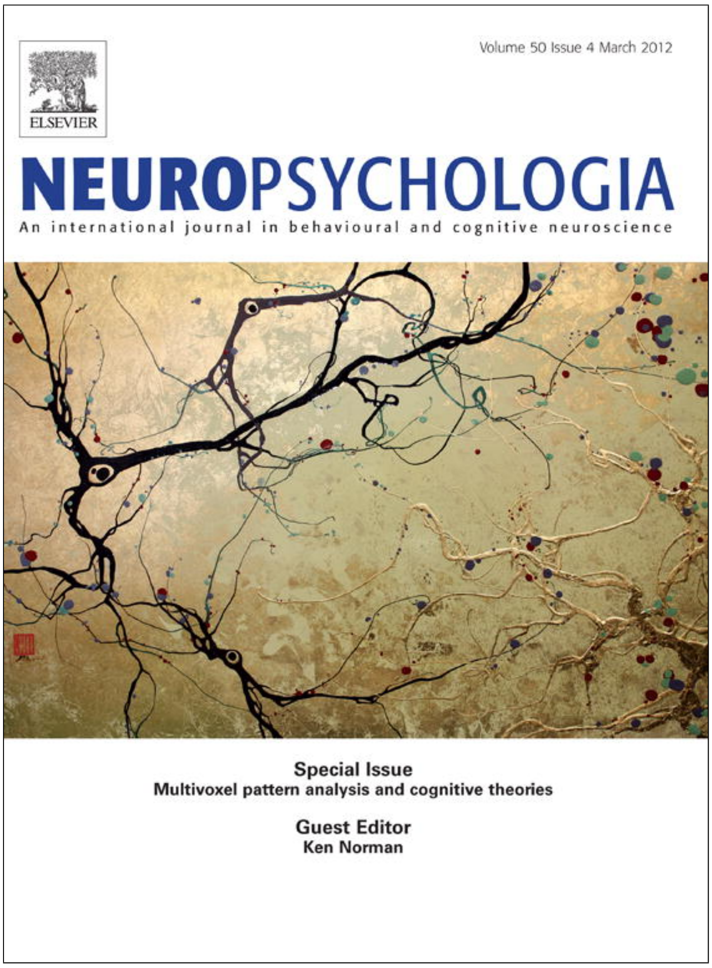


Volume 50 Issue 4 March 2012

ELSEVIER

NEUROPSYCHOLOGIA

An international journal in behavioural and cognitive neuroscience

Special Issue
Multivoxel pattern analysis and cognitive theories

Guest Editor
Ken Norman

Neuropsychologia 50 (2012) 499–513

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses

Evelina Fedorenko*, Alfonso Nieto-Castañon, Nancy Kanwisher

Brain & Cognitive Sciences Department, MIT, 43 Vassar Street, 46-4141C, Cambridge, MA 02139, United States

ARTICLE INFO

Article history:
Received 16 April 2011
Received in revised form 18 August 2011
Accepted 12 September 2011
Available online 17 September 2011

Keyword:
fMRI MVPA language syntax lexical processing

ABSTRACT

Work in theoretical linguistics and psycholinguistics suggests that human linguistic knowledge forms a continuum between individual lexical items and abstract syntactic representations, with most linguistic representations falling between the two extremes and taking the form of lexical items stored together with the syntactic/semantic contexts in which they frequently occur. Neuroimaging evidence further suggests that no brain region is selectively sensitive to only lexical information or only syntactic information. Instead, all the key brain regions that support high-level linguistic processing have been implicated in both lexical and syntactic processing, suggesting that our linguistic knowledge is plausibly represented in a distributed fashion in these brain regions. Given this distributed nature of linguistic representations, multi-voxel pattern analyses (MVPAs) can help uncover important functional properties of the language system. In the current study we use MVPAs to ask two questions: (1) Do language brain regions differ in how robustly they represent lexical vs. syntactic information? and (2) Do any of the language bran regions distinguish between "pure" lexical information (lists of words) and "pure" abstract syntactic information (jabberwocky sentences) in the pattern of activity? We show that lexical information is represented more robustly than syntactic information across many language regions (with no language region showing the opposite pattern), as evidenced by a better discrimination between conditions that differ along the lexical dimension (sentences vs. jabberwocky, and word lists vs. nonword lists) than between conditions that differ along the syntactic dimension (sentences vs. word lists, and jabberwocky vs. nonword lists). This result suggests that lexical information may play a more critical role than syntax in the representation of linguistic meaning. We also show that several language regions reliably discriminate between "pure" lexical information and "pure" abstract syntactic information in their patterns of neural activity.

1. Introduction

A primary goal of language research is to understand the representations (data structures) and the computations (algorithms) that enable us to produce and understand language. However, these are difficult questions, and even in domains like vision, where we have access to animal models and single-cell recording data, we are only beginning to get glimpses of what the representations might look like that allow us to recognize an object or a face (e.g., DiCarlo & Cox, 2007; Freiwald & Tsao, 2010; see Kanwisher, 2010, for discussion). Nevertheless, behavioral and neuroimaging investigations can place constraints on the architecture of the language system by revealing (i) cognitive and neural dissociations between specific mental processes, (ii) the time-course of each mental process, and (iii) whether specific pairs of mental processes are independent or interactive.

The current paper is concerned with the processing of two kinds of information present in the linguistic signal: lexical information and syntactic information. In the rest of the Introduction, we discuss what is currently known about the nature of our linguistic representations (Section 1.1), and about the spatial organization of these representations in the brain (Section 1.2). Based on work in theoretical linguistics and available evidence from psycholinguistic investigations, we argue that our linguistic knowledge forms a continuum between individual lexical items and abstract syntactic representations, with most linguistic representations falling between the two extremes and taking the form of lexical items stored together with the syntactic/semantic contexts in which they frequently occur (the context could further vary in the degree of abstractness from a string of specific lexical items to an abstract rule or part of a rule stored in terms of syntactic categories). Based on the available neuroimaging evidence, we argue that these linguistic representations are organized in a distributed fashion throughout the language system, such that no brain region is selectively sensitive to only lexical or only abstract syntactic information.

---

* Corresponding author. Tel.: +1 617 253 8423.
E-mail address: Evelina9@mit.edu (E. Fedorenko).

0028-3932/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2011.09.014

Given this distributed nature of linguistic representations, we use multi-voxel pattern analyses (Haxby et al., 2001) to ask two questions: (1) Do language brain regions differ in how robustly they represent lexical vs. syntactic information? and (2) Do any of the language brain regions distinguish between conditions that contain "pure" lexical information vs. "pure" syntactic information? In Section 1.3 we provide a brief general introduction to multi-voxel pattern analyses.

1.1. *The nature of linguistic representations*

Knowing the meanings of individual words and understanding how these words can combine with one another to create new, complex meanings are core components of our language knowledge. In the early days of language research, the lexicon and syntactic rules were conceived of as distinct components of the human cognitive architecture. However, over the years, the distinction between the lexicon and syntax has become blurred. We briefly review the historical trajectory of this relationship in the theoretical linguistic, computational linguistic, and psycholinguistic communities.

Theoretical linguistics. At the outset of the generative linguistics enterprise (initiated by the "Chomskyan revolution" that took place in the late 1950s/early 1960s), syntax was conceived of as a set of abstract rules for combining words into complex structures. The grammatical frameworks developed in this tradition did not incorporate lexical information into the syntactic rules (e.g., "Aspects", Chomsky, 1965). However, throughout the 1960s, 1970s and 1980s language researchers observed that fine-grained properties of words have consequences for how words can combine with other words in various syntactic constructions ("X-bar theory", Jackendoff, 1977; "Government & Binding", Chomsky, 1981), making it clear that abstract rules based on syntactic categories alone (like noun or verb) are not the most natural way to characterize the richness of linguistic knowledge. In the 1980s and 1990s a number of grammatical frameworks were developed where lexical knowledge was tightly integrated with combinatorial rules (e.g., LFG, Bresnan, 1982; HPSG, Pollard & Sag, 1994), thus blurring the distinction between the lexicon and syntax. Most recent grammatical frameworks do not draw a sharp distinction between the lexicon and syntax. Instead, lexical entries contain rich information about how a given word can combine with other words across various constructions, and the abstract combinatorial rules are reduced to a minimum (e.g., Bybee, 1998; Culicover & Jackendoff, 2005; Goldberg, 1995; Jackendoff, 2002, 2007; Joshi, Levy, & Takahashi, 1975; Schabes, Abeille, & Joshi, 1988).

Computational linguistics. In parallel with the efforts in the field of theoretical linguistics to construct grammars that best capture our knowledge of language, in the 1980s and 1990s computational linguists worked on parsers that could provide syntactic analyses for arbitrary sentences. Initial attempts that used grammars akin to those in Chomsky's earliest proposals that do not track and make use of phrasal co-occurrence relations among words or between words and syntactic categories (e.g., simple context-free grammars) were not successful. However, when lexical knowledge was taken into consideration (e.g., Charniak, 1996, 1997; Collins, 1996, 1997; Magerman, 1994), performance greatly improved, suggesting that accurate analysis of new linguistic input is only possible when fine-grained information about co-occurrences among words, and/or between words and syntactic contexts, is incorporated into the parser.

Psycholinguistics. In the early days of psycholinguistic research in the area of sentence processing, the lexicon and syntax were treated as two distinct information sources that are used to extract meaning from the linguistic signal.[1] However, many studies have now shown that comprehenders are sensitive not only to unigram frequencies of lexical items (e.g., Morton, 1969; Rayner & Duffy, 1986) or to overall frequencies of particular syntactic constructions (e.g., Frazier & Fodor, 1978), but also to frequencies of lexical items in particular syntactic contexts. For example, comprehenders are sensitive to frequencies of words occurring in specific syntactic constructions, and to co-occurrences between verbs and particular arguments (e.g., Clifton, Frazier, & Connine, 1984; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Gennari & MacDonald, 2008; MacDonald, Pearlmutter, & Seidenberg, 1994; Reali & Christiansen, 2007; Traxler, Morris, & Seely, 2002; Trueswell, Tanenhaus, & Garnsey, 1994). Such findings suggest that comprehenders keep track of co-occurrences at different grain sizes, crossing the boundary between individual lexical items and abstract combinatorial rules.

In summary, most current proposals in theoretical linguistics agree that lexical knowledge must be tightly incorporated into the grammar in order to account for the richness of people's linguistic knowledge, and much psycholinguistic evidence suggests that comprehenders keep track of fine-grained distributional information about lexical items and the syntactic contexts in which they occur.

In spite of this highly integrated view of the lexicon and syntax, it is worth pointing out that some syntactic abstraction – beyond particular lexical items or groups of lexical items – must exist in the language system. One kind of evidence comes from the processing of jabberwocky sentences (e.g., "All mimsy were the borogoves, and the mome raths outgrabe"; Carroll, 1871). For example, Kako (2006) showed that people can make grammaticality judgments on jabberwocky materials (see also Yamada & Neville, 2007, for evidence of ERP responses to syntactic violations in jabberwocky sentences, and see also Fedorenko, Frank, & Gibson, 2009, for evidence of syntactic complexity effects in jabberwocky sentences). Similarly, some representations are plausibly purely lexical with little/no syntactic information associated with them (e.g., certain nouns may have this form). Let us now consider this view of linguistic knowledge in the context of what is known about the neural architecture of language.

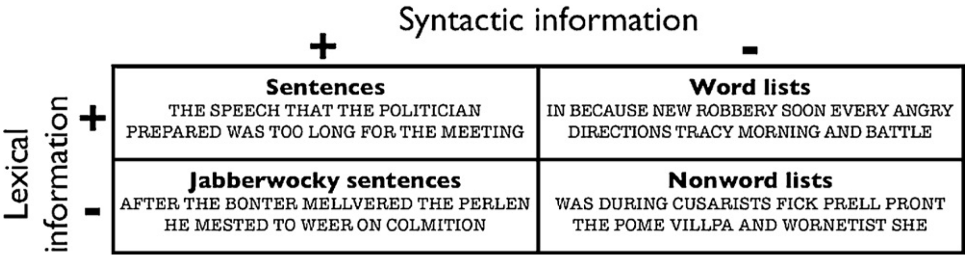
1.2. *The organization of linguistic knowledge in the brain*

If our knowledge of language indeed consists of some "pure" lexical representations and some abstract syntactic representations, with most representations falling in between and taking the form of combinations of lexical items and syntactic rules or part-rules associated with those lexical items, then we might expect our language system to contain some brain regions that are selectively sensitive to lexical information, other brain regions that are selectively sensitive to abstract syntactic information, with perhaps most regions being sensitive to both lexical and syntactic information.

Before discussing the available evidence, let us define what it would mean for a brain region to be sensitive, as well as *selectively* sensitive, to lexical vs. to syntactic information. To do so, let us consider four conditions, which have been commonly used in previous neuroimaging investigations to ask questions about the brain basis of lexical and syntactic processing (e.g., Friederici, Meyer, & von Cramon, 2000; Humphries, Binder, Medler, & Liebenthal, 2006, among others): sentences, lists of

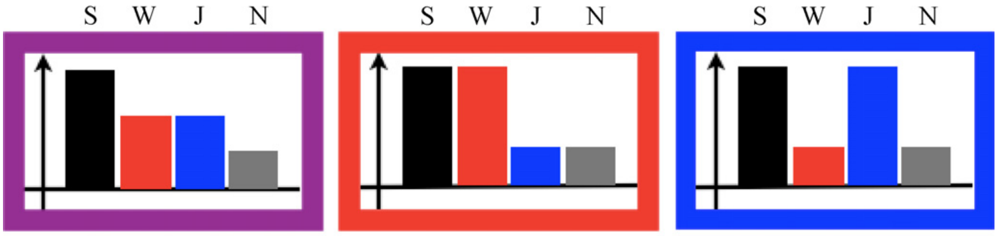
---

[1] In fact, some early studies argued that syntactic information is somehow privileged (i.e., considered at an earlier point in time and/or weighted more strongly) in deriving the meaning of an utterance (e.g., Frazier, 1978; Ferreira and Clifton, 1986). Later studies have shown, however, that this is not the case: lexical information is used as early as can be measured and can guide interpretation, just like syntactic information (e.g., Trueswell et al., 1994).



	Syntactic information +	Syntactic information -
Lexical information +	**Sentences** THE SPEECH THAT THE POLITICIAN PREPARED WAS TOO LONG FOR THE MEETING	**Word lists** IN BECAUSE NEW ROBBERY SOON EVERY ANGRY DIRECTIONS TRACY MORNING AND BATTLE
Lexical information -	**Jabberwocky sentences** AFTER THE BONTER MELLVERED THE PERLEN HE MESTED TO WEER ON COLMITION	**Nonword lists** WAS DURING CUSARISTS FICK PRELL PRONT THE POME VILLPA AND WORNETIST SHE

Fig. 1. Four experimental conditions that have been used extensively in previous neuroimaging studies to study lexical and syntactic processes. Sample items are taken from Fedorenko et al. (2010; Experiment 1).



A voxel/region sensitive to both lexical and syntactic information

A voxel/region sensitive to lexical information only

A voxel/region sensitive to syntactic information only

Fig. 2. Idealized functional profiles of response for a voxel or a region that is sensitive to both lexical and syntactic information (purple frame), to lexical information only (red frame), and to syntactic information only (blue frame). S = Sentences, W = Word-lists, J = Jabberwocky sentences, N = Nonword lists.

unconnected words, jabberwocky sentences and lists of unconnected pronounceable nonwords. These four conditions can be conceived of in a 2 × 2 design (see Fig. 1) crossing the presence/absence of lexical information (present in the Sentences and Word-lists conditions, absent in the Jabberwocky and Nonword-lists conditions) with the presence/absence of syntactic information (present in the Sentences and Jabberwocky conditions, absent in the Word-lists and Nonword-lists conditions).

In principle then, using a contrast between +lexical and −lexical conditions we should be able to find regions sensitive to lexical-level information, and using a contrast between +syntactic and −syntactic conditions we should be able to find regions sensitive to combinatorial (syntactic and/or semantic) information (for brevity, we will refer to this information as "syntactic" in the rest of the paper). Of course, these contrasts may not target *only* the regions sensitive to lexical and syntactic information, respectively. For example, some regions may respond more to +lexical and/or +syntactic conditions because those conditions are more engaging. Furthermore, different language researchers may have strong opinions about why these functional contrasts are not ideally suited for investigating lexical and/or syntactic processing in the brain, and/or about other kinds of contrasts that may be more appropriate. Nevertheless, these simple and maybe somewhat crude conditions are highly useful for investigating lexical and syntactic processing because clear predictions can be made about the responses that these conditions should elicit in brain regions with different functional properties. For example, most researchers would agree that a brain region that stores lexical knowledge should show a stronger response to words than nonwords, and a region that stores syntactic knowledge should show a stronger response to sentences than lists of unconnected words, and to jabberwocky than lists of nonwords (see e.g., Rogalsky & Hickok, 2011, for similar arguments). We therefore use these conditions in the current investigation.

Furthermore, if a brain region is *selectively sensitive* to lexical information, then it should be relatively insensitive to the presence/absence of structural information (e.g., Fig. 2, red frame), and if a brain region is *selectively sensitive* to abstract syntactic information, then it should be relatively insensitive to the presence/absence of lexical meanings (e.g., Fig. 2, blue frame). (Of course, if a brain region is sensitive to both lexical and syntactic information, it should show a profile like the one in Fig. 2, purple frame.)

Let us now turn to the available neuroimaging evidence. Inspired by the early theorizing in the field of language research, numerous PET and fMRI studies have looked for neural dissociations between lexico-semantic and syntactic processing (e.g., Dapretto & Bookheimer, 1999; Friederici et al., 2000; Humphries et al., 2006; Mazoyer et al., 1993; Stowe et al., 1999; Vandenberghe, Nobre, & Price, 2002, among others), and many of these claimed to have observed such dissociations. However, considering the literature as a whole, no brain region has been convincingly shown to selectively respond to lexical information, or to selectively respond to syntactic information.[2] Instead, across studies, all key language regions on the lateral surfaces of left frontal and temporal cortices have been implicated in both lexical and syntactic processing, suggesting that lexical knowledge and syntactic knowledge may be stored in the same brain regions.

Consistent with this picture, our own recent work (Fedorenko, Hsieh, Nieto-Castañon, Whitfield-Gabrieli, & Kanwisher, 2010; see Bedny, Pascual-Leone, Dodell-Feder, Fedorenko, & Saxe, 2011, for similar results) shows that each brain region that is sensitive to high-level aspects of the linguistic signal, as evidenced by a stronger response to sentences than nonword lists (i.e., regions on the lateral

---

[2] It is worth noting that a simple failure to observe activation in some brain region when examining the brain's response to a particular manipulation, especially in traditional group-based analyses, cannot be taken as evidence that said brain region is not sensitive to the manipulation in question. Such failures could be due to insufficient power to detect the effect and/or to variability in the locations of the relevant brain region across subjects (see Fedorenko et al., 2010; Fedorenko, Nieto-Castanon, & Kanwisher, in press; Nieto-Castañon, Kanwisher, & Fedorenko, submitted for publication, for recent discussions). However, the case can be strengthened if (i) the most sensitive methods available are used (e.g., functional regions-of-interest defined in individual subjects), (ii) other brain regions *are* responsive to the manipulation in question, and (iii) the brain regions that are not sensitive to the manipulation in question are sensitive to other manipulations.

surfaces of left frontal and left temporal and parietal cortices, and some additional regions; see Fig. 5), is sensitive to both lexical and syntactic information. In fact, every region within this high-level linguistic processing network shows a similar response profile: the response is strongest to sentences, weaker to word lists and jabberwocky, and weakest to nonword lists (Fig. 2, purple frame). The Word-lists and Jabberwocky conditions elicit a similar-strength BOLD response throughout the language system (see Fig. D1 in Fedorenko et al., 2010; see Bedny et al., 2011, for similar results).

In summary then, among the brain regions within the language network there does not appear to be a robust dissociation between sensitivity to lexical vs. syntactic information. However, the functional profile observed in these regions (i.e., sensitivity to both lexical and syntactic information; Fig. 2, purple frame) could arise from several distinct underlying activation patterns at a finer spatial grain within these regions. We illustrate some of these possibilities in Fig. 3, where each 4 × 4 grid represents a region, and each square in the grid represents a voxel.

One possibility (Fig. 3a) is that these regions are functionally homogeneous, and the profile observed for the region as a whole is also observed in each individual voxel comprising this region (of course, some of these voxels may be relatively more sensitive to lexical information and others to syntactic information). Another possibility, however, is that at a finer spatial grain there is functional heterogeneity, such that some voxels are selectively sensitive to lexical information, and other voxels are selectively sensitive to syntactic information (Fig. 3b–e). This heterogeneity can take the form of sub-regions that have some systematicity in their spatial locations within a region across subjects (Fig. 3b), sub-regions whose spatial locations within a region are arbitrary across subjects[3] (Fig. 3c), or distributed patterns of voxels where voxels that are selectively sensitive to lexical information are spatially interleaved with voxels that are selectively sensitive to syntactic information (Fig. 3d–e).

Distinguishing among these various possibilities is critical for understanding how lexical and syntactic knowledge is organized in the human mind and brain. Some existing evidence renders some of these alternatives less likely. In particular, neuroimaging evidence and evidence from brain-damaged populations does not appear consistent with the existence of sub-regions that are selectively sensitive to lexical vs. syntactic information (i.e., Fig. 3b and c). As discussed above, previous neuroimaging evidence failed to identify brain regions that are selectively sensitive to lexical or to syntactic information. Consistent with this picture, when we were developing a functional "localizer" for high-level language regions (Fedorenko et al., 2010), we spent a great deal of time and effort in a search for regions selectively sensitive to lexical vs. syntactic information. Our efforts included both (i) a careful examination of many individual subjects' activation maps and (ii) applying the group-constrained subject-specific analysis methods (GSS; see Fedorenko et al., 2010, for details) to various functional contrasts (e.g., Word-lists > Nonword-lists, Jabberwocky > Nonword-lists, Word-lists > Jabberwocky, and Jabberwocky > Word-lists). GSS-style analyses are a more sensitive alternative to the traditional random-effects analyses and enable discovering spatially similar regions without the requirement of voxel-level overlap among subjects (a version of these analyses can also look for functionally stable sub-regions within a particular volume without the constraint of spatial systematicity across subjects). None of these analyses revealed regions that are selectively sensitive to lexical information or selectively sensitive to syntactic information (i.e., with response profiles like those shown in the red and blue frames in Fig. 2), with or without the constraint of spatial consistency across individual brains.

Similarly, the existence of such sub-regions predicts the existence of patients with selective difficulties with individual word meanings (with no syntactic difficulties), or selective difficulties in combining words with one another or interpreting meaning units above the word level (with no lexical difficulties). Although a detailed discussion of the available patient evidence is beyond the scope of the current paper, to the best of our knowledge no strong dissociations between difficulties with individual word meanings and difficulties with combinatorial processing have been reported.[4] The absence of such dissociations could of course simply mean that the specialized sub-regions are so small that typical kinds of brain damage (e.g., due to stroke) always affect multiple nearby sub-regions. However, taken in the context of the neuroimaging evidence (where we were unable to discover functionally selective sub-regions in a relatively large dataset, with a lot of data per subject, and sensitive analysis methods), the patient data provide further evidence against the possibilities illustrated in Fig. 3b and c. This leaves us with the possibilities illustrated in Fig. 3a, d and e: distributed representations that either do (Fig. 3d and e) or do not (Fig. 3a) contain voxels that are selectively sensitive to lexical or syntactic information, and that either do (Fig. 3a and e) or do not (Fig. 3d) contain voxels that are sensitive to both lexical and syntactic information. Given these plausibly distributed representations, we can investigate functional properties of language regions using multi-voxel pattern analysis methods (e.g., Haxby et al., 2001; see Section 1.3 below for a brief overview), which look at the spatial patterns of neural activity across voxels in a region in response to different experimental manipulations.

In the current investigation we use pattern analyses to ask two questions. First, we ask whether brain regions engaged in high-level linguistic processing differ in how robustly they represent lexical vs. syntactic information. Although we have found that all the regions that respond to high-level linguistic stimuli are sensitive to both lexical and syntactic information, it is possible that some regions are relatively *more* sensitive to lexical information and other regions are relatively *more* sensitive to syntactic information. Although such a dissociation would be weaker than the strongest possible dissociation with different brain regions *selectively sensitive* to lexical vs. syntactic information, it would still be an important property of the language system. Alternatively, it is possible that one kind of information is represented more robustly than the other kind of information throughout the language system. This may tell us about the relative importance of different kinds of information for representing linguistic meanings.

---

[3] In either the scenario illustrated in Fig. 3b or the one in Fig. 3c there may be sub-regions that are sensitive to both lexical and syntactic information, but the critical feature of these scenarios is that they contain some sub-regions that are selectively sensitive to lexical information, and others that are selectively sensitive to syntactic information.

[4] Agrammatic Broca's aphasics are sometimes discussed as an example of a selective syntactic deficit (e.g., Caramazza and Zurif, 1976; Grodzinsky, 2000; Grodzinsky & Santi, 2008) based on their difficulty with understanding semantically reversible infrequent constructions where syntax is the only cue to meaning. However, Linebarger, Schwartz, and Saffran (1983) have convincingly demonstrated that syntactic knowledge is largely preserved in these patients as evidenced by high performance in a grammaticality judgment task (i.e., deciding whether a sentence is well-formed, according to the rules of the language) on a large number of syntactic constructions. This result suggests that difficulties with reversible passives and other similar constructions in agrammatic aphasics are not due to the loss of syntactic knowledge. In fact, behavioral and neuroimaging work has related these kinds of syntactic complexity effects to domain-general factors like working memory (e.g., Fedorenko, Gibson, & Rohde, 2007; Gibson, 1998; Gordon, Hendrick, & Levine, 2002; McElree, Foraker, & Dyer, 2003), cognitive control (e.g., Novick, Trueswell, & Thompson-Schill, 2005), or general non-verbal intelligence (Gibson & Fedorenko, 2011).

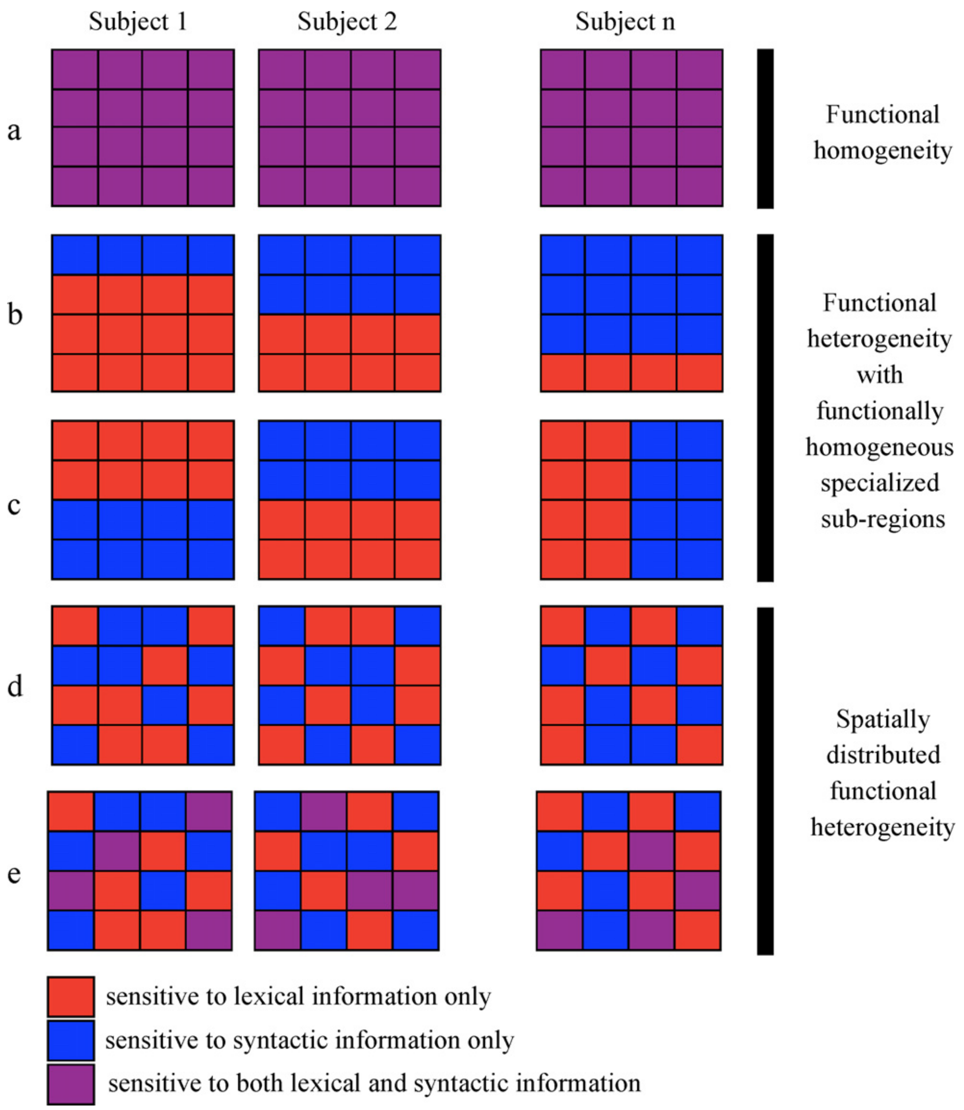


Fig. 3. Five sample scenarios for how a functional profile for a region that is sensitive to both lexical and syntactic information (e.g., purple frame in Fig. 2) can arise. Each 4 × 4 grid represents a brain region, and each square in the grid represents a voxel. Red voxels are sensitive to lexical information only, blue voxels are sensitive to syntactic information only, and purple voxels are sensitive to both lexical and syntactic information.

And second, we ask whether lexical and syntactic information are so tightly integrated that even at the level of individual voxels there is no difference in the degree to which each voxel is sensitive to lexical vs. syntactic information. This question may help us rule one version of the possibility illustrated in Fig. 3a. In general, distinguishing among different kinds of distributed representations is challenging: a failure to find units (voxels in this case) that show a particular functional property (e.g., selective sensitivity to lexical information) is always consistent with the possibility that such selectivity is present at a finer spatial scale (after all, each voxel encompasses a few hundred thousand neurons). However, to the extent that dissociations are discovered, important inferences can be drawn. As discussed above, we found that in the mean BOLD signal, no brain region in the language network can distinguish between “pure” lexical information (the Word-lists condition) and “pure” syntactic information (the Jabberwocky condition). Instead, every region responds to these conditions in a similar way and this response is lower than that elicited by sentences with real words, and higher than that elicited by lists of nonwords (Fig. 2, purple frame). However, if spatial patterns of neural activity within some language region can distinguish between the Word-lists and Jabberwocky conditions, this would argue against the most extreme version of the integration of lexical and syntactic knowledge (i.e., the extreme version of the architecture shown in Fig. 3a, where not only is each voxel sensitive to both lexical and syntactic information but this sensitivity is identical across voxels).

1.3. *MVPAs: a brief introduction*

Traditional fMRI analyses treat each voxel in the brain as an independent data point and examine differences in the strength of the BOLD signal between conditions in each voxel individually or across voxels in some region(s) of interest. In contrast, multi-voxel pattern analyses examine the pattern of activity across multiple voxels (i) across the whole brain or (ii) in some pre-defined anatomical/functional region of interest (ROI) (see Norman, Polyn, Detre, & Haxby, 2006; Schwarzkopf & Rees, 2011, for overviews of these methods). These methods can distinguish between conditions that are not differentiated in terms of the mean BOLD response, and are therefore perfectly suited for asking where in the brain a particular cognitive dimension is represented, or whether a particular brain region is sensitive to a particular dimension of knowledge.

A wide range of classification algorithms developed in the field of machine learning (e.g., Duda, Hart, & Stork, 2001; Kotsiantis,

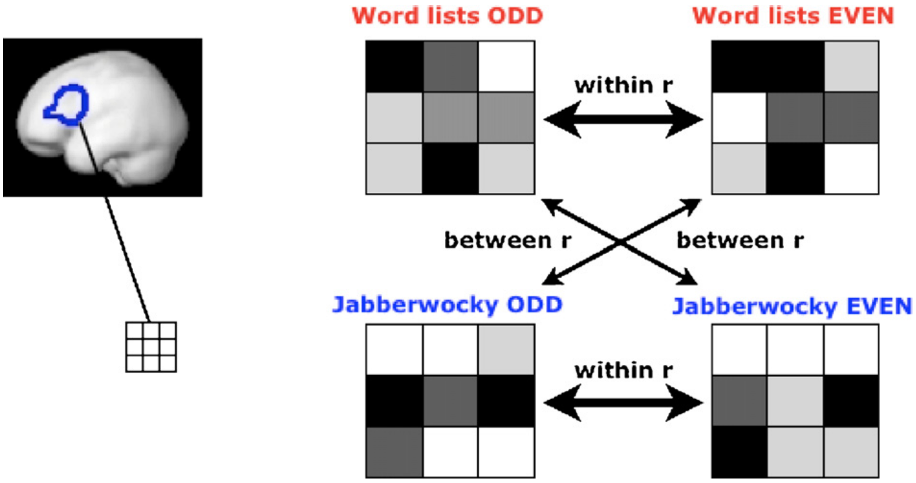


Fig. 4. A schematic illustration of the logic of the correlation-style multi-voxel pattern analyses (Haxby et al., 2001). Adapted from a figure by Julie Golomb.

2007; Pereira, Mitchell, & Botvinick, 2009) can be applied to neuroimaging data. Although complicated algorithms (e.g., nonlinear support vector machines or neural nets with hidden layers) may achieve high classification accuracies, interpreting the classification results in terms of possible underlying patterns of brain activity becomes difficult as the algorithms get more complex (see Schwarzkopf & Rees, 2011, for a discussion). As a result, we here adopt simple linear correlation-based classification methods introduced to fMRI by Haxby et al. (2001). In this method the functional data are separated into two independent sets (e.g., odd vs. even runs). Across these two sets of the data, the similarity between the patterns of activity from the same condition (within-condition correlation) is compared to the similarity between the patterns of activity from different conditions (between-condition correlation). Such comparisons are often performed on some pre-defined region(s) of interest. For example, Haxby et al. (2001) focused on regions within ventral visual cortex that respond to high-level visual categories to investigate the information represented in each region.

Fig. 4 schematically illustrates this correlation-based method. Suppose you want to find out whether a brain region implicated in linguistic processing (e.g., parts of the left inferior frontal gyrus that roughly correspond to Broca's area; Broca, 1861) can discriminate between lists of unconnected words and jabberwocky sentences. Reading or listening to the former involves processing the meanings of individual words. In contrast, the latter involves putting nonwords together according to the syntactic rules and perhaps constructing a coarse semantic representation of who is doing what to whom. As discussed above, these two conditions are not distinguished in the mean BOLD signal in any of the regions within the language network (e.g., Bedny et al., 2011; Fedorenko et al., 2010), eliciting a similar-strength response that is lower than that elicited by sentences and higher than that elicited by unconnected nonwords. However, we can ask whether this distinction is represented in the pattern of activity. We would then divide the data for each condition in half and compare the data halves within each condition vs. between the Word-lists condition and the Jabberwocky condition. If, across subjects, the within-condition correlations are higher than the between-condition correlations, we take this to mean that some aspect of the distinction between word-lists and jabberwocky sentences is indeed represented in the relevant brain region.

Kriegeskorte, Goebel and Bandettini (2006) extended this basic idea to whole-brain methods where a sphere of a certain size (e.g., a few millimeters) is iteratively moved across the brain and within- vs. between-condition similarity values are computed on the voxels located within each sphere. This "searchlight"-style MVPA can then produce whole-brain maps (similar to the activation maps we get with the traditional fMRI methods) that show where in the brain the conditions in question can be discriminated. This style of analysis is nicely complementary to ROI-based MVPAs in much the same way that whole-brain activation maps are complementary to ROI-based analyses in traditional, activation-based, fMRI methods. For example, it can tell us whether we are missing some important regions outside the borders of our ROIs that contain information about our conditions in question. Similarly, it can discover potential heterogeneity within the ROIs, so that e.g., different portions of the ROI represent different dimensions of the stimuli.

2. Materials and methods

We conducted a series of MVPA analyses on a dataset ($n = 25$) whose basic univariate results were reported in Fedorenko et al. (2010). We here provide some basic information about the design and procedure. In a blocked design participants read sentences, lists of words, jabberwocky sentences and lists of nonwords.[5] The sentences were 12 words long (for 12/25 subjects) or 8 words long (for 13/25 subjects). Two non-overlapping sets of linguistic materials were used across two subsets of subjects to ensure generalizability across different sets of materials. Across both sets, the sentences were created so as to include a variety of syntactic constructions and to vary in terms of content. The word lists were created by scrambling the words across sets of sentences, so that "reconstructing" a sentence out of individual words in a word list was not possible. The jabberwocky sentences were created by replacing the content words (nouns, verbs, adjectives and adverbs) in the sentences by pronounceable nonwords that were matched to the words in length (in letters and syllables) and bigram letter frequency. Finally, the nonword lists were created by scrambling the nonwords across sets of jabberwocky sentences in a similar way to how words were scrambled in the Word-lists condition (this method of creating the Word-lists and Nonword-lists conditions – by scrambling the Sentences and Jabberwocky conditions, respectively – means that some function words are present in both of these conditions, as shown in Fig. 1; separate experiments where function words were replaced with length-matched words/nonwords revealed similar activation patterns to those reported in Fedorenko et al. (2010)). 12/25 subjects were run on a passive-reading version of the task, and 13/25 subjects were run on a version of the task where after each sentence/word-list/jabberwocky sentence or

---

[5] In Experiment 3 in Fedorenko et al. (2010), we demonstrated that all the key regions identified with the sentences > nonwords contrast behave similarly regardless of whether materials are presented visually or auditorily, which is to be expected given that the sentences > nonwords contrast is aimed at regions that support high-level aspects of linguistic processing; see also Braze et al. (2011), for recent evidence and a review of modality-independence of brain regions that support high-level linguistic processing.

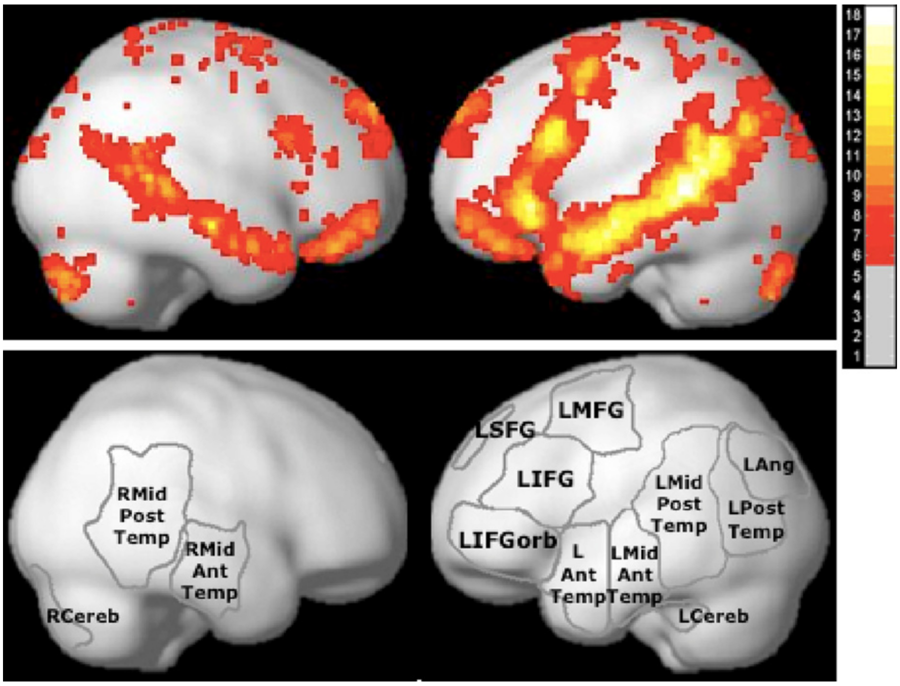


Fig. 5. Top: A probabilistic overlap map showing in each voxel how many of the 25 individual subjects show a significant (at $p < .05$, FDR-corrected) effect for the Sentences > Nonwords contrast. Bottom: The main functional parcels derived from the probabilistic overlap map using an image parcellation (watershed) algorithm, as described in more detail in Fedorenko et al. (2010).

nonword-list a memory probe was presented (a word in the sentences and word-list conditions, and a nonword in the jabberwocky and nonword-list conditions), and participants had to decide whether the probe was present in the preceding stimulus. As discussed in Fedorenko et al. (2010), the two versions of the task (passive reading vs. reading with a memory probe at the end) produced similar activation patterns; we therefore collapsed across the two subsets of the subjects in our analyses in that paper and we do the same here. Each participant completed between 6 and 8 runs (i.e., between 24 and 32 blocks per condition; see Fedorenko et al., 2010, for details of the timing).

In Section 3, we report the results of: (a) region-of-interest-based (ROI-based) MVPA analyses on a set of key language-sensitive regions and (b) whole-brain searchlight-style analyses (Kriegeskorte et al., 2006).

2.1. ROI-based analyses

We chose to use as ROIs for our MVPA analyses the thirteen group-level functional parcels[6] (Fig. 5, bottom) that were derived from the probabilistic overlap map for the Sentences > Nonword-lists activations[7] (Fig. 5, top), as described in Fedorenko et al. (2010). These group-based ROIs represent the locations where individual activations tend to be found most consistently across subjects. So, for any given subject, a parcel will include some voxels that respond reliably more strongly to Sentences than Nonwords, and some voxels that do not show this property. We chose to use these group-level parcels instead of subject-specific functional ROIs in these analyses for two reasons. First, it has been previously demonstrated (Haxby et al., 2001; Kriegeskorte et al., 2006) that even voxels that do not show a particular functional signature relevant to the to-be-discriminated conditions can contribute to classification accuracy. For example, Haxby et al. (2001) showed that removing voxels from the ventral visual regions that respond most strongly to some visual category does not strongly affect the ability to discriminate that category from other categories. Consequently, voxels in the vicinity of language-sensitive regions in each individual subject may contain information about various aspects of linguistic knowledge even though they do not show the functional signature of language-sensitive voxels. And second, because we wanted to examine neural activity patterns across all four conditions, we could not use any of the conditions for defining subject-specific fROIs. (However, in addition to these whole-parcel-based analyses, we did conduct one analysis where we looked at the ability of subject-specific functional ROIs (fROIs), defined by the Sentences > Nonword-lists contrast, to discriminate between word lists and jabberwocky sentences. The results of this analysis are reported in Appendix A.)

For each condition we divided the data into odd-numbered and even-numbered runs (each subject performed between 6 and 8 runs total). Then, for each subject and for each ROI, and across the two independent halves of the data, we computed the within- vs. between-condition spatial correlations for each pair of conditions (as schematically shown in Fig. 4 above), considering all the voxels within the parcel. For example, to see how well the pattern of activity for the Sentences condition is discriminated from the pattern of activity for the Word-lists condition, we computed (i) a within-condition correlation value for the Sentences condition by comparing the pattern of activity for the Sentences condition in the odd vs. even runs (all the r values are Fisher-transformed); (ii) a within-condition correlation value for the Word-lists condition by comparing the pattern of activity for the Word-lists condition in the odd vs. even runs; and (iii) a between-condition correlation value by comparing the pattern of activation for the Sentences condition in the odd/even runs and for the Word-lists condition in the even/odd runs (these two values are averaged to create one between-condition value). Finally, for each ROI we performed an F-test on the within vs. between-condition correlation values across subjects to see whether the within-condition values are reliably higher than the between-condition values. If so, this would suggest that the distinction between the two conditions in question is represented in the relevant ROI.

We deviated from Haxby's analysis strategies in one way. In particular, Haxby applied centering to his data by subtracting the mean level of activation of a voxel from the activation level for each of the conditions. This is equivalent to considering the activation from each condition with respect to a baseline activation level computed as the average activation across all conditions, instead of using an independent fixation baseline as we used in our analyses. The centering procedure potentially increases sensitivity of the MVPAs by removing one source of variance from across the voxels and leaving only between-condition differences in play. However, centering also introduces between-condition dependencies in the estimation of the within-condition similarity measures, which complicates their interpretation.

---

[6] These parcels were created in order to systematize and automate the procedure for defining subject-specific functional ROIs (fROIs): in particular, for any given region, an individual subject's fROI is defined by intersecting the relevant parcel with the subject's thresholded activation map. In other words, these functional parcels serve as spatial constraints on the selection of subject-specific voxels, akin to using borders of anatomical regions (see Julian, Fedorenko, & Kanwisher, submitted, for an extension of this method to ventral visual regions).

[7] Although these group-level functional parcels were created from the 25 subjects whose data we examine here, non-independence issues (Vul & Kanwisher, 2009) do not arise in examining the discriminability between word lists and jabberwocky sentences because the data from those conditions were not used in creating the parcels. Some non-independence is present when we examine the discriminability among all four conditions (Section 3.1). This non-independence should be taken into consideration when interpreting the results from the ROI-based analyses. However, the fact that the results of the whole-brain searchlight analyses, which do not suffer from such non-independence problems, look similar to those of the ROI-based analyses largely alleviates the concerns.

Table 1

F and *p* values for the tests evaluating the discriminability for different pairs of conditions. Degrees of freedom are (2,23) in all cases. Asterisks indicate the comparisons that survive the FDR correction for the number of ROIs. Sentences vs. Jabberwocky (S vs. J) and Word-lists vs. Nonword-lists (W vs N) contrasts involve pairs of conditions that differ along the lexical dimension, and Sentences vs. Word-lists (S vs. W) and Jabberwocky vs. Nonword-lists (J vs. N) contrasts involve pairs of conditions that differ along the syntactic dimension.

	S vs. N	S vs. J	W vs. N	S vs. W	J vs. N	W vs. J
LIFGorb	20.1; <.001*	18.2; <.001*	11.9; <.001*	4.57; <.05*	2.70; .088	6.38; <.01*
LIFG	39.7; <.001*	25.1; <.001*	7.28; <.005*	15.5; <001*	6.77; <.005	4.69; <.05
LMFG	27.3; <.001*	18.0; <.001*	8.61; <.005*	8.07; <.005*	<1; n.s.	4.19; <.05
LSFG	16.5; <.001*	9.73; <.001*	2.43; n.s.	5.73; <.01*	<1; n.s.	<1; n.s.
LAntTemp	15.1; <.001*	13.4; <.001*	4.29; <.05*	8.13; <.005*	<1; n.s.	1.52; n.s.
LMidAntTemp	41.6; <.001*	27.5; <.001*	16.0; <.001*	16.7; <.001*	<1; n.s.	1.56; n.s.
LMidPostTemp	74.2; <.001*	42.3; <.001*	32.1; <.001*	15.4; <.001*	1.71; n.s.	9.79; <.001*
LPostTemp	31.5; <.001*	32.1; <.001*	11.2; <.001*	19.2; <.001*	<1; n.s.	8.21; <.005*
LAngG	15.8; <.001*	16.3; <.001*	7.89; <.005*	7.09; <.005*	<1; n.s.	5.55; <.05*
RMidAntTemp	21.1; <.001*	7.03; <.005*	1.61; n.s.	6.29; <.01*	<1; n.s.	<1; n.s.
RMidPostTemp	11.5; <.001*	7.63; <.005*	3.52; <.05	5.65; <.05*	<1; n.s.	3.15; .061
RCereb	22.0; <.001*	16.8; <.001*	<1; n.s.	7.11; <.005*	<1; n.s.	1.50; n.s.
LCereb	9.02; <.005*	5.23; <.05*	4.42; <.05*	5.69; <.01*	2.74; .086	3.63; <.05

In all the analyses, we report uncorrected significance values and indicate the ones that survive the FDR correction.

2.2. Whole-brain searchlight-style analyses

As discussed in the introduction, whole-brain searchlight-style analyses are a useful complement to ROI-based analyses. In these analyses, a sphere of a certain size is moved iteratively across the brain and within- vs. between-condition correlation values are computed on the voxels located within each sphere. Then, for each voxel (at the center of the sphere), we can obtain a measure of how well surrounding neural tissue can distinguish between some conditions of interest in the pattern of activity. The group-level statistics in this method are computed as described above, i.e., by performing *F*-tests on the within- vs. between-condition correlation values within each sphere.

Instead of using a spherical ROI (with hard edges), as in the original Kriegeskorte et al. (2006) analyses, we used a probabilistic ROI defined as a Gaussian kernel (8 mm FWHM). The use of a Gaussian kernel decreases the MVPA method's dependency on the choice of ROI size, by smoothly de-emphasizing the influence of voxels at increasing distances from the reference voxel when computing the spatial correlation measures (see Appendix B for implementation details).

We applied the searchlight method to individual activation maps that have been previously smoothed with a 4 mm kernel. Individual searchlight maps were then subjected to a second-level analysis, with an *F*-test performed on the within- vs. between-condition correlations values within each kernel.

For the comparisons among all four conditions (Section 3.1) we are interested in both ROI-based and searchlight analysis results and treat them as equally informative. However, for the comparison between word lists and jabberwocky sentences (Section 3.2), we present the searchlight results for completeness only. Our main analyses in that section are ROI-based analyses because (a) there is a strong a priori expectation that these will be the regions that distinguish between lexical and syntactic information and (b) ROI-based analyses have considerably more power than whole-brain analyses (see e.g., Saxe, Brett, & Kanwisher, 2006, for discussion).

3. Results and discussion

3.1. Do language brain regions differ in how robustly they represent lexical vs. syntactic information in the spatial patterns of neural activity?

We here consider the pattern information across all four conditions: sentences, lists of words, jabberwocky sentences, and lists of nonwords. The results of the ROI-based analyses for different pairs of conditions are presented in Table 1.

In Fig. 6 (left) we present the whole-brain maps from the searchlight analyses. We present the results for the contrast between Sentences and Nonword lists, for two contrasts that differ along the lexical dimension (S vs. J and W vs. N), and for two contrasts that differ along the syntactic dimension (S vs. W and J vs. N). These analyses tell us where in the brain the different pairs of conditions are reliably discriminated in the pattern of activity. Alongside the searchlight (information-based) maps we present activation-based maps from the random-effects group analyses, for comparison.[8]

Consistent with Kriegeskorte et al.'s (2006) findings in the ventral visual cortex, the information-based maps look similar to the traditional activation-based maps (but see fn. 8). Furthermore, the results of the searchlight analyses closely mirror those from the ROI-based analyses, albeit being somewhat weaker.[9]

The question we asked in this section is whether different brain regions engaged in high-level linguistic processing differ in how robustly they represent lexical vs. syntactic information. As discussed above, although the mean BOLD measures show that each of the language regions is sensitive to both lexical and syntactic information, it is possible that spatial patterns of neural activity within each region would reveal that some of these regions represent lexical information more robustly while other brain regions represent syntactic information more robustly. If that turned out to be the case, then we would still be able to make a case for a functional dissociation between lexical and syntactic information at some level. It would not be a dissociation in the standard sense of the word where two brain regions support distinct mental processes (and consequently, each can be selectively damaged/preserved), but this kind of a data pattern would tell us that different regions within the language system may be relatively more sensitive to one aspect of the linguistic signal than other aspects (e.g., more sensitive to individual word meanings than the structure, or vice versa).

However, we do not find such a pattern of results. Instead, across the different ROIs, pairs of conditions that differ along the lexical dimension (i.e., S vs. J and W vs. N) are distinguished better than pairs of conditions that differ along the

---

[8] As noted above and discussed in our previous work, group maps are often not a good summary representation of individual subjects' data. In particular, activations at the individual subject level are often more robust and extensive than those in group-level maps. So, in order to directly compare the power of information-based analyses to that of the traditional activation-based analyses (as e.g., Kriegeskorte et al., 2006, did in their original paper), we would need to examine individual maps for both analyses. Furthermore, we would want to run the searchlight analyses on the unsmoothed individual maps (cf. the current analyses, which are run on maps smoothed with a 4 mm kernel). Nevertheless, we present the two kinds of whole-brain maps side by side for our key contrasts in order to show the broad similarities between them.

[9] It is not surprising that group-based searchlight-style analyses produce weaker results than ROI-based analyses. As discussed extensively in Fedorenko et al. (2010; see also Fedorenko et al., in press; Fedorenko & Kanwisher, 2011; Saxe et al., 2006), due to anatomical variability across brains language activations do not line up well, often making it difficult to detect group-level effects in traditional activation-based analyses even in cases where every subject shows activation in the vicinity of some anatomical landmark. This problem applies to group-based whole-brain MVPAs in a similar way.

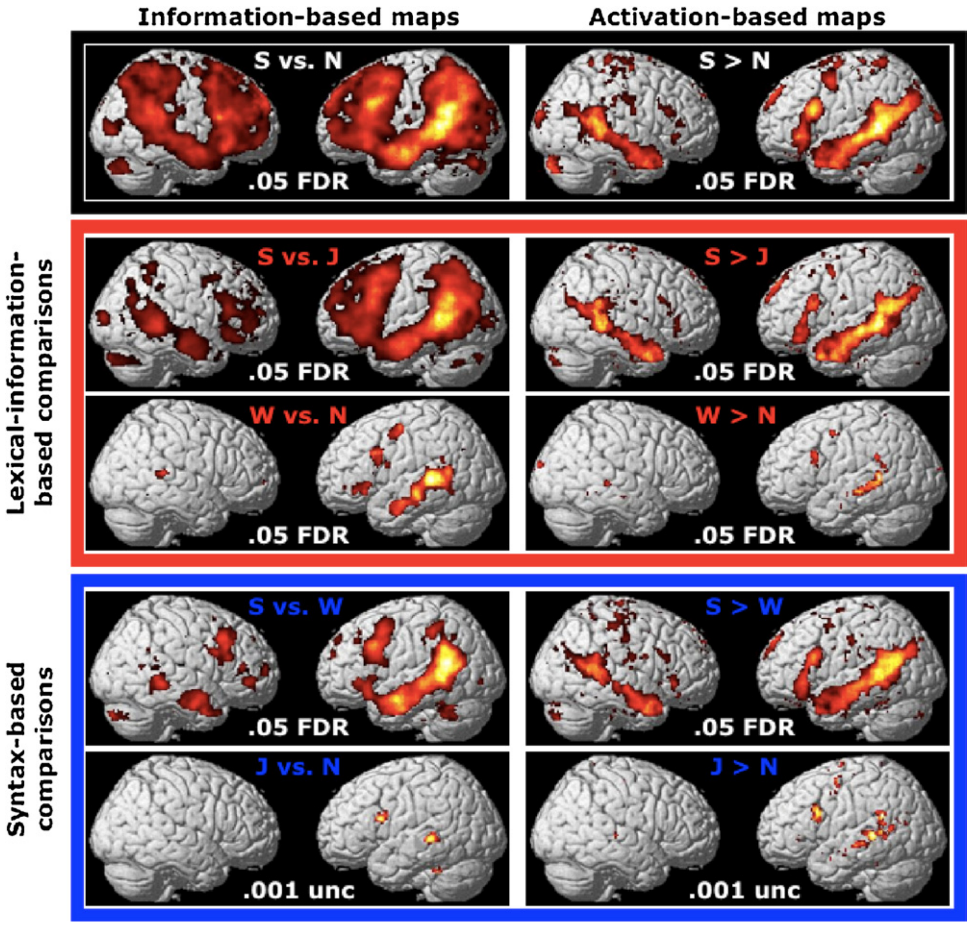


Fig. 6. Left: The results of the searchlight analyses showing discriminability between Sentences and Nonword-lists conditions (black frame), and pairs of conditions that differ along the lexical dimension (S vs. J and W vs. N; red frame) or along the syntactic dimension (S vs. W and J vs. N; blue frame). [Note of course that the contrasts that involve the Sentences condition – S vs. N, S vs. J and S vs. W – additionally involve compositional semantic processes, which may be contributing to discriminability.] Right: The group-level (random-effects) activation maps for the corresponding contrasts. All the maps are thresholded at $p < .05$, FDR-corrected (i.e., the darker red colors show voxels that reach this significance level, with the brighter colors showing voxels that reach higher significance levels), except for the Jabberwocky > Nonwords maps which are thresholded at $p < .001$, uncorrected, because no voxels emerged – either for the searchlight- or activation-based analysis – for this contrast at the FDR .05 threshold.

syntactic dimension (i.e., S vs. W and J vs. N). To quantify this observation we computed interaction statistics comparing the within- and between-condition correlation values for (a) pairs of conditions that differ along the lexical dimension (i.e., S vs. J and W vs. N) and (b) pairs of conditions that differ along the syntactic dimension (i.e., S vs. W and J vs. N). The results are shown in Table 2 and Fig. 7. We see that all of the regions show a trend such that pairs that differ along the lexical dimension are discriminated better than

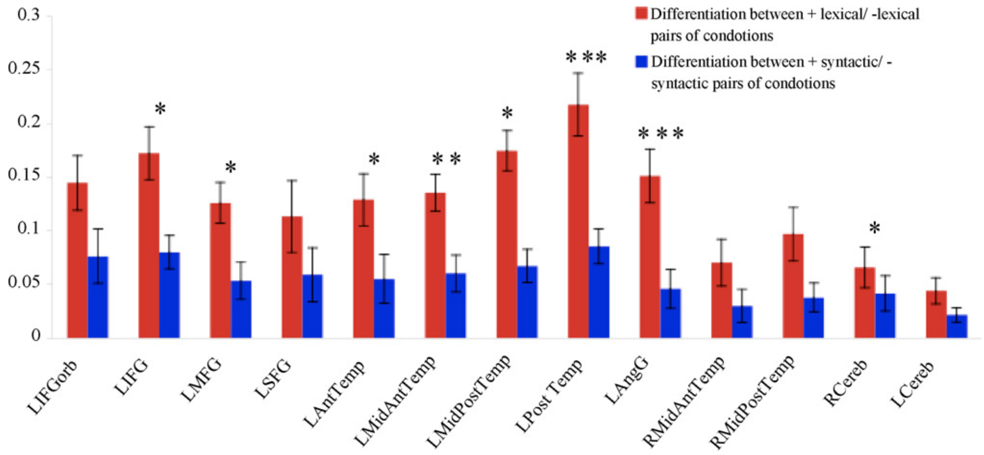


Fig. 7. Differences between within- and between-condition correlation values for pairs of conditions that differ along the lexical dimension (red bars) and pairs of conditions that differ along the syntactic dimension (blue bars). For each ROI and each subject we averaged the within- vs. between- difference scores for pairs of conditions that differ along the lexical dimension (i.e., S-SJ, W-WN, J-JS, N-NW) and for pairs of conditions that differ along the syntactic dimension (i.e., S-SW, W-WS, J-JN, N-NJ). We then averaged these values across subjects for each ROI. Error bars represent standard errors of the mean over subjects. Asterisks indicate significance levels for the interaction (see Table 2 for details): *<.05; **<.005; ***<.001.

Table 2
F and *p* values for the interaction tests comparing the within- and between-condition correlation values for pairs of conditions that differ along the lexical dimension (i.e., S vs. J and W vs. N), and pairs of conditions that differ along the syntactic dimension (i.e., S vs. W and J vs. N). Asterisks indicate the comparisons that survive the FDR correction for the number of ROIs.

	Interaction statistics
LIFGorb	$F(3,22) = 1.45$; n.s.
LIFG	$F(3,22) = 4.49$; $p < .05^*$
LMFG	$F(3,22) = 4.03$; $p < .05^*$
LSFG	$F(3,22) = 1.44$; n.s.
LAntTemp	$F(3,22) = 4.53$; $p < .05^*$
LMidAntTemp	$F(3,22) = 6.54$; $p < .005^*$
LMidPostTemp	$F(3,22) = 8.19$; $p < .005^*$
LPostTemp	$F(3,22) = 10.5$; $p < .001^*$
LAngG	$F(3,22) = 12$; $p < .001^*$
RMidAntTemp	$F(3,22) = 1.95$; n.s.
RMidPostTemp	$F(3,22) = 2.72$; $p = .07$
RCereb	$F(3,22) = 3.84$; $p < .05^*$
LCereb	$F(3,22) = 3.22$; $p < .05$

those that differ along the syntactic dimension, and many of the regions – including the LIFG and LMFG ROIs and all of the regions in the left temporal lobe – show this effect reliably. Similarly, in the searchlight analyses (Fig. 6), the maps for pairs of conditions that differ along the lexical dimension are more robust and extensive than those for pairs of conditions that differ along the syntactic dimension (the latter is essentially a subset of the former). These results suggest that throughout the language system lexical information is represented more robustly than syntactic information. We discuss the implications of this finding in Section 3.3.

Two additional points are worth making. First, several brain regions have been previously implicated in syntactic processing. These include parts of the left inferior frontal gyrus (e.g., Ben-Shachar, Hendler, Kahn, Ben-Bashat, & Grodzinsky, 2003; Caramazza & Zurif, 1976; Friederici, Fiebach, Schlesewsky, Bornkessel, & von Cramon, 2006; Grodzinsky, 2000; Stromswold, Caplan, Alpert, & Rauch, 1996) and, more recently, anterior temporal regions (e.g., Brennan et al., 2011; Humphries, Love, Swinney, & Hickok, 2005; Noppeney & Price, 2004; Rogalsky & Hickok, 2009). Note that even regions in the left IFG and in the anterior temporal cortices clearly represent lexical information more strongly than syntactic information in the patterns of neural activity.

And second, whereas current results are still consistent with the role of inferior frontal regions in representing abstract syntactic information (even though these regions are clearly not *selective* for abstract syntactic processing), they undermine the claims (e.g., Rogalsky & Hickok, 2009) that anterior temporal regions are important for representing abstract syntactic information. In natural language, syntax is inherently confounded with compositional semantics. As a result, many findings that are interpreted as indexing syntactic processing can be interpreted in terms of the effects of compositional semantic processing. In the current data, anterior temporal brain regions (in contrast to regions in the left inferior frontal gyrus or in the posterior MTG) show no significant ability to discriminate between the Jabberwocky and Nonword-lists conditions (see the J vs. N contrast in Fig. 6), which differ along the syntactic dimension, and yet discriminate well between the Sentences and Word-lists conditions, which differ along both the syntactic and the compositional semantic dimension. This finding suggests that the anterior temporal structures are more likely to subserve compositional semantic processing, consistent with some earlier hypotheses about the function of these regions (e.g., Vandenberghe et al., 2002).

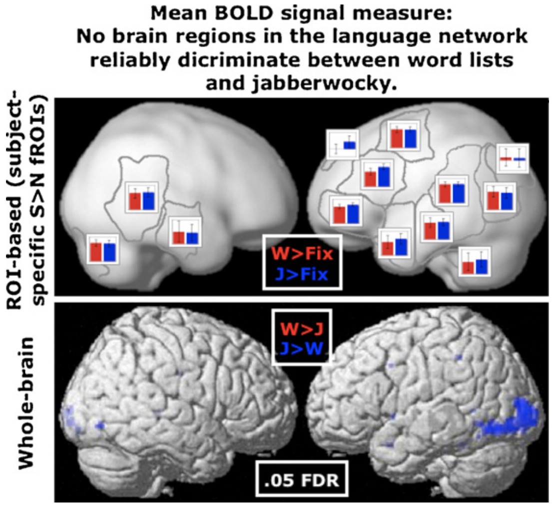


Fig. 8. Top: Mean BOLD responses (in PSC units; see Fig. D1 in Fedorenko et al., 2010, for more details) in the language ROIs to word lists (red bars) and jabberwocky sentences (blue bars) relative to the fixation baseline. The ROIs were defined by intersecting the parcels, whose outlines are shown in grey (see also Fig. 5, bottom), with subject-specific thresholded (at $p < .05$, FDR-corrected) activation maps for the Sentences > Nonword-lists contrast, as described in Fedorenko et al. (2010). Each of these regions replicates the Sentences > Nonwords effect in independent data (left out runs), but shows no difference in response between word lists and jabberwocky. Bottom: The results of random-effects group analyses for the Word-lists > Jabberwocky contrast (red) and the Jabberwocky > Word-lists contrast (blue). The activation maps are thresholded at $p < .05$, FDR-corrected. (For the W > J contrast, no voxels emerged at this threshold. As noted in the text, for the J > W contrast, only the inferior posterior temporal/occipital regions – that fall outside of the classical language network – emerge at this threshold.)

3.2. Do any of the language brain regions distinguish between "pure" lexical information (lists of words) and "pure" syntactic information (jabberwocky sentences)?

We first present key results from the univariate analyses for the Word-lists and Jabberwocky conditions. In Fig. 8 (see Figs. D1 and D2 in Fedorenko et al., 2010, for more details) we present the results from the ROI-based and whole-brain analyses for the Word-lists and Jabberwocky conditions, which show that the Word-lists and Jabberwocky conditions elicit a similar-strength BOLD response in each of the ROIs. In the whole-brain analysis we do not see any differential effects in the left frontal or temporal/parietal cortices (see also Fig. D2 in Fedorenko et al., 2010),[10] except for a reliably stronger response to Jabberwocky in the inferior posterior temporal and occipital regions, i.e., regions that fall outside of the classical language network. In Fedorenko et al. (2010) we hypothesized that these activations reflect greater visual difficulty in processing nonwords compared to real words rather than anything to do with high-level linguistic (e.g., syntactic) processing.[11] In

---

[10] It is worth pointing out that the whole-brain maps in Fig. 8 are from a traditional random-effects group analysis. We are using these maps here as a summary representation because careful analyses of individual subjects' activation maps for these contrasts revealed a similar picture, with no robust J > W or W > J effects anywhere in the left frontal or temporo/parietal cortices.

[11] In particular, during the processing of jabberwocky stimuli phonological representations must be constructed in a bottom-up way from individual graphemes. In contrast, in processing real words, top-down influences (whole-word template matching; e.g., Coltheart et al., 1993) may facilitate word processing. Consistent with this hypothesis, looking at individual subjects' ROIs defined by the Jabberwocky > Words contrast, we found that these inferior posterior temporal/occipital regions also respond more to nonwords than sentences. Furthermore, in the auditory presentation of the same materials, the Jabberwocky > Word-lists contrast does not activate these regions (Fedorenko & Kanwisher, unpublished data), indicating that these activations cannot reflect abstract high-level linguistic processing, which should be similar for visual and auditory presentation.

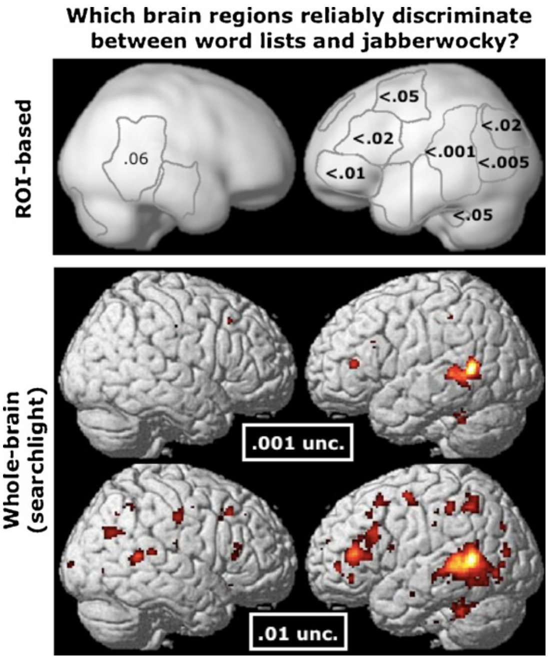


Fig. 9. The results of *F*-tests evaluating which regions can discriminate between word lists and jabberwocky sentences in the pattern of activity. Top: The results of ROI-based analyses. We show uncorrected values here (see Table 1 for details). If no value is shown for an ROI, this means that the region does not discriminate significantly above chance between the two conditions. Bottom: The results of the searchlight analyses at two different thresholds: .001, uncorrected, and .01, uncorrected.

summary, nowhere within the language network are the Word-lists and Jabberwocky conditions distinguished in terms of the mean BOLD response.

To see whether the Word-lists and Jabberwocky conditions may be distinguished in the pattern of neural activity, we turned to MVPAs. In Fig. 9 we present the results of ROI-based (see Table 1, last column, for the statistics) and whole-brain searchlight-style analyses. Both ROI-based and searchlight analyses demonstrate that a number of brain regions can reliably distinguish between these two conditions, although the results of the searchlight analyses are weaker. This includes regions on the lateral surface of the left frontal lobe and left posterior temporal regions. (See Appendix A for the results of ROI-based analyses where subject-specific functional ROIs are used.)

The searchlight analyses demonstrate that the between-region differences in the ROI-based analyses are not explainable by the different sizes of the ROIs. For example, it could be the case that some regions show better discrimination simply because they include more voxels (like e.g., the LMidPostTemp ROI). But to the extent that we find similar results in the searchlight whole-brain analyses, we can be more confident that the between-region differences in the ROI-based analyses are not driven by the differences in the size of the ROIs, but rather by some intrinsic functional properties of those regions.

The fact that several brain regions reliably discriminate between the Word-lists and Jabberwocky conditions in the spatial patterns of neural activity, in the absence of a difference in the mean BOLD response, argues against the most extreme version of the integration of lexical and syntactic knowledge, where each voxel is equally sensitive to both lexical and syntactic information (Fig. 3a). Instead, it appears that in several of the language regions voxels differ in their relative sensitivity to lexical vs. syntactic information, with some voxels possibly even being selectively sensitive to one vs. the other kind of information.

Note that the regions that can discriminate between word lists and jabberwocky sentences in the pattern of activity are the brain regions that have been classically implicated in linguistic processing, i.e., regions on the lateral surface of the left frontal lobe and regions in the posterior temporal lobe (e.g., Broca, 1861; Geschwind, 1970; Wernicke, 1874). Damage to these regions most commonly leads to linguistic deficits. Looking at the discrimination results across all four conditions (Section 3.1), we can see that these same regions perform better than other regions within the extended language network across different pairwise comparisons, suggesting that this subset of brain regions represents linguistic information in a more robust and fine-grained manner.

3.3. Summary and discussion

We used multi-voxel pattern analyses to ask two questions about how lexical and syntactic information is represented in the brain. First, we asked whether language brain regions differ in how robustly they represent lexical vs. syntactic information. We found that pairs of conditions that differ along the lexical dimension (i.e., sentences vs. jabberwocky sentences, and word lists vs. nonword lists) are discriminated more reliably in pattern analyses than pairs of conditions that differ along the syntactic dimension (i.e., sentences vs. word lists, and jabberwocky vs. nonword lists) throughout the language system. And second, we asked whether any of the language regions can discriminate between "pure" lexical information and "pure" abstract syntactic information, i.e., the Word-lists and Jabberwocky conditions in our experiment. We found that regions in the inferior frontal gyrus and posterior temporal cortices can discriminate between these two conditions, in spite of the fact that these conditions elicit a similar-strength BOLD response throughout the language system. We now discuss the implications of these findings.

3.3.1. The relative importance of lexical vs. syntactic information for representing linguistic meaning

Our results suggest that lexical information is represented more robustly than syntactic information across different brain regions within the language system. We were able to uncover this pattern using MVPA-style analyses, but not using univariate analysis methods, which showed that each high-level language region is similarly sensitive to lexical and syntactic information (e.g., Fedorenko et al., 2010).

Although syntactic information provides an important cue to meaning, lexical-level information, especially given the linguistic and extralinguistic (visual, social, etc.) context, is often sufficient for inferring the intended meaning. As an intuitive example, think about talking to kids or foreigners, who have not yet mastered the grammar of the language. Although such speakers may occasionally omit articles, fail to add the required functional morphemes (e.g., the third singular -s marker on the verb), and/or put the words in an incorrect order, we can often understand the intended meanings perfectly well based on just the lexical items and contextual cues. In contrast, the meaning carried by syntax alone (i.e., word order and functional morphology) is considerably weaker. A Jabberwocky string does convey some meaning (abstract structure of an event). For example, in a string like "The florper blimmed a mabe", we can infer that some event ("blimming") took place, where a (probably animate) entity "florper" acted upon a (probably inanimate) entity "mabe" in some way. However, this kind of meaning on its own would be insufficient for communication (cf., a string of unconnected uninflected content words "boy apple eat", where it is relatively easy to infer the meaning of "a boy eating an apple").

These results fit nicely with the existing psycholinguistic evidence that suggests that lexical information is an important source of information that can guide sentence interpretation (e.g., MacDonald et al., 1994; cf. earlier proposals according to which only syntactic information can guide interpretation; e.g., Frazier, 1978; Ferreira & Clifton, 1986).

The goal of the language system is to construct and decode meanings. Because the bulk of the meaning of an utterance can be inferred from the lexical items, it makes sense for the language system to represent lexical information more robustly.

3.3.2. *Lexical and syntactic information is differentiated in the spatial patterns of neural activity in some of the language regions*

Our previous work (Bedny et al., 2011; Fedorenko et al., 2010), as well as earlier imaging work and considerable evidence from brain-damaged patients, has suggested that the same brain regions support lexical representations of individual words and abstract syntactic representations. The current work demonstrated, however, that conditions that selectively target lexical processing vs. abstract syntactic processing can be discriminated in the pattern of neural activity in some regions within the language system, even though throughout the language network these two conditions elicit a similarly strong level of BOLD response (that falls in between that elicited by sentences and that elicited by nonwords). This result helps rule out an extreme version of an integrated view of lexical and syntactic information where each voxel is sensitive to both kinds of information, and voxels do not differ in their relative sensitivity to lexical vs. syntactic information. Furthermore, this result highlights the promise of MVPAs for discovering dissociations in cases where a brain region may not be working harder during the processing of one condition than another and yet may show different patterns of activation for the two conditions.

In particular, one common strategy in neuroimaging studies of language is to compare a condition that places a greater burden on some mental process of interest (e.g., retrieving lexical representations from long-term semantic memory, or forming syntactic dependencies between words) to a condition that places a lower burden on that process (e.g., Ben-Shachar et al., 2003; Graves, Grabowski, Mehta, & Gordon, 2007; Keller, Carpenter, & Just, 2001; Stromswold et al., 1996). However, many important dimensions of our language knowledge may not be dissociable with such manipulations. As a result, multi-voxel pattern analyses are perfectly suited for complementing traditional univariate fMRI methods in investigating some theoretical distinctions in language where the relevant conditions do not differ in processing difficulty.

3.3.3. *Concluding remarks*

We conclude by laying out three important open questions about lexical and syntactic processing that future behavioral and neuroimaging work should aim to address. These questions will be best answered by a tighter integration of neuroimaging with behavioral psycholinguistic work.

First, given that quite a number of brain regions appear to be sensitive to both lexical and syntactic information, it will be important to characterize the precise role of each of these regions in high-level linguistic processing. We here showed that several of these regions (inferior frontal and posterior temporal regions) appear to represent linguistic information in a more fine-grained and robust manner than other regions. In particular, these regions (a) are better at discriminating among different linguistic conditions and (b) can discriminate between conditions that are indistinguishable in the mean BOLD response (i.e., the Word-lists and Jabberwocky conditions). These regions happen to be the regions that have been classically implicated in linguistic processing (Broca, 1861; Wernicke, 1874). Perhaps these regions are the ones that actually store our linguistic knowledge or conduct linguistic computations, whereas the other regions in the extended linguistic network respond to lexical and syntactic information for more artifactual reasons (e.g., due to post-linguistic conceptual processing). Future work will tell. Of course, it is worth keeping in mind that the current study was conducted in English. It is possible that cross-linguistic differences in the richness of morphology or in the rigidity of word order within and across constructions may affect the nature of the linguistic representations. As a result, it will be important to extend the current findings to other languages.

Second, and relatedly, in order to understand the precise function of any given language region, detailed information on the time-course of information processing in that region is needed. Combining methods like fMRI with more temporally sensitive methods (e.g., EEG, TMS, eCOG) is likely to afford substantial progress (e.g., Dale et al., 2000; Kuperberg et al., 2003; Sahin, Pinker, Cash, Schomer, & Halgren, 2009).

And third, as we discussed above, the goal of the language system is to construct (in production) and decode (in comprehension) meanings. An important question therefore concerns the relationship between the language system and brain regions that store abstract conceptual knowledge (e.g., Patterson, Nestor, & Rogers, 2007). It seems clear that there must be a close relationship between language regions and abstract conceptual regions, but whether this relationship is implemented through neural overlap between these sets of regions, or through patterns of connections between them is still an open question.

In conclusion, analysis methods that do not require one condition to elicit an overall stronger response than another condition (like MVPAs, or neural adaptation; e.g., Grill-Spector & Malach, 2001; Krekelberg, Boynton, & van Wezel, 2006) are likely to provide substantial insights into the representations that underlie language comprehension and production. These methods allow us to characterize the information represented in each region and to determine whether different aspects of our language knowledge are neurally dissociable even if they are represented in the same brain regions and activate these regions to the same extent. We here showed that although all the key brain regions within the language system are sensitive to both lexical and syntactic information, (1) lexical information is represented more robustly than syntactic information across these regions, which suggests perhaps that lexical information plays a more important role in representing linguistic meaning and (2) some of these regions distinguish between "pure" lexical and "pure" syntactic information in the spatial patterns of neural activity, which rules out one kind of an architecture where lexical and syntactic information are so tightly integrated that each voxel is sensitive to both kinds of information, and voxels do not differ in their relative sensitivity to lexical vs. syntactic information. The recent advances in the development of functional localizers for brain regions engaged in high-level linguistic processing (e.g., Fedorenko et al., 2010; Pinel et al., 2007) should help increase the power of MVPA-style methods and neural adaptation paradigms. As a result, future work may be able to investigate the neural basis of sophisticated linguistic distinctions, like different aspects of syntactic knowledge, different lexico-semantic properties of word classes, or subtle meaning differences between words or sentences.

Acknowledgements

We are grateful to Ted Gibson, Roger Levy, and members of the TedLab and SaxeLab for insightful discussions about the relationship between the lexicon and syntax. We thank two anonymous reviewers, Ray Jackendoff, Roger Levy, Kyle, Mahowald, Tim O'Donnell, Ivan Sag and Hal Tily, for comments that helped to greatly improve this manuscript. We thank Christina Triantafyllou, Steve Shannon, and Sheeba Arnold for technical support with fMRI data acquisition. We also would like to acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, Massachusetts Institute of Technology. This research was funded by Eunice Kennedy Shriver National Institute of Child Health and Human Development Award K99HD-057522 (to E.F.).

Appendix A. Discrimination between word lists and jabberwocky sentences in functionally defined subject-specific fROIs.

In addition to the analyses reported in Section 3.2, where we used group-level functional parcels, we examined the performance of subject-specific functional ROIs. In particular, we used the contrast between sentences and nonword lists to define subject-specific fROIs within each parcel and then tested how well those fROIs can discriminate between the Word lists and Jabberwocky conditions. Subject-specific fROIs were defined as described in Fedorenko et al. (2010): each individual subject's thresholded (at $p < .001$, uncorrected) activation map for the relevant contrast (Sentences > Nonword-lists) was intersected with the parcels, and all the supra-threshold voxels within the borders of the parcel were taken as that subject's fROI. The results are presented in the right column of Table 3 (the left column is repeated from Table 1, for easier comparison). In spite of the fact that these analyses have *considerably* less power because they only use a small subset of the voxels that were used in the whole-parcel-based analyses above (20% of voxels, on average, across regions), for a couple of ROIs the discrimination results look as robust or more robust with subject-specific functional masking. This is true of the following regions: LIFG, LMidPostTemp, and LAngG. This pattern suggests that in these regions the discrimination we see when examining the group-level parcels is largely carried by the voxels that respond more strongly to sentences than nonwords. It is interesting that these regions are once again the regions that are classically implicated in linguistic processing (see main text for more discussion). In addition to these regions, we see a similar pattern in the L and R MidAntTemp ROIs, but even with subject specific masking the results do not reach significance. However, this suggests in that some cases adding subject-specific functional masking may reveal effects that are otherwise obscured by the noise from the "irrelevant" voxels.

Table 3
F and *p* values for the tests evaluating the discriminability between Word-lists and Jabberwocky conditions in two kinds of ROIs. For the whole-parcel-based analyses, degrees of freedom are (2,23) in all cases. Asterisks indicate the effects that survive the FDR correction for multiple ROIs.

	Whole-parcel-based ROI analysis (repeated from Table 1)	Subject-specific fROI analysis
LIFGorb	$F = 6.38$; $p < .01^*$	$F(2,21) < 1$; n.s.
LIFG	$F = 4.69$; $p < .05$	$F(2,23) = 6.69$; $p < .01^*$
LMFG	$F = 4.19$; $p < .05$	$F(2,22) = 3.08$; $p = .066$
LSFG	$F < 1$; n.s.	$F(2,17) < 1$; n.s.
LAntTemp	$F = 1.52$; n.s.	$F(2,21) = 1.02$; n.s.
LMidAntTemp	$F = 1.56$; n.s.	$F(2,22) = 2.87$; $p = .078$
LMidPostTemp	$F = 9.79$; $p < .001^*$	$F(2,23) = 17.9$; $p < .001^*$
LPostTemp	$F = 8.21$; $p < .005^*$	$F(2,22) < 1$; n.s.
LAngG	$F = 5.55$; $p < .05^*$	$F(2,20) = 5.83$; $p < .05^*$
RMidAntTemp	$F < 1$; n.s.	$F(2,19) = 2.78$; $p = .088$
RMidPostTemp	$F = 3.15$; $p = .061$	$F(2,21) = 1.25$; n.s.
RCereb	$F = 1.50$; n.s.	$F(2,20) = 1.03$; n.s.
LCereb	$F = 3.63$; $p < .05$	$F(2,15) = 2.26$; n.s.

In a couple of regions, however, the use of subject-specific functional masking renders the effects less robust. This pattern is true of the LIFGorb, LMFG, LPostTemp and LCereb ROIs. This suggests that in these regions voxels that do not respond more strongly to sentences than nonwords contribute in an important way to classification. For example, we speculate that in the LPostTemp region the discrimination may at least in part be driven by voxels that respond more to jabberwocky than word lists (and also to nonwords than sentences) located in the posterior inferior portions of the temporal lobe (see footnote 11). Some of these voxels (see Fig. 8) may be included in the LPostTemp parcel.

Appendix B. Whole-brain searchlight-style analyses.

For each voxel, with spatial coordinates (x,y,z) represented in voxel space, the activation during conditions A and B estimated from two independent halves of the functional data can be represented by the three-dimensional spatial maps $f_A(x,y,z)$ and $f_B(x,y,z)$, respectively. For any given voxel of interest, with spatial coordinates (u,v,w), we can consider an ROI around this voxel whose shape and distribution is characterized by a function $h(x,y,z,u,v,w)$. This function can take discrete values 0 and 1 for 'hard' ROIs, or continuous values between 0 and 1 for 'probabilistic' ROIs. For simplicity and without loss of generality we will assume that h is scaled so that it sums up to one across all voxels (x,y,z). In addition, for ROIs that share the same shape and size independently of the voxel of interest (u,v,w), the function h can be more simply represented by a three-dimensional kernel function $h(x-u,y-v,z-w)$. As a measure of the similarity between the spatial patterns of activation during conditions A and B restricted to voxels within a region centered around the voxel (u,v,w), the spatial correlation $r(u,v,w)$ can be computed as:

$$r(u,v,w) = \frac{\sigma_{A,B}(u,v,w)}{\sigma_{A,A}(u,v,w)^{1/2} \cdot \sigma_{B,B}(u,v,w)^{1/2}}$$

where :

$$\sigma_{A,B}(u,v,w) \equiv \sum_{x,y,z} (f_A(x,y,z) - \mu_A(u,v,w)) \cdot (f_B(x,y,z) - \mu_B(u,v,w)) \cdot h(x-u, y-v, z-w)$$

$$\mu_\bullet(u,v,w) \equiv \sum_{x,y,z} f_\bullet(x,y,z) \cdot h(x-u, y-v, z-w)$$

The computations involved can be implemented as a combination of spatial convolutions and simple voxel-wise operations directly on the three-dimensional maps f_A and f_B:

$$r = \frac{(f_A \cdot f_B) * h - \mu_A \cdot \mu_B}{(f_A^2 * h - \mu_A^2)^{1/2} \cdot (f_B^2 * h - \mu_B^2)^{1/2}}$$

where :

$$\mu_\bullet = f_\bullet * h$$

The "*" symbol represents the convolution operator, and the rest of the operations represent voxel-wise operations. In addition, when the kernel function is defined as a Gaussian kernel, the convolutions with the kernel h above represent simple spatial smoothing operations, which are implemented in SPM as well as other packages in an efficient manner taking advantage of the separability of this kernel (three-dimensional smoothing is implemented as the combination of three one-dimensional smoothing operations).

References

Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E. & Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Sciences of the United States of America*, in press.

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D. & Grodzinsky, Y. (2003). The neural reality of syntactic transformations: Evidence from functional magnetic resonance imaging. *Psychological Science*, *14*(5), 433–440.

Braze, D., Mencl, W. E., Tabor, W., Pugh, K. R., Constable, R. T., Fulbright, R. K., et al. (2011). Unification of sentence processing via ear and eye: An fMRI study. *Cortex*, *47*, 416–431.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J. & Pylkkänen, L. (2011). Syntactic complexity predicts anterior temporal activity during natural story listening. *Brain and Language*, in press.

Bresnan, J. (1982). *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.

Broca, P. (1861). Notes on the seat of the faculty of articulate language, followed by an observation of aphemia. *Bulletin de la Société Anatomique*, *6*, 330–357 [translated from French].

Bybee, J. (1998). A functionalist approach to grammar and its evolution. *Evolution of Communication*, *2*, 249–278.

Caramazza, A. & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain & Language*, *3*, 572–582.

Carroll, L. (1871). *Through the looking glass and what Alice found there*.

Charniak, E. (1996). Tree-bank grammars. Technical Report CS-96-02, Department of Computer Science, Brown University.

Charniak, E. (1997). Statistical parsing with a context-free grammar and word statistics. In *Proceedings of the fourteenth national conference on artificial intelligence (AAAI 1997)*.

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

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures*. Holland: Foris Publications. Reprint. 7th ed. Berlin and New York: Mouton de Gruyter, 1993.

Clifton, C., Jr., Frazier, L. & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *23*, 696–708.

Collins, M. (1996). A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th annual meeting of the ACL, Santa Cruz*.

Collins, M. (1997). Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th annual meeting of the ACL (jointly with the 8th conference of the EACL)* Madrid.

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). Models of reading aloud: Dual route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.

Culicover, P. W. & Jackendoff, R. (2005). *Simpler syntax*. Oxford University Press.

Dale, A. M., Liu, A. K., Fischl, B., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high resolution imaging of cortical activity. *Neuron*, *26*, 55–67.

Dapretto, M. & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron*, *24*, 427–432.

DiCarlo, J. J. & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341.

Duda, R. O., Hart, P. E. & Stork, D. G. (2001). *Pattern classification* (2nd ed.). Wiley.

Fedorenko, E. & Kanwisher, N. (2011). Functionally localizing language-sensitive regions in individual subjects with fMRI: A reply to Grodzinsky's critique of Fedorenko & Kanwisher (2009). *Language and Linguistics Compass*, *5*(2), 78–94.

Fedorenko, E., Frank, M. & Gibson, E. (2009). Syntactic complexity effects in Jabberwocky sentences. In *The 22nd CUNY conference on human sentence processing* Davis, CA, March 2009.

Fedorenko, E., Gibson, E. & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language*, *56*(2), 246–269.

Fedorenko, E., Hsieh, P.-J., Nieto-Castañon, A., Whitfield-Gabrieli, S. & Kanwisher, N. (2010). A new method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, *104*, 1177–1194.

Fedorenko, E., Nieto-Castanon, A., & Kanwisher, K. Syntactic processing in the human brain: What we know, what we don't know, and a suggestion for how to proceed. *Brain and Language*, in press.

Ferreira, F. & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.

Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies. PhD dissertation, University of Connecticut.

Frazier, L. & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*, 291–325.

Freiwald, W. A. & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*(6005), 845–851.

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. & von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex*, *16*(12), 1709–1717.

Friederici, A. D., Meyer, M. & von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, *74*, 289–300.

Garnsey, S. M, Pearlmutter, N. P., Myers, E. & Lotocky, M. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58–93.

Gennari, S. & MacDonald, M. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*, 161–187.

Geschwind, N. (1970). The organization of language and the brain: Language disorders after brain damage help in elucidating the neural basis of verbal behavior. *Science*, *170*, 940–944.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76.

Gibson, E. & Fedorenko, E. (2011). The domain generality of working memory resource for language. In *Architectures and mechanisms of language processing conference* Paris, France.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gordon, P. C., Hendrick, R. & Levine, W. H. (2002). Memory load interference in syntactic processing. *Psychological Science*, *13*, 425–430.

Graves, W., Grabowski, T., Mehta, S. & Gordon, J. (2007). A neural signature of phonological access: Distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of Cognitive Neuroscience*, *19*, 617–631.

Grill-Spector, K. & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*, 293–321.

Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, *23*(1), 1–71.

Grodzinsky, Y. & Santi, A. (2008). The battle for Broca's region. *Trends in Cognitive Sciences*, *12*(12), 474–480.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L. & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2429.

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.

Humphries, C., Love, T., Swinney, D. & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping*, *26*, 128–138.

Jackendoff, R. (1977). X-bar-Syntax: A study of phrase structure. In *Linguistic inquiry monograph 2*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, *1146*, 2–22.

Joshi, A., Levy, L. & Takahashi, M. (1975). Tree adjunct grammars. *Journal of the Computer and System Sciences*, *10*(1), 136–163.

Julian, J. B., Fedorenko E., Webster, J. & Kanwisher, N. (submitted). An algorithmic method for functionally defining regions of interest in the ventral visual pathway.

Kako, E. (2006). The semantics of syntactic frames. *Language and Cognitive Processes*, *21*(1), 562–575.

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11163–11170.

Keller, T. A., Carpenter, P. A. & Just, M. A. (2001). The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. *Cerebral Cortex*, *11*, 223–237.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 3–24.

Krekelberg, B., Boynton, G. M. & van Wezel, R. J. (2006). Adaptation: From single cells to BOLD signals. *Trends in Neuroscience*, *29*(5), 250–256.

Kriegeskorte, N., Goebel, R. & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 3863–3868.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M. & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience*, *15*, 272–293.

Linebarger, M. C., Schwartz, M. F. & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, *13*, 361–392.

MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

Magerman, D. (1994). Natural language parsing as statistical pattern recognition. Doctoral dissertation, Stanford University.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., et al. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*, 467–479.

McElree, B., Foraker, S. & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*, 67–91.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165–178.

Nieto-Castañon, A., Kanwisher, N., & Fedorenko, E. Functional localizers increase the sensitivity and functional resolution of multi-subject analyses, submitted for publication.

Noppeney, U. & Price, C. J. (2004). An FMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience*, *16*(4), 702–713.

Norman, K. A., Polyn, S. M., Detre, G. J. & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430.

Novick, J. M., Trueswell, J. C. & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*, 263–281.

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*(12), 976–987.

Pereira, F., Mitchell, T. & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, *45*(1 Suppl.), S199–S209.

Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Le Bihan, D., et al. (2007). Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neuroscience*, *8*, 91.

Pollard, C. & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Rayner, K. & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191–201.

Reali, F. & Christiansen, M. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*, 1–23.

Rogalsky, C. & Hickok, G. (2009). Selective attention modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex*, *19*, 786–796.

Rogalsky, C. & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience*, *23*, 1664–1680.

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*, *326*, 445–449.

Saxe, R., Brett, M. & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *Neuroimage*, *30*(4), 1088–1096, discussion 1097–1099.

Schabes, Y., Abeille, A. & Joshi A.K. (1988). Parsing strategies with 'lexicalized' grammars: Application to tree adjoining grammars. Proceedings of the 12th International Conference on Computational Linguistics. Budapest, Hungary.

Schwarzkopf, S. D. & Rees, G. (2011). Pattern classification using functional magnetic resonance imaging. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 568–579.

Stowe, L. A., Paans, A. M. J., Wijers, A. A., Zwarts, F., Mulder, G. & Vaalburg, W. (1999). Sentence comprehension and word repetition: A positron emission tomography investigation. *Psychophysiology*, *36*, 786–801.

Stromswold, K., Caplan, D., Alpert, N. & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, *52*, 452–473.

Traxler, M. J., Morris, R. K. & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69–90.

Trueswell, J. C., Tanenhaus, M. K. & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Vandenberghe, R., Nobre, A. C. & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, *14*(4), 550–560.

Vul, E. & Kanwisher, N. (2009). Begging the question: The non-independence error in fmri data analysis. *Foundations and Philosophy for Neuroimaging*.

Wernicke, C. (1874). Der Aphasische Symptomencomplex (Max Cohn and Weigert, Breslau, Germany). In R. S. Cohen, & M. W. Wartofsky (Eds.), *Boston studies in the philosophy of science* (Vol. 4). Dordrecht, The Netherlands: Reidel (German, trans (1969)).

Yamada, Y. & Neville, H. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Research*, *1130*, 167–180.