Neural representation of minimal syntactic units

Charles Roger Bradley (bradley4@purdue.edu)
Purdue University, Linguistics Program, 100 N. University Street, Beering Hall
West Lafayette, Indiana 47907-2098 USA

Jeffrey Mark Siskind (qobi@purdue.edu)
Purdue University, School of Electrical and Computer Engineering, 465 Northwestern Avenue, Electrical Engineering Building
West Lafayette, Indiana 47907-2035 USA

Ronnie B. Wilbur (wilbur@purdue.edu)
Purdue University, Linguistics Program and Speech, Language, and Hearing Sciences, 715 Clinic Drive, Lyles-Porter Hall
West Lafayette, Indiana 47907-2122 USA

Background

How and why humans can create and understand an infinite number of novel sentence remains a linguistic mystery, especially given the number and diversity of languages. Despite the apparent complexity of the problem, Generative linguists claim that the answer can be reduced to a single, simply defined and simply implemented function, Merge (Chomsky, 1995). Merge takes two syntactic objects (*e.g.*, words) and joins them, forming a larger syntactic object, called a constituent. Merge operates iteratively, either applying to yet unmerged items or to the product of previous applications of Merge, building hierarchical, recursive structures (Chomsky, 2001). Merge is argued to be category-neutral, such that the derivation of noun phrases (NPs) is identical to the derivation of verb phrases (VPs), and so on. Little evidence comes from studies on language processing and brain function. Instead, studies of neural processing of syntax has focused on large-scale sentential phenomena, involving several applications of Merge, and further processing requirements.

Studies implicating Merge indicate that constituents of different sizes elicit activation in left inferior frontal gyrus (LIFG) during comprehension (Pallier, Devauchelle, & Dehaene, 2011) and production (Indefrey et al., 2001; Indefrey, Hellwig, Herzog, Seitz, & Hagoort, 2004). However, direct evidence for Merge is scarce: Bemis and Pylkkänen (2011) looked at effects of combining adjectives with nouns to form NPs, finding significant activation in left anterior temporal lobe (LATL). However, their design made it difficult to tease apart contributions of syntactic operations (*i.e.*, Merge) and semantic composition (*i.e.*, how meanings of the words combine).

Zaccarella and Friederici (2015) report that Merge is processed in a small cluster within BA44. They use minimal compositions (*i.e.*, two-word phrases) and two-word word lists, as well as pseudo-words to avoid effects of semantic composition (Bemis & Pylkkänen, 2011; Humphries, Binder, Medler, & Liebenthal, 2006; Pallier et al., 2011). However, they only investigated NPs. Here, we attempt to replicate Z&F's findings across 3 further categories: Verb Phrases (VPs), Adjective Phrases (APs), and Prepositional Phrases (PPs) (:XPs). However, our aim is not just to see whether constituent XPs all engage the same (sub)regions of LIFG, but also to address whether the processing of each category is identical. That is, Merge is theoretically not category-specific, so we should not find category-specific patterns of activation. While we test Z&F's claim that Merge is localized in Broca's Area, our focus here is not to expand knowledge of the region, but a finer, neurological characterization of constituency distinctions and lexical categories.

Questions

We ask the following:

1. Is there a pattern of neural activation that characterizes constituents *vs.* non-constituents?
2. Is that pattern the same across lexical categories?

We predict that the answer to both is *yes*, given the robustness of Merge in the modeling of data from hundreds of languages. However, we are presupposing that lexical category is relevant at the neural level. So, we additionally ask:

3. Can we distinguish between lexical categories, independent of lexical items?

Taken together, we can show that lexical category, while neurologically real, is not relevant to constituent building.

Method

We employed functional magnetic resonance imaging (fMRI) analyzed with multi-voxel pattern analysis (MVPA).

Subjects Data was acquired for a single adult female, a student at Purdue University. Informed consent was obtained. All protocols, experiments, and analyses were approved by the Institutional Review Board at Purdue University.

Stimuli Six English words were selected for each of six categories (Table 1). The words were partitioned into two disjoint sets. These were formed into bigrams, half of which were syntactic constituents and half of which were not (Table 2). The non-constituent bigrams were obtained by reversing the word order of the constituent bigrams. Bigrams were always formed from words selected from the same set.

A rapid event-related design (Just, Cherkassky, Aryal, & Mitchell, 2010) was employed. There were 16 runs, with each containing 72 stimulus presentations for a total of $16 \times 72 =$

Table 1: Words used to construct bigram stimuli.

Category	Set A			Set B		
determiner (D)	*both*	*most*	*the*	*few*	*many*	*no*
verb (V)	*abandon*	*buy*	*catch*	*adore*	*carry*	*choose*
intensifier (I)	*really*	*so*	*too*	*super*	*quite*	*very*
preposition (P)	*about*	*of*	*with*	*among*	*on*	*without*
noun (N)	*apples*	*cars*	*dogs*	*books*	*chairs*	*flowers*
adjective (A)	*big*	*bright*	*smelly*	*loud*	*soft*	*sweat*

Table 2: Constituent and non-constituent bigram stimuli.

Category	Constituent	non-Constituent
noun	D N	N D
verb	V N	N V
adjective	I A	A I
preposition	P N	N P

1152 stimulus presentations. Each stimulus presentation consisted of a bigram presented visually as text for 2 s in a random font, with random point size, and random position in the field of view, synced to the TR trigger. The even numbered runs (from zero) used bigrams only from set A ('*set*'). The odd numbered runs used bigrams only from set B. For each run, half of the stimuli were constituents and half were non-constituents ('*constituency*'). One quarter were noun, verb, adjective, and preposition bigrams respectively ('*category*'). Each run contained a single stimulus presentation for possible combination of all three first words in the bigram paired with all three second words, *i.e.*, $2 \times 4 \times 3 \times 3 = 72$ stimulus presentations, randomly presented. TR was 2 s. Each run comprised 278 TRs (9:16), beginning with four TRs of fixation, ending with ten TRs of fixation, with a minimum of two TRs fixation between stimuli. An additional 48 TRs of jitter fixation were randomly distributed between stimulus presentations. Presentation order and jitter varied randomly by run. Font, point size, and position in the field of view varied randomly by stimulus presentation.

Data Acquisition Imaging was performed using a 3T GE scanner with 16 channel brain array to collect whole-brain volumes via a gradient-echo EPI sequence. Thirty-five axial slices were acquired with a 3.0mm slice thickness using a 64×64 acquisition matrix resulting in 3.125mm×3.125mm×3.0mm voxels. The subject was given no task or instructions except to read the stimuli, as presented, in their head, but not to vocalize them.

Preprocessing Whole-brain scans were processed using AFNI to drop the first two TRs of each run, skull-strip each volume, motion correct, slice-timing correct, and detrend each run, and align all scans for a given subject to a subject-specific reference volume. Voxels within a run were z-scored. Since each brain volume has very high dimension, 143,360 voxels, voxels were eliminated by computing a per-voxel Fisher score on the dataset and keeping the 1,024 highest-scoring voxels.[1] The Fisher score of a voxel v for a classification task with C classes where each class c has n_c examples was computed as

$$\frac{\sum_{c=1}^{C} n_c(\mu_{c,v} - \mu_c)^2}{\sum_{c=1}^{C} n_c \sigma_{c,v}^2}$$

where $\mu_{c,v}$ and $\sigma_{c,v}^2$ are the per-class per-voxel means and variances and μ_c was the per-class mean for the entire brain volume. Voxel selection varied by analysis, since the classes differed, but was performed on the entire dataset containing both the training and test set for each analysis.

Analysis A two-layer perceptron with the same number of hidden units as voxels was used to classify the selected voxels. Rectified linear units were used as the activation function for the first layer, 50% dropout was employed at the input to each layer, and classification was performed with log soft-max and a negative log likelihood class criterion. A single TR for each stimulus, 3 TRs after stimulus onset, was used to train and test classifiers, to compensate for the hemodynamic response function (HRF). Three types of classifiers were trained.

word Twelve 1-out-of-3 word classifiers (chance 33%) were trained, one for each of the six categories (determiner, verb, intensifier, preposition, noun, and adjective) for each set (A and B). (*E.g.*, *abandon vs. buy vs. catch*.) Thus there were $\frac{1152}{4\times2} = 144$ samples for the determiner, verb, intensifier, preposition, and adjective, classifiers, while there were $3 \times \frac{1152}{4\times2} = 432$ samples for the noun classifier. Eight-fold leave-one-run-out cross-validation was performed, training on data from 7 of the 8 runs, testing on the eighth, cycling among all eight runs as test set. Thus the noun classifiers were trained on $\frac{7\times432}{8} = 378$ samples and tested on $\frac{432}{8} = 54$ samples, while the classifiers for other categories were trained on $\frac{7\times144}{8} = 126$ samples and tested on $\frac{144}{8} = 18$ samples.

category Two 1-out-of-4 category classifiers (noun *vs.* verb *vs.* adjective *vs.* preposition; chance 25%) were trained, one trained on set A and tested on set B, the other vice versa. TRs corresponding to all 1152 stimuli were used, half for training and half for test.

constituency Eight binary constituency classifiers (constituent *vs.* non-constituent; chance 50%) were trained. Half were trained on set A and tested on set B; half vice versa. For each half, classifiers were trained; each one trained on samples from three categories and tested on samples of the fourth. (*E.g.*, 1 of the 8 classifiers was trained on noun, verb, and adjective samples from set A but tested on preposition samples from set B.) Thus each

[1] An alternate analysis with only the 128 highest-scoring voxels yielded similar results.

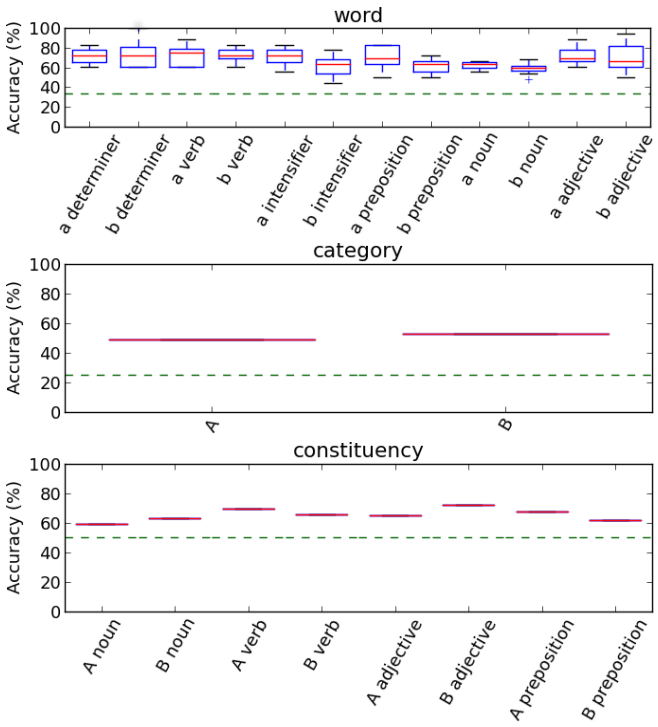

Figure 1: Accuracies of word classifiers (top), category classifiers (middle), and constituency classifiers (bottom). Each classifier in the above has $p < 0.02$.

classifier was trained on $\frac{3\times1152}{4\times2} = 432$ samples and tested on $\frac{1152}{4\times2} = 144$ samples.

Results

Classifier accuracies are shown in Figure 1. All classifiers determine their target with statistical significance ($p < 0.02$), except for the set A noun and set B preposition constituency classifiers, $p < 0.001$.

- The accuracies of the word classifiers show that the identity of the lexical items is manifest in brain activity.
- The accuracies of the category classifiers show that lexical category information is manifest in brain activity independent of the lexical items, since the classifiers were tested on different words than they were trained on.
- The accuracies of the constituency classifiers show that constituency is manifest in brain activity independent of the lexical items and the lexical categories, since the classifiers were tested on different lexical items of different lexical categories than they were trained on.

Interpretation

The results confirm our hypotheses, suggesting that the derivation of complex syntactic behavior can be reduced to a simple concatenative operation, Merge. By using machine-learning techniques, we do not assume that the neural implementation of Merge is itself simple (*e.g.*, the function is located in BA44; Zaccarella & Friederici, 2015). To our knowledge, we are the first to do so w.r.t. characterizing Merge. Machine-learning techniques allow us access to fine-grained linguistic distinctions that may otherwise be undetectable (Allen, Pereira, Botvinick, & Goldberg, 2012), while much of neurolinguistics is concerned with coarse-grained linguistic effects (*e.g.*, active *vs.* passive sentences; see Poeppel & Embick, 2005; Poeppel, 2012).

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1522954-IIS. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Allen, K., Pereira, F., Botvinick, M., & Goldberg, A. E. (2012). Distinguishing grammatical constructions with fMRI pattern analysis. *Brain and Language*, *123*(3), 174–182.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience*, *31*(8), 2801–2814.

Chomsky, N. (1995). *The minimalist program*. MIT press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (p. 1-52). MIT Press.

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*(4), 665–679.

Indefrey, P., Brown, C. M., Hellwig, F., Amunts, K., Herzog, H., Seitz, R. J., & Hagoort, P. (2001). A neural correlate of syntactic encoding during speech production. *Proceedings of the National Academy of Sciences*, *98*(10), 5933–5936.

Indefrey, P., Hellwig, F., Herzog, H., Seitz, R. J., & Hagoort, P. (2004). Neural responses to the production and comprehension of syntax in identical utterances. *Brain and language*, *89*(2), 312–319.

Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PloS One*, *5*(1), e8622.

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, *108*(6), 2522–2527.

Poeppel, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cognitive neuropsychology*, *29*(1-2), 34–55.

Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. *Twenty-first century psycholinguistics: Four cornerstones*, 103–118.

Zaccarella, E., & Friederici, A. D. (2015). Merge processing in the human brain: a cluster-based functional investigation in the left pars opercularis. *Frontiers in Psychology*, *6*, 1818.