Language and Linguistics Compass 10/7 (2016): 299–313, 10.1111/lnc3.12198

Naturalistic Sentence Comprehension in the Brain

Jonathan Brennan*
Department of Linguistics, University of Michigan

Abstract

The cognitive neuroscience of language relies largely on controlled experiments that are different from the everyday situations in which we use language. This review describes an approach that studies specific aspects of sentence comprehension in the brain using data collected while participants perform an everyday task, such as listening to a story. The approach uses 'neuro-computational' models that are based on linguistic and psycholinguistic theories. These models quantify how a specific computation, such as identifying a syntactic constituent, might be carried out by a neural circuit word-by-word. Model predictions are tested for their statistical fit with measured brain data. The paper discusses three applications of this approach: (i) to probe the location and timing of linguistic processing in the brain without requiring unnatural tasks and stimuli, (ii) to test theoretical hypotheses by comparing the fits of different models to naturalistic data, and (iii) to study neural mechanisms for language processing in populations that are poorly served by traditional methods.

1. Language comprehension inside and outside of the lab

Research in the cognitive neuroscience of language has mapped many of the relevant brain regions and has begun to reveal the dynamic interplay between these regions that underlies language comprehension and production (for reviews, see Friederici and Gierhan 2013; Hagoort and Indefrey 2014; see Kemmerer 2014 for a textbook introduction). There is growing interest in whether these results extend beyond constrained laboratory settings to natural, everyday uses of language like listening to a story or having a face-to-face conversation (see the papers collected in Willems 2015). The stimuli and tasks used in these new efforts are 'naturalistic' in that they are drawn from how language is used outside of the laboratory but are constrained by the equipment needed to record brain signals. This review describes an approach to studying the neural bases of specific sub-processes of sentence processing during naturalistic comprehension.

The majority of work using naturalistic stimuli has addressed language processing at a coarse-grained level. Studies have outlined a range of brain regions that are engaged during natural reading (Yarkoni et al. 2008; Speer et al. 2009; Xu et al. 2005; Wehbe et al. 2014), listening (Brennan et al. 2012; Whitney et al. 2009; Lerner et al. 2014), and audio-visual processing (Wilson et al. 2008; Skipper et al. 2009). Extensions of this work have identified neural signals associated with higher-order processes that are shared across speech rates (Lerner et al. 2014) and across production and comprehension (Stephens et al. 2010; Silbert et al. 2014). Other work using naturalistic stimulation has focused on aspects of discourse comprehension (Whitney et al 2009; Egidi and Caramazza 2013; Kurby and Zacks 2013). These efforts have proved fruitful in building bridges between cognitive neuroscience and humanistic studies including the study of literature (Willems 2013), poetics (Jacobs 2015), and cinema (Hasson et al. 2008). In contrast, there have been relatively few studies of more fine-grained linguistic processes at the sentence level and below.

Research at or below the sentence level has relied largely on controlled experiments that use minimal pairs of stimuli or tasks to isolate relevant neural signals. The primary experimental logic is that of *subtraction*. This research has been undoubtedly successful but faces two limitations. Firstly, the subtraction approach necessarily requires linguistic stimuli and tasks that are rather different from everyday language. Secondly, the great majority of these studies are framed in terms of qualitative processing models that limit their generalizability.

Most sentence-level experiments present participants with a series of carefully constructed sentences, phrases, or words with little or no context. To encourage attention, participants may be asked to render a judgment about the sensicality, acceptability, or some other linguistic property of a stimulus. Such meta-linguistic judgments require processing other than that used in everyday settings, which often require one to just interpret the input. However, there has been little discussion of how meta-linguistic processes relate to everyday comprehension.

The stimuli that are used in sentence-level experiments also often deviate from everyday language. For example, a common technique to isolate aspects of syntactic parsing uses stimuli that do or do not violate expectations about syntactic structure (e.g. Neville et al. 1991; Hagoort et al. 1993 and many others; see Gouvea et al. 2010 for discussion). Stimuli that violate expectations are taken to elicit more syntactic processing with debate over what kinds of operations are involved (e.g. Kim and Osterhout 2005; Chow and Phillips 2013). These studies do not make explicit how the processing of ungrammatical sentences might relate to naturalistic comprehension. Are ungrammatical utterances recognized as such and corrected to some grammatical form? Or, might ungrammatical components be ignored or backgrounded, as when listening to false starts in a natural utterance? The latter is an implicit assumption associated with another common experimental design which compares well-formed sentences with lists of unstructured words (e.g. Mazoyer et al. 1993; Stowe et al. 1998; Humphries et al. 2006). In these experiments, it is assumed that sentence-level operations, broadly construed, are evoked by full sentences but not by (ungrammatical) lists of words. As with artificial meta-linguistic tasks, studies do not make explicit how the processing of isolated or unusual sentences relates to everyday comprehension.

With some recent and notable exceptions (Gibson et al. 2013; Brouwer 2014), experimental manipulations are typically framed in terms of qualitative distinctions. For example, a condition does, or does not, require 'syntactic reanalysis', or is more or less 'semantically demanding'. Processing models based on these studies are similarly qualitative; they might distinguish the function of brain regions in terms of 'linear' vs. 'hierarchical processing' or 'syntactic' vs. 'semantic combinatorics' (e.g. Bornkessel-Scheslewsky et al. 2015; Friederici and Gierhan 2013). Qualitative descriptions like these are valuable, but they amplify the challenges posed by unnatural stimuli and tasks as they do not allow for robust generalization from controlled experiments to naturalistic contexts. This is because such models do not indicate with sufficient rigor how target processes will be engaged beyond the narrow domain of the stimuli and task that constitute a specific experiment.

In sum, predominant neurolinguistic approaches to sentence- and word-level processing deviate from everyday language use and typically lack rigorous processing models that are necessary to extend conclusions drawn from controlled experiments to more natural settings. While everyday language can be studied in qualitative terms (e.g. Lerner et al. 2014; Egidi and Caramazza 2013), naturalistic data at the sentence level and below demand the use of rigorously specified theoretical models. To adequately take advantage of the rich data offered by everyday language, theories must be specified with enough detail to broadly cover most, or all, of the words found in a naturalistic corpus.

The next section introduces an approach that addresses these limitations by using computationally precise and broad coverage cognitive models of naturalistic processing. These 'neurocomputational' models allow fine-grained aspects of sentence comprehension to be studied

with naturalistic neural data. Properties of linguistic representations are mapped to neural signals by specifying the algorithm by which those representations are used in real time by the brain. By making such a mapping explicit, this approach addresses a chief barrier to integrating (psycho) linguistic theories with neuroscientific data (for discussion, see Poeppel 2012; Poeppel and Embick 2005; Embick and Poeppel 2015).

Algorithmic theories of how linguistic knowledge is deployed in real time have been developed within computational psycholinguistics. Computational psycholinguistic models offer quantitative estimates of how specific syntactic, semantic, or other processes are engaged word-by-word. For example, a model might specify the syntactic parse states that unfold incrementally during comprehension (Hale 2014). Actual brain signals that are recorded while participants encounter stimuli within the domain of the computational model provide the relevant data. Models are then put to the test by comparing the estimated sequence of cognitive states with the measured brain signals.

Neuro-computational models can be tested against data collected with many techniques used in cognitive neuroscience. Each technique has its own strengths and weaknesses. For example, data collected with functional magnetic resonance imaging (fMRI) reflect changes in blood oxygenation in response to neuronal activity. While fMRI has high millimeter-level spatial resolution, it has a low temporal resolution. This is because blood oxygenation changes slowly, on the order of several seconds, compared to the speed of language. Data collected with electroencephalography (EEG), similar to the related technique of magnetoencephalography (MEG), reflect neuroelectric activity as it happens millisecond-by-millisecond. These techniques do so at the cost of poor (EEG) to moderate (MEG) spatial resolution. As detailed in the next section, adjustments necessary to query different data types are incorporated into the models.

The neuro-computational model approach adds to the neurolinguistic toolbox. It augments traditional experimental methods by facilitating the analysis of data collected during naturalistic processing. By addressing fine-grained aspects of processing at and below the sentence level, neuro-computational models complement other approaches that use naturalistic neural data to study language processing at a coarse grain (e.g. Lerner et al. 2014). Section 2 provides a formal description of the approach, and Sections 3–5 highlight several applications.

2. Analyzing naturalistic neural data using neuro-computational models

A neuro-computational model operates over linguistic input, such as sequences of words, and returns mental states. Aspects of these mental states are quantified in order to link them to measured brain signals. Running the model generates sequences of estimated brain signals. To evaluate the model, the sequence of estimated brain signals is compared with actual measured brain signals. For instance, the number of syntactic nodes in the mental state might be taken to vary proportionally with fMRI signals as a way to study constituent structure processing.

The model can be understood as a set of functions. A parser $P_{G,A,O}$ consists of a grammar G that defines well-formed representations, an algorithm A that specifies how the grammar is applied word-by-word, and an oracle O that reconciles indeterminacies, for instance by choosing to evaluate the most common grammatically licensed representation of all that are possible for a given sub-string. Applied to a sequence of words w_1, w_2, w_3… $P_{G,A,O}$ yields a sequence of mental states m_1, m_2, m_3…. For syntax, these might be partial tree structures. While the term 'parsing' commonly refers to syntactic processing, the present usage is not intended to be limited in this way. Other examples include identifying whether a sequence of phonemes is phonotactically well-formed or computing the implicatures that are triggered by a sequence

of utterances. In all cases, the parameters *G*, *A*, and *O* jointly characterize a sequence of mental states that connect the linguistic input and desired output. A complexity metric *C* quantifies this sequence of mental states in some way. For example, *C* may count the number of tree nodes, open dependencies, or the change in probability mass between states. These values serve as estimators for brain states. Lastly, estimated brain states are aligned with measured brain signals with a response function *R* which stands in for whatever mediates between the actual physiological state of the brain and the signal that is measured with a technique such as fMRI, EEG, or MEG. For instance, *R* might be the hemodynamic response function (HRF) used in fMRI research to account for the delay between neuroelectric activity and measured changes in blood oxygenation. The composition of *C* with *R* constitutes a linking hypothesis in the sense of Embick and Poeppel (2015): together these functions connect the properties of a theoretical mental state with an observable brain signal. The connection between these components is shown on the left-hand side of the schematic in example (1).

The right-hand side of (1) shows the data against which the models are evaluated. The function *H* stands for the human brain's response to a sequence of words to yield a sequence of internal brain states $b_1, b_2, b_3 \ldots$. The term *Measure* is a function that returns an observable signal for a particular brain state. This term stands in for all of the choices involved in the measurement and analysis of functional brain data, examples of which are given in the sections that follow. The final line of the schematic in represents a statistical test of equivalence, such as linear regression, between the output of the model and the measured brain data.

$$
\begin{array}{cc}
\underline{\textit{model}} & \underline{\textit{data}} \\
P_{G,A,O}(w_1, w_2, w_3, \ldots) & H(w_1, w_2, w_3, \ldots) \\
\downarrow & \downarrow \\
\text{parser interprets} & \text{human listens} \\
\downarrow & \downarrow \\
m_1, m_2, m_3 \ldots & b_1, b_3, b_3 \ldots
\end{array}
$$

$$
R \circ C(m_1, m_2, m_3 \ldots) \stackrel{?}{\equiv} \text{Measure}(b_1, b_3, b_3 \ldots)
\tag{1}
$$

Figure 1 provides a concrete illustration of the model-based approach. This example comes from an fMRI experiment about syntactic processing by Brennan et al. (in press), which is discussed in Section 4 below. Step 1 shows a sequence of mental states defined by a particular parser, and Step 2 shows how the complexity metric and response function together quantify these mental states to create an estimated brain signal. In this example, the estimate is for a brain signal associated with computing syntactic expectations. Steps 3 and 4 on the right-hand side illustrate the measured fMRI data against which the model is evaluated.

The model-based approach that is illustrated in Figure 1 describes a multi-dimensional hypothesis space. The space is defined by the *Model* parameters $< G,A,O,R,C >$ on the left-hand side and *Measure* on the right-hand, or *Data*, side of the schematic in example (1). Research in the neuroscience of sentence comprehension, whether naturalistic or not, aims to identify which points in this space best match the actual operations of the mind and brain. It is not typically feasible to explore all hypotheses at once. Rather, studies focus on a limited sub-space. Most studies in the literature fall into one of two categories which are schematized in example (2). A neurally focused 'N-study' tests hypotheses about the location, timing, or other

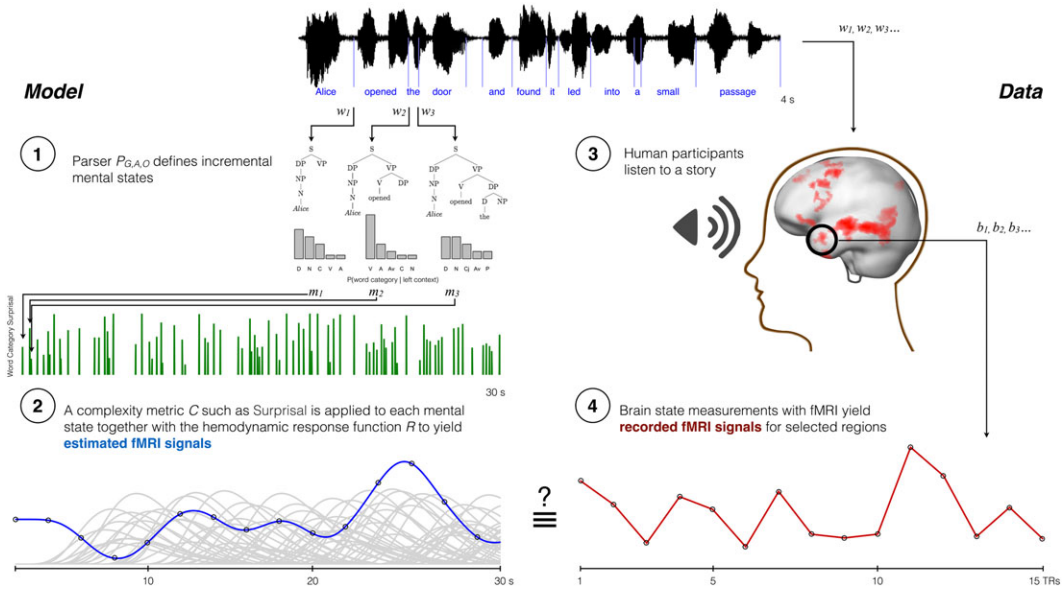

Fig. 1. An illustration of the model-based approach. The top of the figure shows a segment of a naturalistic audiobook stimulus. Word boundaries are indicated in blue. (1) The parser defines word-by-word mental states that reflect syntactic constituency and expectations for upcoming words, shown in the gray probability distributions. Here, the grammar *G* is set to a *context-free phrase structure grammar*, the algorithm *A* is set to *top-down enumeration*, and the oracle *O* is set to resolve temporary ambiguities to the structures that are ultimately correct. (2) These mental states are quantified to derive processing predictions by applying the *surprisal* complexity metric, *C*, shown in green. This quantity reflects the conditional probability of each word in the stimulus and is higher when words are unexpected. A response function *R* aligns word-by-word complexity values with fMRI-measured brain signals. The overlapping gray traces show the result of convolving each word's surprisal with the *hemodynamic response function*. The sum of these, shown in blue, is an estimate of the time-course of fMRI signals that reflect syntactic expectation. (3) The data for this illustration come from recordings while participants passively listened to the audiobook. (4) The measured signal from a particular region of interest is extracted (red trace) and correlated against the estimated signal (blue trace) to test how well the estimated fMRI signal aligns with the recorded fMRI signal.

brain-based property of some linguistic process. It does so by varying the brain signal being tested, represented by the *Measure* parameter(s), while keeping the model side of the equation constant. For instance, a single parsing model might be tested against brain data collected with fMRI from across many different brain regions. Alternatively, a linguistically focused 'L-study' tests linguistic or psycholinguistic theories by comparing different parameterizations of the *Model* side against one, or just a few, settings for *Measure*. For example, different parsers could be tested against fMRI data from a single region of interest.

N-Study			L-Study		
$mode$	$\overset{?}{\equiv}$	$data$	$model$	$\overset{?}{\equiv}$	$data$
		$Measure_1$	$< G_1, A, O, C, R >$		
$< G, A, O, C, R >$		$Measure_2$	$< G_2, A, O, C, R >$		$Measure$
		$Measure_3$	$< G_3, A, O, C, R >$		

(2)

Of course, it is also conceivable to evaluate hypotheses on both sides of the equation simultaneously. While such an approach has not been pursued in neurolinguistics, there are potential examples elsewhere in the neurosciences. Kumar and Penney (2014) develop a method for

estimating 'Neuronal Response Functions', or mappings from stimuli to neuronal events (equivalent to the $P_{G,A,O}$ and C functions in the present framework), across a large number of brain regions using Bayesian model comparison, and they demonstrate its application to studying basic auditory processing. The sections below illustrate the model-based approach with both N- and L-type studies and with several different cognitive neuroscience techniques.

3. The location and timing of naturalistic linguistic processing in the brain

This section establishes the model-based approach as valid for isolating specific aspects of sentence processing during naturalistic comprehension. Syntactic parsing is a sensible starting point for applying neuro-computational models to naturalistic data. It is a well-studied domain within computational psycholinguistics which offers numerous models (see Hale 2014 for an introduction). In addition, naturalistic stimuli are a focus of current debates (see Section 4). Other levels of linguistic processing and representation may be studied with the same framework but present some challenges. At lexical and sub-lexical levels, the ubiquity of context effects (see, for example, DeLong et al. 2005 on lexical prediction and Dikker et al. 2010 on pre-lexical prediction) means that sentence- and higher-level context should be modeled in addition to lexical-level factors for naturalistic data (but see Huth et al., 2016, for a sophisticated approach in the domain of conceptual representation). At semantic and pragmatic levels, computational models of incremental interpretation are currently less well developed than equivalent incremental syntactic models. Given these considerations, syntactic processing is well suited to begin an investigation of naturalistic comprehension.

A first step is to verify that the model-based approach can successfully target specific components of the brain's sentence processing network. Brennan et al. (2012) test whether specific aspects of syntactic processing can be localized in the brain during naturalistic listening using fMRI. To provide a brief background, the inferior frontal lobe of the left hemisphere ('Broca's Area') has long been associated with various sentence-level processes with debate about what specific computations are carried out in this region (see Rogalsky and Hickok 2010 for a critical overview). Other research implicates the anterior portion of the left temporal lobe in processing simple sentences, but not isolated words (see Pylkkänen 2016 for discussion). These findings have been difficult to reconcile as they rely on different sets of stimuli and tasks.

The cognitive process that Brennan et al. focus on is the identification of phrase structure, or basic sentence composition. They use a single model of syntactic parsing to test different hypotheses about where this process occurs in the brain; this is an N-study. The parsing model constructs a syntactic tree word-by-word according to a broad coverage context-free phrase-structure grammar (Marcus et al. 1993). This grammar is paired with an algorithm that enumerates nodes from the bottom up, after identifying each child of a given node, and a 'perfect' oracle that enumerates only the correct tree. Word-by-word parse states are converted to brain state estimates by counting the number of new nodes created upon encountering each word. These node counts, time-aligned with the offset of each word in the auditory stimulus, are matched to the fMRI signal via a response function which takes into account hemodynamic lag and the fMRI sampling rate.

In the notation introduced in Section 2 above, Brennan et al. (2012) evaluate a set of propositions that rely on a single model with the parameters G = *Context-free phrase-structure grammar*, A = *Bottom-up enumeration*, O = *Perfect*, C = *Node count*, and R = *Hemodynamic Response Function*. The propositions differ in the settings for *Measure*: whole-brain fMRI scanning is used to test for any regions that showed sensitivity to their model of basic sentence composition. The first row of Table 1 summarizes the parameter values for this study.

Activity from the left anterior temporal lobe significantly correlates with model estimates for basic sentence composition. Activity from the left inferior frontal gyrus shows no correlation,

Table 1. Comparison of parameter values for the model-based studies referenced in Sections 3 and 4.

	Parser			*Linking hypothesis*		*Measure*
	G	***A***	***O***	***C***	***R***	
1. Brennan et al. 2012	Context-free	Bottom-up	Perfect	Node count	Hemodynamic response	Whole-brain fMRI
2. Willems et al. 2015	Trigram	–	–	Surprisal, entropy reduction	Hemodynamic response	Whole-brain fMRI
3. Bachrach 2008	Context-free	Top-down	Perfect	Node count, Surprisal	Hemodynamic response	Whole-brain fMRI
4. Henderson et al. 2016	Context-free	–	–	Surprisal	Hemodynamic response	Whole-brain fMRI
5. Wehbe et al. 2014	Lexical, syntactic, and discourse features	*ad hoc*	Perfect	Presence/ absence	Hemodynamic response	Whole-brain fMRI classification
6. Brennan and Pylkkänen submitted	Bi-/tri-gram, context-free, minimalist	Left-corner	Perfect	Node count	Identity	MEG source-localized regions of interest
7. Brennan et al. in press	Trigram, context-free	Bottom-up, top-down	Perfect	Node count, surprisal	Hemodynamic response	fMRI regions of interest
8. Brennan et al. 2016 in prep.		–	–	Surprisal	Identity	EEG

Some complexity metrics do not require fully specified models; underspecified parameters are indicated with '–'. The parameters are described in Section 2.

and no other region shows a positive correlation with basic sentence composition. The observed correlation with the left anterior temporal lobe aligns well with the many studies that show a systematic response in this same brain area for simple sentences and phrases but not for lists of words (see Pylkkänen 2016 and studies cited therein). In light of this alignment, the result serves as a 'proof of concept' that the model-based approach can isolate specific sub-processes of cognition during naturalistic language processing.

Similar work by Willems et al. (2015; see also Bachrach 2008) tests for brain regions involved in another aspect of sentence comprehension: making linguistic predictions. This is another N-study as the model parameters were fixed to just a few values while neural measures from the whole brain were tested. Prediction has long been recognized as central to the speed and rapidity with which complex linguistic utterances are understood (see e.g. Federmeier 2007; Van Petten and Luka 2012; Hagoort and Indefrey 2014 for reviews from a neurolinguistic perspective). Most relevant neurolinguistic research comes from electrophysiological studies in which different kinds of predictions are violated. Willems et al. map the spatial localization of brain regions that are sensitive to predictions and use naturalistic stimuli to avoid complications posed by unnatural stimuli that sharply violate expectations.

Willems et al. use a Markov model, a sequence-based grammar, to compute the conditional probability of each word's part-of-speech given just the two immediately preceding words. They use two complexity metrics to quantify these conditional probabilities in terms of

expectation: (1) the degree to which a word is unexpected given the context, or its *lexical surprisal* (Hale 2001), and (2) the degree to which a word decreases uncertainty about the overall sentence, or *entropy reduction* (Hale 2006; see Hale In press, for an introduction to these metrics). By considering all possible outcomes of a given context ('full parallelism'), the metrics in this study can be computed without specifying a parser algorithm or oracle. They test these two models against whole-brain fMRI data while participants listen to three narratives (see Table 1, row 2). Entropy reduction, reflecting predictive strength, correlates with brain activity in fronto-parietal regions including the right inferior frontal gyrus, the left ventral pre-motor cortex, the left supplementary motor area, and the left inferior parietal lobule. Surprisal, reflecting unexpectedness, correlates with activity from temporal regions bilaterally, including the left inferior temporal regions, left and right posterior superior temporal gyrus, and the right anterior temporal pole. They also report effects in the right inferior frontal gyrus, right amygdala, and right brain stem. These spatially distributed results are consistent with models in which linguistic predictions propagate from the top down through multiple levels of representation (e.g. Dikker and Pylkkänen 2013; Molinaro et al. 2013).

Henderson et al. (2016) extend the results of Willems et al. to reading by recording fMRI data while participants read narratives (Table 1, row 4). They query the data with a computational model that estimates *syntactic surprisal* based on a context-free grammar (Roark et al. 2009); this phrase-structure model contrasts with the sequence-based Markov model used by Willems et al. above. Henderson et al. report that structurally unexpected words lead to greater activation in the left inferior frontal gyrus and left anterior temporal lobe. Considered together, the results of Willems et al., Brennan et al., and Henderson et al. show that neuro-computational models with different parameter settings, such as node count or surprisal-based complexity metrics, can successfully target distinct sub-processes of sentence comprehension during both naturalistic listening and reading.

The studies discussed above each target one or two specific aspects of sentence comprehension in isolation. One advantage of the model-based approach is that sufficiently complex neuro-computational models allow multiple sub-processes to be studied simultaneously. Wehbe et al. (2014) illustrate this possibility by combining fMRI data from story-reading with a multi-faceted model that simultaneously addresses aspects of lexical semantics, syntactic features like subject and object, and discourse-level features like character reference.[1] Features at each level comprise the grammar of the model, which is combined with an *ad hoc* algorithm to link features with specific words; a perfect oracle ensures that only the correct feature is associated with each word. The parser states are quantified in terms of the presence or absence of features at a given time-point (see Table 1, row 5). In addition, rather than using linear regression to evaluate their model against brain data (see the final line of the schematic in (1)) they use a statistical classifier to iteratively test whether a particular brain region carries information about some feature of the stimulus. Whereas the studies described above assume that higher values for some complexity metric map linearly to numerically higher neural measurements, the classifier approach used by Wehbe et al. makes no such assumption. Their results reveal a highly articulated map in which the large-scale network traditionally associated with reading is divided into sub-regions specific to lexical semantic, syntactic, and discourse-level features. While too detailed for this brief review, Wehbe et al.'s model permits this network to be even further sub-divided in terms of specific reading processes at each level.

The studies discussed thus far rely on fMRI, which provides high-resolution spatial pictures of brain activity but is limited by the sluggish hemodynamic response. A natural speech rate of two to four words per second means that processing associated with many words is blurred together. Electrophysiologial tools like EEG and MEG measure brain activity millisecond-by-millisecond. Brennan and Pylkkänen (2012) outline the temporal characteristics of naturalistic

sentence processing, broadly construed, by testing for brain activity that is unique to reading stories, compared to lists of words. They use a single-trial analysis in which the data from a time-span following each word are correlated against word-by-word estimates from a model (cf. Hauk et al. 2006). They find broad activation across the temporal and frontal lobes in both hemispheres for story reading which emerges between 0.25 and 0.5 s after word onset. To tease out basic sentence composition from this broad spatio-temporal network, Brennan and Pylkkänen (submitted) construct a simple context-free grammar and combine this with a predictive left-corner parsing algorithm (Hale 2011), a perfect oracle, and a node count complexity metric (see Table 1, row 6). Estimates are correlated with MEG signals collected for each word across the time-span consistent with the general 'story comprehension' response described above. Significant correlations between the node count values and brain activity emerge in the left anterior temporal lobe between 0.35 and 0.5 s after word onset. This result is consistent with the fMRI study by Brennan et al. (2012) and shows that the model-based approach can resolve the rapid temporal dynamics of fine-grained linguistic processes in naturalistic data.

This section reviewed several studies that use a model-based approach to query hemodynamic and electrophysiological brain activity during naturalistic stimulation. Alignments between the results described here and those from more traditional methods validate the approach and show, among other things, that the anterior temporal lobe plays an important role in basic sentence composition during naturalistic comprehension.

4. Testing theoretical hypotheses by comparing models against naturalistic data

In addition to uncovering the neural signatures of naturalistic comprehension, the model-based approach is well suited to addressing questions about mental representations and computations that are specific to everyday processing. By using models that differ in parameter settings for the grammar or other factors, alternative cognitive theories may be evaluated against naturalistic neural data. This section describes one such application.

Psycholinguists have debated whether abstract hierarchical syntactic representations play a role in everyday online comprehension. The hierarchical representations posited in syntactic theories have been developed based on offline data such as acceptability judgments and cross-linguistic comparisons. While much evidence suggests that such detailed syntactic representations do guide syntactic processing in carefully controlled experiments (see Lewis and Phillips 2015 for a review), existing naturalistic data present a mixed picture. Based on interpretation errors where aspects of syntactic structure are ignored, some have argued that surface-based 'good enough' syntactic processing may suffice in many circumstances (Sanford and Sturt 2002; Ferreira et al. 2002; Ferreira and Patson 2007). In support of this view, Frank and Bod (2011) found that eye-movements from naturalistic reading are sensitive to simple word-to-word dependencies, such as can be computed from a Markov model, but not to hierarchical dependencies such as can be computed from a context-free grammar (but cf. Fossum and Levy 2012). Similar results have been found for ERP components sensitive to linguistic expectations (Frank et al. 2015), although this study relied on single-sentence experimental materials and not narrative stimuli. Prior studies have not tested for abstract hierarchical representations using naturalistic neural data.

To probe the nature of the syntactic representations that underlie fMRI-measured brain responses during story comprehension, Brennan et al. (in press) test many models against just a few brain measures collected while participants listen to a chapter from a children's story. Models vary in terms of grammar, algorithm, and complexity metric. Brennan et al. test three grammars with different levels of syntactic detail: sequence-based Markov models that use only word-to-word dependencies, a context-free hierarchical grammar that incorporates hierarchical

structure but not abstractions like syntactic movement (Marcus et al. 1993), and a minimalist grammar that allows for syntactic movement and derives X-bar trees with empty categories (Stabler 1997; Sportiche et al. 2013). They also test several complexity metrics: surprisal and node count via either a bottom-up or a predictive top-down algorithm. The model space thus spans a range of syntactic processing hypotheses, from simple surface-based accounts to abstract proposals from generative syntax (see Table 1, row 7; Figure 1 provides an illustration based on one of the models used in this study).

Models are tested using regression against fMRI data from six brain regions that have been previously implicated in sentence-level processing: left and right anterior temporal lobes, left inferior frontal gyrus, left posterior temporal lobe, left inferior parietal lobe, and left pre-motor cortex. Surprisal from the non-hierarchical Markov models correlates with frontal and temporal brain regions, and surprisal from a context-free grammar, which is hierarchical, significantly improves the statistical fit of a regression model in all regions but the left inferior frontal gyrus. In contrast, node counts from minimalist grammars further improve the statistical fits in only the left anterior and posterior temporal lobes after controlling for effects of the non-hierarchical and context-free models. This result suggests that the left temporal lobe is involved in constructing abstract hierarchical representations during everyday language comprehension

The limited temporal resolution of fMRI leaves open whether abstract hierarchical representations are recruited rapidly during incremental processing, or whether they play a role only during later processing stages. Brennan et al. (2016, in prep) address this question using EEG. Participants passively listen to the same story stimulus during EEG recording. Prior work has found that stimuli that deviate from an expected word category elicit two event-related potential (ERP) components: an early left anterior negativity (ELAN) between 0.1 and 0.3 s after stimulus onset and a late positivity over central electrodes beginning around 0.6 s (e.g. Friederici et al. 1993). Using the same models as described above and focusing just on part-of-speech surprisal, which quantifies syntactic expectations along the same lines as previous studies, Brennan et al. find that estimates from the hierarchical context-free grammar significantly improve model fit against late centrally distributed signals (0.5–0.7 s), compared with a baseline model that includes word-to-word dependencies, lexical properties such as word frequency, and acoustic control predictors (see Table 1, row 8). These preliminary findings suggest that indices of syntactic expectations may reflect the rapid computation of hierarchical dependencies during naturalistic comprehension.

In sum, sets of neuro-computational models can be compared using data collected during naturalistic comprehension. Varying whether the models are based on hierarchical or non-hierarchical grammars yields different word-by-word estimates for brain activity. Results suggest that hierarchical representations play a role online during naturalistic comprehension.

5. Neural mechanisms for language processing in populations that are poorly served by artificial tasks

A final example of potential applications for the model-based approach comes from the study of language disorders. Combining naturalistic neural data with a model-based approach opens the door to new methods for studying language processing in populations that have difficulty performing standard laboratory tasks. Models can pinpoint specific cognitive processes that are implicated in neural disorders, and naturalistic methods are easy to use with almost any participant. This section describes initial steps in this direction with ongoing research on autism spectrum disorder (ASD).

Individuals with ASD show moderate to severe social deficits in addition to a wide range of well-documented language production deficits (see Groen et al. 2008; Kelley 2011 for reviews).

However, a clear understanding of language comprehension in ASD remains elusive in part because comorbid social deficits lead to problems with behavioral task compliance (Tager-Flusberg and Caronna 2007).

Passively collected neural signals offer a way to overcome this challenge. MEG and EEG data collected from task-free subtraction experiments that compare, for instance, brain responses to frequent or infrequent speech sounds (Roberts et al. 2011) or to plausible or implausible sentences (Pijnacker et al. 2010), have revealed intriguing neural differences between individuals with ASD and typically developing children. These effects might implicate difficulties with linguistic predictions, however, behavioral studies of linguistic prediction which rely on constrained experimental tasks show conflicting results (Brock et al. 2008). Just as comorbid deficits impact task compliance in behavioral experiments, they may also have ill-understood effects on processing the unusual and unfamiliar stimuli used in controlled experiments. Data collected during passive naturalistic processing offer an appealing alternative (Lombardo et al. 2015). To this end, our group has begun to collect MEG data from children with and without ASD while they perform passive naturalistic tasks.

For example, in one experiment participants listen passively to a children's story. The computational models described in Section 2 above can be used to probe for activity that is associated with syntactic predictions for this narrative. Figure 2 shows a preliminary set of results from these data in which expectation, quantified in terms of surprisal from three-word part-of-speech sequences (trigrams), is correlated word-by-word with whole-head neuromagnetic signals.[2] Panel A shows the effect of surprisal across all participants: there is a significant correlation between expectations and right temporal activity around 200 ms after word onset. The same pattern of activity is seen for both typically developing (TD, $N = 13$) children and for those with ASD ($N = 14$), as shown in Panel B (statistically, there is a main effect for surprisal but no interaction with participant group). In other words, the online neural response to sequence-based expectations appears to be similar between the two groups.

These preliminary data are presented here not to support a particular conclusion about language processing in ASD but rather to illustrate a methodological point. Pairing the models of syntactic prediction described in Section 4 with passively collected naturalistic brain data grants researchers a new tool to test whether and how neural differences in ASD impact prediction and other aspects of comprehension during sentence processing.

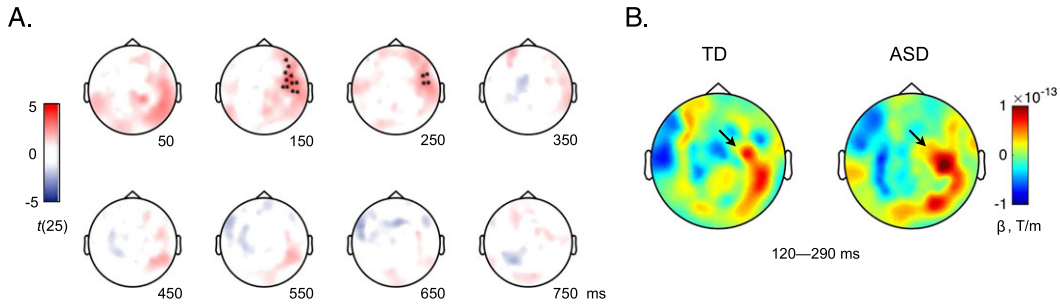

Fig. 2. Preliminary analysis of neural signals reflecting syntactic expectations during naturalistic listening in children with and without ASD. (a) Statistical maps for a group-level effect of *trigram surprisal* on MEG sensor signals. Each panel shows an average over 100 ms; asterisks indicate right temporal sensors that show a significant effect from 120–290 ms ($p < 0.01$, corrected for multiple comparisons). (b) The effect for *trigram surprisal* by group within the same time-window shows statistically indistinguishable right temporal activation patterns for typically developing (left; $N = 13$) children and those with ASD (right; $N = 14$).

6. Conclusions

Naturalistic stimuli and tasks are necessary to identify the brain mechanisms that underlie everyday language processing. The model-based approach allows specific sub-processes to be studied using naturalistic data. Models can be tested across many different neural measurements (e.g. Brennan et al. 2012; Willems et al. 2015), or different models can be evaluated against highly circumscribed brain data (e.g. Brennan et al. in press). While this review has focused on results pertaining to basic sentence composition, the framework can be applied much more broadly; indeed, multiple levels of processing can be probed simultaneously (Wehbe et al. 2014). Fine-grained cognitive resolution is made possible by appealing to rigorous neuro-computational models. These models serve as a bridge between the representations and computations of linguistic theory and the data furnished by the tools of cognitive neuroscience.

Acknowledgement

This work was funded in part by a grant from the University of Michigan M-Cubed Initiative. The author is grateful for feedback from Richard L. Lewis, John T. Hale, Julie Van Dyke, and from two anonymous reviewers.

Notes

* Correspondence address: Jonathan Brennan, Department of Linguistics, University of Michigan, 440 Lorch Hall, 611 Tappan St. Ann Arbor, MI 48109, USA. E-mail: jobrenn@umich.edu

[1] See also Yarkoni et al. (2008) for an early example of naturalistic reading that is focused on lexical processes.

[2] Twelve minutes of story-listening data were collected with 148 magnetometers in a quiet magnetically shielded room. Data were epoched from −0.3 to 1 s around the onset of each word and then cleaned of artifacts with visual inspection and independent component analysis. For each participant, regression weights were estimated for *trigram surprisal* along with covariates of word length and word frequency across all sensors and time-points. For group analysis, regression weights were converted from magnetic flux to the planar magnetic gradient and were statistically compared across all participants and between-groups using a non-parametric cluster test (Maris and Oostenveld 2007).

Works Cited

Bachrach, A. (2008). Imaging neural correlates of syntactic complexity in a naturalistic context. PhD thesis, Massachusetts Institute of Technology.

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., & Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: common computational properties. *Trends in Cognitive Sciences*, 19(3), 142–50.

Brennan, J., Cantor, M., and Hale, J. T. (In prep). EEG indices of syntactic expectation reflect both linear and hierarchical dependencies.

Brennan, J., Cantor, M., Eby, R., and Hale, J. T. (2016). EEG correlates of syntactic expectation reflect both word-to-word and hierarchical dependencies. *Talk presented at the 2016 CUNY Conference on Human Sentence Processing*, Gainesville FL.

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

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

—— (Submitted). MEG evidence for incremental sentence composition in the anterior temporal lobe.

Brennan, J., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., and Hale, J. T. (In press). Abstract linguistic structure correlates with left temporal activity during naturalistic comprehension. *Brain and Language*.

Brock, J., Norbury, C., Einav, S., and Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, 108(3):896–904.

Brouwer, H. (2014). The electrophysiology of language comprehension: A neurocomputational model. PhD thesis, University of Groningen.

Chow, W.-Y. and Phillips, C. (2013). No semantic illusion in the 'Semantic P600' phenomenon: ERP evidence from Mandarin Chinese. *Brain Research*, 1506:76–93.

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8):1117– 1121.

Dikker, S. and Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language*, 127:55–64.

Dikker, S., Rabagliati, H., Farmer, T., and Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science*, 21(5):629–634.

Egidi, G. and Caramazza, A. (2013). Cortical systems for local and global integration in discourse comprehension. *NeuroImage*, 71:59–74.

Embick, D. and Poeppel, D. (2015). Towards a computational(ist) neurobiology of language: correlational, integrated, and explanatory neurolinguistics. *Language, Cognition, and Neuroscience*, 30(4):357– 366.

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, 44(4):491–505.

Ferreira, F., Bailey, K. G. D., and Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1):11–15.

Ferreira, F., and Patson, N. (2007). The 'Good Enough' approach to language comprehension. *Language and Linguistics Compass*, 1(1–2):71–83.

Fossum, V. and Levy, R. (2012). Sequential vs. hierarchical syntactic models of human incremental sentence processing. In *Proceedings of the 3rd Annual Workshop on Cognitive Modeling and Computational Linguistics*. 61–69.

Frank, S. L. and Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, 22(6):829–34.

Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140(0):1–11.

Friederici, A. D., and Gierhan, S. M. E. (2013). The language network. *Current Opinions in Neurobiology*, 23(2):250–4.

Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, 1(3):183–192.

Gibson, E., Bergen, L., and Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of the Sciences USA*, 110(20):8051– 6.

Gouvea, A. C., Phillips, C., Kazanina, N., and Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25(2):149–188.

Groen, W. B., Zwiers, M. P., van der Gaag, R.-J., and Buitelaar, J. K. (2008). The phenotype and neural correlates of language in autism: an integrative review. *Neuroscience Biobehavioral Review*, 32(8):1416–25.

Hagoort, P., Brown, C., and Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8(4):439–483.

Hagoort, P. and Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, 37:347–62.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *North American Chapter of the Association for Computational Linguistics*, 1–8. Association for Computational Linguistics Morristown, NJ, USA.

—— (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4):643–672.

—— (2011). What a rational parser would do. *Cognitive Science*, 35(3):399–443.

—— (2014). Automaton theories of human sentence comprehension. CSLI Publications.

—— (In press). Information-theoretical complexity metrics. *Language and Linguistics Compass*.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. J. (2008). Neurocinematics: the neuroscience of film. *Projections*, 2(1):1–26.

Hauk, O., Davis, M. H., Ford, M., Pulvermuller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage*, 30(4), 1383–1400.

Henderson, J. M., Choi, W., Lowder, M. W., and Ferreira, F. (2016). Language structure in the brain: a fixation-related fMRI study of syntactic surprisal in reading. *NeuroImage*.

Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, 18(4):665–679.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453–458

Jacobs, A. M. (2015). Towards a neurocognitive poetics model of literary reading. In Willems, R. M., editor, *Cognitive neuroscience of natural language use*. Cambridge, UK: Cambridge University Press.

Kelley, E. (2011). Language in ASD. In Fein, D. A., editor, *The neuropsychology of autism*. Oxford: Oxford University Press.

Kemmerer, D. (2014). *Cognitive neuroscience of language*. New York, NY: Psychology Press.

Kim, A., and Osterhout, L. (2005). The independence of combinatory semantic processing: evidence from event-related potentials. *Journal of Memory and Language*, 52:205–225.

Kumar, S., and Penny, W. (2014). Estimating neural response functions from fMRI. *Frontiers in Neuroinformatics*, 8, 48.

Kurby, C. A., and Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and Language*, 126(3):338–49.

Lerner, Y., Honey, C. J., Katkov, M., and Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *Journal of Neurophysiology*, 111(12):2433–44.

Lewis, S., and Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research*, 44(1):27–46.

Lombardo, M. V., Pierce, K., Eyler, L. T., Carter Barnes, C., Ahrens-Barbeau, C., Solso, S., Campbell, K., and Courchesne, E. (2015). Different functional neural substrates for good and poor language outcome in Autism. *Neuron*, 86(2):567–77.

Marcus, M., Marcinkiewicz, M., & Santorini, B. (1993). Building a large annotated corpus of English: The penn treebank. *Computational linguistics*, 19(2), 313–330.

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Sala- mon, G., Dehaene, S., Cohen, L., and Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, 5(4):467–479.

Molinaro, N., Barraza, P., and Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*, 72(0):120–132.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., and Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3(2):151–165.

Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., and Hagoort, P. (2010). Exceptions and anomalies: an ERP study on context sensitivity in autism. *Neuropsychologia*, 48(10):2940– 51.

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. and Embick, D. (2005). Defining the relation between linguistics and neuroscience. In Cutler, A., editor, *Twenty-first century psycholinguistics: four cornerstones*. Mahwah, NJ: Lawrence Erlbaum.

Pylkkänen, L. (2016). Composition of complex meaning: interdisciplinary perspectives on the left anterior temporal lobe. In Hickok, G. and Small, S., editors, *Neurobiology of language*. Academic Press, London: Elsevier.

Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* pp. 324–333.

Roberts, T. P. L., Cannon, K. M., Tavabi, K., Blaskey, L., Khan, S. Y., Monroe, J. F., Qasmieh, S., Levy, S. E., and Edgar, J. C. (2011). Auditory magnetic mismatch field latency: a biomarker for language impairment in Autism. *Biological Psychiatry*, 70(3):263–269.

Rogalsky, C., and Hickok, G. (2010). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience*, 23(7):1–17.

Sanford, A., and Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*, 6(9):382.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., and Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences USA*, 111(43):E4687–96.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Current Biology*, 19(8):661–7.

Speer, N. K., Reynolds, J. R., Swallow, K. M., and Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, 20(8):989–99.

Sportiche, D., Koopman, H., and Stabler, E. (2013). *An introduction to syntactic analysis and theory*. West Sussex: Wiley-Blackwell.

Stabler, E. (1997). Derivational minimalism. In Retoré, editor, *Logical aspects of computational linguistics*, 68–95. Springer.

Stephens, G. J., Silbert, L. J., and Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences USA*, 107(32):14425–30.

Stowe, L. A., Broere, C. A., Paans, A. M., Wijers, A. A., Mulder, G., Vaalburg, W., & Zwarts, F. (1998). Localizing components of a complex task: Sentence processing and working memory. *NeuroReport*, 9(13), 2995–2999.

Tager-Flusberg, H., and Caronna, E. (2007). Language disorders: autism and other pervasive developmental disorders. *Pediatric Clinics of North America*, 54(3):469–81.

Van Petten, C., and Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2):176–90.

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A., and Mitchell, T. (2014). Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. *PLoS One*, 9(11):e112575.

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., and Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *NeuroImage*, 47(1):360–6.

Willems, R. (2013). Can literary studies contribute to cognitive neuroscience? *Journal of LIterary Semantics*, 42(2):217–222.

Willems, R. M., editor (2015). *Cognitive neuroscience of natural language use*. Cambridge, UK: Cambridge University Press.

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., and van den Bosch, A. (2015). Prediction during natural language comprehension. *Cerebral Cortex*. doi: 10.1093/cercor/bhv075.

Wilson, S. M., Molnar-Szakacs, I., and Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, 18(1):230–242.

Xu, J., Kemeny, S., Park, G., Frattali, C., and Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. *NeuroImage*, 25(3):1002–1015.

Yarkoni, T., Speer, N. K., Balota, D. A., McAvoy, M. P., and Zacks, J. M. (2008). Pictures of a thousand words: investigating the neural mechanisms of reading with extremely rapid event-related fMRI. *NeuroImage*, 42(2):973–87.