Temporal Lobes as Combinatory Engines for both Form and Meaning

Jixing Li  
Cornell University  
jl2939@cornell.edu

Jonathan Brennan  
University of Michigan  
jobrenn@umich.edu

Adam Mahar  
Cornell University  
ajm348@cornell.edu

John Hale  
Cornell University  
jthale@cornell.edu

Abstract

The relative contributions of meaning and form to sentence processing remains an outstanding issue across the language sciences. We examine this issue by formalizing four incremental complexity metrics and comparing them against freely-available ROI timecourses. Syntax-related metrics based on top-down parsing and structural dependency-distance turn out to significantly improve a regression model, compared to a simpler model that formalizes only conceptual combination using a distributional vector-space model. This confirms the view of the anterior temporal lobes as combinatory engines that deal in both form (see e.g. Brennan et al., 2012; Rogalsky and Hickok, 2009) and meaning (see e.g., Wilson et al., 2014). This same characterization applies to a posterior temporal region in roughly “Wernicke’s Area.”

1 Introduction

Processing complexity in human language comprehension remains a central challenge for computational psycholinguistics. Investigations of this essentially biological phenomenon typically rely on formalized complexity metrics. These metrics reflect some aspect of the language being comprehended: some are form-based in the sense of syntactic structure while others are meaning-based in the sense of conceptual information.

But what is the biological basis of the processing that these metrics index? The clinical syndrome semantic dementia suggests that the anterior temporal lobes (ATLs) perform some sort of conceptual combination (for a review, see Patterson et al., 2007). But it remains unclear whether this conceptual processing overlaps or is separate from form-based processing e.g. based on syntactic phrase structure.

To disentangle the influence of form and meaning in sentence processing in different brain regions, we used stepwise regression against freely-available ROI timecourses (Brennan et al., 2016). The regressors in these statistical models are incremental complexity metrics formalizing several different cognitive and linguistic theories about processing difficulty in form and meaning. The pattern of improvements across these steps suggests a role for syntactic processing, above and beyond conceptual combination. This result is consistent with the experimental work of Rogalsky and Hickok (2009) as well as the correlational work of Brennan et al. (2012) on which we build. The remainder of the paper is organized into four sections: Section 2 reviews our syntactic and semantic complexity metrics; Section 3 describes the material and data analysis methods; Section 4 presents the results and Section 5 discusses the implications of the results and concludes the paper.

2 Quantifying complexity factors

We quantify two different aspects of syntactic complexity: Structural Distance and Node Count (this latter metric previously investigated in Brennan et al., 2016), and we use vector-space model to quantify semantic complexity as Lexical-Semantic Coherence. In evaluating the contribution of these complexity metrics, we control for linear order in two ways: Lexical sequences from Google Book ngrams (Michel, 2011), and the linear order of parts of speech using the same POS trigram model in Brennan et al. (2016).
2.1 Structural Distance

One form-related aspect of processing difficulty derives from memory load induced by integration of two syntactically dependent words (Wanner and Maratsos, 1978; O’Grady, 1997; Gibson, 1998). Following Baumann (2014), we quantify this load as Structural Distance, i.e., the number of phrase-structural tree nodes between two dependent words. We obtained both phrase structures and dependency relations for every sentence using the Stanford Parser (Klein and Manning, 2003; de Marneffe et al., 2006). Structural Distance is then the number of nodes traversed between the head and the dependent in the phrase structural tree. We considered only the rightmost word in any dependency relation. For words in multiple dependency relations, we summed the structural distances.

2.2 Node Count

Another form-based complexity metric is Node Count, which is the number of phrase structural nodes in between successive words in a sentence. This expresses a form of Yngve’s (1960) Depth hypothesis (see also Frazier, 1985). We examined X-bar structures generated by Minimalist Grammars in the sense of Stabler (1997). These structures reflect grammatical analysis by Van Wagenen et al. (2014). We counted the number of nodes in these trees that would be visited by a top-down parser (see Hale, 2014).

2.3 Lexical-Semantic Coherence

Our meaning-based metric Lexical-Semantic Coherence is built on vector-space models. Vector-space models represent word meaning based on co-occurrence statistics from a large text corpus (e.g., Baroni et al., 2014; Erk, 2012). Cosine similarity between the word vectors have been found to influence eye-fixation times (Pynte et al., 2008), word pronunciation duration (Sayeed et al., 2015), and fMRI activation patterns (Mitchell et al., 2008). We used latent semantic analysis (LSA; Landauer and Dumais, 1997) to build our semantic vector space model. The training data were the whole book of Alice in Wonderland. We first built the type-by-document matrix where the rows are all the words in the book and the documents are all the paragraphs. The input vector space was transformed by singular value decomposition (SVD), and truncated to a 100-dimensional vector space. The context vector is the average of the previous 10 word vectors. We used negative cosine between the target word vector and the context vector to represent lexical-semantic coherence: higher negative cosine value indicates less semantic coherence.

2.4 Linear Order

Our control predictors include the lexical and POS trigram models. Linear order of words, as reflected in a Markov chain, has been successful in modeling human reading performance (Frank and Bod, 2011; Frank et al., 2015). We used the freely-available trigram counts from the Google Books project (see e.g. Michel, 2011) and restricted consideration to publication years 1850-1900, i.e., the year surrounding the publication of Alice in Wonderland. We backed off to lower-order grams where necessary: coverage was 1725/2045 for trigrams and 1640/1694 for bigrams. The POS trigram regressor from ?) served as an additional control. We then used surprisal of the trigram probabilities to link the probability of a word in its left-context to BOLD signals (see Hale, 2001; 2016).

3 Methods

3.1 Data acquisition

The ROI timecourses from Brennan et al. (2016) come from twenty-five native English speaker (17 female, 18-24 years old, right-handed) listening to a story while in the scanner. The story was the first chapter of Alice in Wonderland, lasting for about 12.4 minutes. Participants completed twelve multiple-choice questions after scanning. The detailed imaging parameters and preprocessing procedures are described in Brennan et al. (2016).
3.2 Regions of interest
Six regions of interest (ROIs), including the left anterior temporal lobe (LATL), the right anterior temporal lobe (RATL), the left inferior frontal gyrus (LIFG), the left posterior temporal lobe (LPTL), the left inferior parietal lobe (LIPL) and the left premotor region (LPreM).

Both functional and anatomic criteria guided the precise positioning of these ROIs. The functional criterion derives from an atheoretical Word Rate regressor, which has value 1 at the offset of each word in the audio stimulus, and 0 elsewhere. This localizer identified regions whose BOLD signals were sensitive to word presentation. Each ROI sphere (10 mm radius) was centered on a peak \( t \)-value of at least 2.0 within the anatomical areas.

3.3 Data analysis
3.3.1 Estimating hemodynamic response
Following Just and Varma (2007), we convolved each complexity metric's time series with SPM12's canonical hemodynamic response function (HRF). These time series are made orthogonal to the convolved Word Rate vector since it is our localizer for defining the ROIs.

3.3.2 Stepwise regression
We tested the unique contribution of each model by conducting stepwise model comparisons against the ROI timecourses. The null model included fixed effects for head movements (dx, dy, dz, rx, ry, rz) and word rate; We also included fixed effects for word frequency, \( f_0 \), and root mean square (RMS) intensity of the speech into our null model, which were also convolved with the same HRF. Word frequency was based on the SUBTLEXus corpus (Brysbaert and New, 2009), which contains 51 million words from the subtitles of American films and television series. The random effects included a random intercept by participant and a random slope for word rate:

\[
BOLD_{null} = BOLD \sim dx + dy + dz + rx + ry + rz + rate + f_0 + intensity + frequency(1 + rate|subject) \tag{1}
\]

We then added regressors in a particular order: surprisal of trigram lexical, negative cosine similarity between word vector and context vector (semantic coherence), surprisal of trigram pos, top-down node count and structural distance between dependent words. Model fit was assessed using chi-square tests on the log-likelihood values to compare different models. Both the predictors were converted to z-scores before statistical analysis. Statistical significance was corrected for multiple comparisons across six ROIs with the Bonferroni method (the adjusted alpha-level is 0.05/6=0.0083).

4 Results
4.1 Correlation between predictors
The correlation matrix shows highest values for word rate and intensity (\( r = 0.58 \)). This is expected as word rate tracks the presentation of a word, which is generally higher in intensity than silences. Similarly, \( f_0 \) is also moderately correlated with intensity (\( r = 0.39 \)) and word rate (\( r = 0.37 \)). Semantic coherence and word frequency have a correlation coefficient of 0.38; no other two parameters has a correlation coefficient higher than 0.3.

4.2 Model comparison
The complexity parameters are subsequently added to the six baseline models. In the ATLs, an improvement in the goodness of fit is obtained for Lexical-Semantic Coherence, but Structural Distance is also significant for the RATL. All the parameters are highly significant for the LPTL, roughly corresponding to the traditional “Wernicke’s area”. Lexical-Semantic Coherence and Structural Distance also significantly improve model fit in the LIPL. However, only the linear order lexical and POS trigram models are significant for the LIFG. The statistical details for the model comparisons are shown in Table 1.
Table 1: Step-wise model comparison results for all regions of interest.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>df</th>
<th>LogLik</th>
<th>χ²</th>
<th>p</th>
<th>Parameter</th>
<th>df</th>
<th>LogLik</th>
<th>χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>∅</td>
<td>15</td>
<td>-11661</td>
<td></td>
<td></td>
<td>∅</td>
<td>15</td>
<td>-11221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A trigram lexical</td>
<td>16</td>
<td>-11625</td>
<td>71.3</td>
<td>&lt;.001</td>
<td>B semantic coherence</td>
<td>17</td>
<td>-11614</td>
<td>22.8</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>B semantic coherence</td>
<td>17</td>
<td>-11614</td>
<td>22.8</td>
<td>&lt;.001</td>
<td>C trigram pos</td>
<td>18</td>
<td>-11608</td>
<td>12.7</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>C trigram pos</td>
<td>18</td>
<td>-11608</td>
<td>12.7</td>
<td>&lt;.001</td>
<td>D node count</td>
<td>19</td>
<td>-11605</td>
<td>4.3</td>
<td>0.04</td>
</tr>
<tr>
<td>D node count</td>
<td>19</td>
<td>-11605</td>
<td>4.3</td>
<td>0.04</td>
<td>E structural distance</td>
<td>20</td>
<td>-11605</td>
<td>0.9</td>
<td>0.34</td>
</tr>
</tbody>
</table>

5 Discussion & Conclusions

The meaning-based metric Lexical-Semantic Coherence is a significant predictor across a broad network of regions including the ATLs, LPTL, LIPL and LPreM. This is consistent with previous findings implicating bilateral ATL in conceptual combination (Rogalsky and Hickok, 2009; Wilson et al., 2014; Pylkkänen, 2015). The form-related metric Structural Distance accounts for the RATL activity even on top of Lexical-Semantic Coherence, suggesting that the ATLs are also involved in syntactic computation (Humphries et al., 2006; Brennan et al., 2012; Brennan et al., 2016).

The LPTL activity is highly correlated with all the syntactic and semantic complexity metrics. As shown in Wehbe et al. (2014), multiple regions spanning the bilateral temporal cortices represent both syntax or semantics. Our results further confirms their suggestion that syntax and semantics might be non-dissociated concepts.

No semantic or syntactic metric is significantly correlated with the LIFG, or the “Broca’s area”. This fails to support traditional models derived from the deficit-lesion studies that have long associated syntactic computation with the LIFG (e.g., Ben-Shachar et al., 2003; Caplan et al., 2008; Just et al., 1996; Stromswold et al., 1996). .

To sum up, our correlational results from fMRI suggest that the temporal lobes perform a kind of computation that is both syntactic in the classical sense of phrase structure, and semantic in the sense of word-embeddings. One set of questions this work leaves open is the precise relationships between these two predictors – for instance, temporal precedence. Other methods, such as MEG, may provide further insight here as suggested by van Schijndel et al. (2015).

6 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1607441. The authors thank Brian Roark for his assistance with OpenGrm.
References


