (Bemis & Pylkkänen, 2011).

Although the ATL was our main region of interest, we tested two additional left-hemisphere regions as controls: the posterior temporal lobe (PTL) and Broca's area. The two control regions did not show an SCF entropy effect. At the same time, the two regions showed significant frequency effects between 300 ms and 400 ms, in contrast with the ATL. Increased activity in response to frequent words has been observed in previous studies (Brennan et al., 2012; Solomyak & Marantz, 2010; Yarkoni, Speer, Balota, McAcoy, & Zacks, 2008). The difference between the ATL and the PTL in sensitivity to these two lexical variables is in line with models that ascribe a compositional role to the ATL and a lexical role to the PTL (Hickok & Poeppel, 2007). This potential dissociation between the anterior and posterior parts of the temporal lobe should be interpreted with caution, however. While an interaction test showed that the two control regions differed significantly from the ATL in the magnitude of the subcategorization entropy effect, a similar test did not show a significant difference in the effect of frequency. Furthermore, there was a hint of an effect of subcategorization relative entropy in the PTL between 300 ms and 400 ms, and fMRI studies have found effects of verb argument structure in posterior temporal regions (Shetreet, Palti, Friedmann, & Hadar, 2010; Shetreet, Friedmann, & Hadar, 2007).

The facilitatory effect of continuation entropy may reflect a conservative prediction strategy, whereby a potential continuation is only preactivated if the prediction has a high probability of being fulfilled. Dikker and Pylkkänen (in press), for example, observe increased neural activity in constraining contexts relative to non-constraining contexts, in advance of the presentation of a predictable word. This interpretation of the neural effect of uncertainty over continuations predicts that the effect should not be specific to individual word recognition tasks, and should also show up when the verb is embedded in a sentence. In a sentential context, if a verb licenses a specific syntactic prediction, that syntactic category is immediately activated. Otherwise, the parser waits for the complement to establish the syntactic structure of the verb phrase.

An alternative interpretation of the results is that words that are encountered in multiple different contexts are more robustly represented in the brain than words that always occur in one specific context (Adelman, Brown, & Quesada, 2006). In other words, if a verb is strongly associated with a specific syntactic context (in our study, has low SCF entropy), seeing it outside of this context will hinder its recognition. This interpretation would not be compatible with a strong functional interpretation of the anatomical pattern of results: subcategorization entropy only modulated activity in the anterior part of the temporal lobe, and not in the posterior part, which is most strongly associated with the retrieval of lexical representations. In future research, the two interpretations could be distinguished in a head-final language, in which the verb does not typically predict the syntactic structure of its arguments. The conservative prediction hypothesis predicts that the subcategorization entropy effect will be weaker or nonexistent in head-final languages, whereas the robustness of representation hypothesis does not predict any difference between head-final and head-initial languages in this respect.

The results of this study suggest that syntactic information associated with a word is accessed even when structure building is not required by the experimental task. Since words are almost always encountered in context, it is not surprising that the activation of syntactic information is automatic. The subcategorization biases of a single verb prime have been shown to affect subjects' productions in a syntactic priming paradigm (Melinger & Dobel, 2005), and argument-structure based distinctions modulate brain activity in single-word lexical decision (Thompson et al., 2007).

Neither of our subcategorization frame variables affected lexical decision reaction times. This may be due to insufficient power: while there was a robust effect of contextual distinctiveness on the full set of words (including nouns), it did not reach significance on verbs. This indicates that our set of verbs may have been too small to detect behavioral effects of contextual variables. Alternatively, it is possible that syntactic information, while automatically activated upon reading a word, is not recruited to make lexical decisions. Previous studies have also failed to detect effects of the number of subcategorization frames on reaction times in lexical decision (Schmauder, 1991; Shapiro, Zurif, & Grimshaw, 1987). It is not clear why that should be the case, and further research would be necessary to address this point.

Subcategorization frame entropy is not a perfect measure of a verb's typical syntactic context. Verbs also differ in their likelihood of occurring with different classes of adjuncts (*yesterday*, *with a knife).* Furthermore, verbs, and words in general, may vary in the larger syntactic structures in which they tend to appear — e.g., questions, ellipsis constructions and so on. It has been suggested that subcategorization frames are more tightly connected to the lexical representation of a verb than are other types of syntactic contexts (Boland, 2005). It thus remains an open question whether all types of syntactic contexts will have the effect on word recognition observed in this paper.

In conclusion, this study supports the role of a word's typical contexts on its recognition in isolation. In particular, we found an effect of syntactic context, abstracting away from the specific lexical items that typically co-occur with the word (collocational context). We found that verbs that tend to appear with a larger variety of syntactic arguments, as measured by subcategorization frame entropy, elicited less neural activity in the left anterior temporal lobe. This is the opposite of what would be predicted by a competition account, under which the activation of multiple possible frames should lead to higher overall activity. We hypothesized that this effect reflects a conservative prediction strategy: a syntactic frame is only preactivated when the verb licenses a specific prediction.

Note

1. A reviewer points out that in following McDonald and Shillcock's (2001a) methodology in computing Contextual Distinctiveness, we remove from consideration close connections between verbs and prepositions (e.g., *depend on*) and verbs and particles (e.g., *look up*). Since

in these cases the verb predicts a particular preposition or particle in its immediate syntactic environment, these dependencies may fall under the same kind of knowledge for verbs as subcategorization frames. Alternatively, or they might pattern with knowledge of collocational context. The nature of these dependencies should be the topic of further research.

References

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

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*, 814–823.

Allen, K., Pereira, F., Botvinick, M., & Goldberg, A. E. (2012). Distinguishing grammatical constructions with fMRI pattern analysis. *Brain and Language*, *123*, 174–182.

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*, 525–560.

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*, 436–461.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *55*, 290–313.

Baayen, R. H., Milin, P., Djurdjević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*, 438–481.

Baayen, R. H., & Piepenbrock, R. (1995). Th*e CELEX lexical database (Release 2) [CD-ROM]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, *2*, 419–463.

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 285–375). Amsterdam: Academic Press.

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes.* Retrieved from http://cran.r-project.org/package=lme4, R package version 0.999999-0.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience*, *31*, 2801–2814.

Bemis, D. K., & Pylkkänen, L. (2013). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, *23*, 1859–1873.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research*, *24*, 57–65.

Bi, Y., Wei, T., Wu, C., Han, Z., Jiang, T., & Caramazza, A. (2011). The role of the left anterior temporal lobe in language processing revisited: Evidence from an individual with ATL resection. *Cortex*, *47*, 575–587.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python.* Sebastopol, CA: O'Reilly Media.

Boguraev, B., & Briscoe, T. (1987). Large lexicons for natural language processing: Utilising the grammar coding system of LDOCE. *Computational Linguistics*, *13*, 203–218.

Boland, J. E. (2005). Visual arguments. *Cognition*, *95*, 237–274.

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

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, *120*, 163–173.

Brennan, J., & Pylkkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *Neuroimage*, *60*, 1139–1148.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera 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.

Cappelle, B., Shtyrov, Y., & Pulvermüller, F. (2010). *Heating up* or *cooling up* the brain: MEG evidence that phrasal verbs are lexical units. *Brain and Language*, *115*, 189–201.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT press.

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*, 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., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*, 968–980.

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition*, *110*, 293–321.

Dikker, S., & Pylkkänen, L. (In press). Predicting language: MEG evidence for lexical preactivation. *Brain and Language*. http://dx.doi.org/10.1016/j.bandl.2012.08.004.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*, 145–177.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429–446.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, *360*, 815–836.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*, 58–93.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*, 194–205.

Halgren, E., Dhond, R., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like MEG responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*, *17*, 1101–1116.

Halle, M., & Marantz, A. (1993). Distributed Morphology and the pieces of inflection. In K. L. Hale & S. J. Keyser (Eds.), *The view from building 20* (pp. 111–176). Cambridge, MA: MIT press.

Halle, M., & Marantz, A. (1994). Some key features of Distributed Morphology. *MIT Working Papers in Linguistics*, *21*, 275–288.

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*, 413–497.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, *48*, 281–303.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, *18*, 665–679.

Korhonen, A., Krymolowski, Y., & Briscoe, T. (2006). A large subcategorization lexicon for natural language processing applications. In *Proceedings of the 5th international conference on language resources and evaluation*. Genova, Italy.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O, Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*, 467–479.

McDonald, S. A., & Shillcock, R. C. (2001a). *Contextual distinctiveness: A new lexical property computed from large corpora*. Informatics Research Report EDI-INF-RR-0042, University of Edinburgh.

McDonald, S. A., & Shillcock, R. C. (2001b). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*, 295–323.

Melinger, A., & Dobel, C. (2005). Lexically-driven syntactic priming. *Cognition*, *98*, B11–B20.

Milin, P., Djurdjević, D. F., & Moscoso del Prado Martín, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, *60*, 50–64.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*, 39–41.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Pallier, C., Devauchelle, A. D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, *108*, 2522–2527.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

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

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987.

Pulvermüller, F., Shtyrov, Y., Hasting, A. S., & Carlyon, R. P. (2008). Syntax as a reflex: neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language*, *104*(3), 244–253.

Pulvermüller, F., Cappelle, B., & Shtyrov, Y. (2013). Brain basis of meaning, words, constructions, and grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 397–417). Oxford: OUP.

Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: An MEG study. *Cognition*, *91*, B35–B45.

Pylkkänen, L., Llinas, R., & Murphy, G. (2006). Representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, *18*, 1–13.

Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, *7*, 187–189.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology Section A*, *55*, 1339–1362.

Rao, R. P. N., & Ballard, D. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*, 245–266.

Rodd, J., Davis, M., & Johnsrude, I. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*, 1261–1269.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*, 205–235.

Roland, D., & Jurafsky, D. (2002). Verb sense and verb subcategorization probabilities. In P. Merlo & S. Stevenson (Eds.), *The lexical basis of sentence processing: Formal, computational, and experimental issues* (pp. 325–346). Amsterdam: John Benjamins.

Schmauder, A. R. (1991). Argument structure frames: A lexical complexity metric? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 49–65.

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

Shapiro, L. P., Zurif, E., & Grimshaw, J. (1987). Sentence processing and the mental representation of verbs. *Cognition*, *27*, 219–246.

Shetreet, E., Friedmann, N., & Hadar, U. (2010). The neural correlates of linguistic distinctions: Unaccusative and unergative verbs. *Journal of Cognitive Neuroscience*, *22*, 2306–2315.

Shetreet, E., Palti, D., Friedmann, N., & Hadar, U. (2007). Cortical representation of verb processing in sentence comprehension: Number of complements, subcategorization, and thematic frames. *Cerebral Cortex*, *17*, 1958–1969.

Simon, D. A., Lewis, G., & Marantz, A. (2012). Disambiguating form and lexical frequency effects in MEG responses using homonyms. *Language and Cognitive Processes*, *27*, 275–287.

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.
Thompson, C. K., Bonakdarpour, B., Fix, S. C., Blumenfeld, H. K., Parrish, T. B., Gitelman, D. R., & Mesulam, M. M. (2007). Neural correlates of verb argument structure processing. *Journal of Cognitive Neuroscience*, *19*, 1753–1767.
Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language*, *60*, 368–392.
Wurm, L. H., Ernestus, M. T. C., 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.
Yarkoni, T., Speer, N. K., Balota, D. A., McAvoy, M. P., & Zacks, J. M. (2008). Pictures of a thousand words: Investigating the neural mechanisms of reading with extremely rapid event-related fMRI. *NeuroImage*, *42*, 973–987.

Corresponding Address

Tal Linzen
Department of Linguistics, New York University
10 Washington Place, New York, NY 10003
Tel: 1-212-998-8371

linzen@nyu.edu