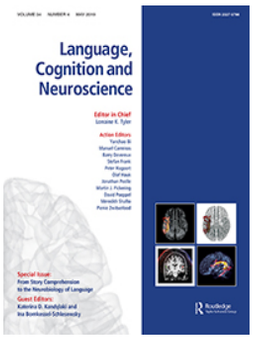

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension

Shohini Bhattasali, Murielle Fabre, Wen-Ming Luh, Hazem Al Saied, Mathieu Constant, Christophe Pallier, Jonathan R. Brennan, R. Nathan Spreng & John Hale

To cite this article: Shohini Bhattasali, Murielle Fabre, Wen-Ming Luh, Hazem Al Saied, Mathieu Constant, Christophe Pallier, Jonathan R. Brennan, R. Nathan Spreng & John Hale (2019) Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension, Language, Cognition and Neuroscience, 34:4, 491-510, DOI: 10.1080/23273798.2018.1518533

To link to this article: https://doi.org/10.1080/23273798.2018.1518533

LANGUAGE, COGNITION AND NEUROSCIENCE
2019, VOL. 34, NO. 4, 491–510
https://doi.org/10.1080/23273798.2018.1518533

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension

Shohini Bhattasali [a*], Murielle Fabre[a], Wen-Ming Luh[a*], Hazem Al Saied[b], Mathieu Constant[b], Christophe Pallier[c], Jonathan R. Brennan [d], R. Nathan Spreng[e] and John Hale[a#]

[a]Cornell University, Ithaca, NY, USA; [b]Université de Lorraine, CNRS, ATILF, Nancy, France; [c]INSERM-CEA, Paris-Saclay, France; [d]University of Michigan, Ann Arbor, MI, USA; [e]McGill University, Montreal, Canada

ABSTRACT

This study examines memory retrieval and syntactic composition using fMRI while participants listen to a book, The Little Prince. These two processes are quantified drawing on methods from computational linguistics. Memory retrieval is quantified via *multi-word expressions* that are likely to be stored as a unit, rather than built-up compositionally. Syntactic composition is quantified via *bottom-up parsing* that tracks tree-building work needed in composed syntactic phrases. Regression analyses localise these to spatially-distinct brain regions. Composition mainly correlates with bilateral activity in anterior temporal lobe and inferior frontal gyrus. Retrieval of stored expressions drives right-lateralised activation in the precuneus. Less cohesive expressions activate well-known nodes of the language network implicated in composition. These results help to detail the neuroanatomical bases of two widely-assumed cognitive operations in language processing.

ARTICLE HISTORY
Received 28 February 2018
Accepted 10 August 2018

KEYWORDS
Memory retrieval; structural composition; syntax; lexical cohesion; naturalistic comprehension; fMRI

1. Introduction

Our human ability to comprehend natural language relies on two fundamental of cognitive operations. One involves the retrieval of memorised elements, while the other composes those elements together into novel expressions. Despite a growing body of work on the brain's language network (see e.g. Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004; Fedorenko et al., 2016; Friederici & Gierhan, 2013; Hagoort & Indefrey, 2014; Pallier, Devauchelle, & Dehaene, 2011) the neural bases of these two operations within this network remain imprecisely specified.

This study disentangles these two cognitive processes with an analysis of fMRI time-courses observed during an extended episode of naturalistic spoken story understanding. Like prior work, it suggests a localisation of memory retrieval and composition to largely distinct regions of the brain. But by linking these operations with quantitatively precise computational models, this study paves the way for more granular research into the relationship between these operations. Further, this localisation is based upon brain responses to a literary narrative, one that attests a variety of different linguistic phenomena in an ecologically valid setting rather than one or two constructions repeated out of context.

The processing of retrieving memorised elements is operationalised using a specific type of word sequence called "multiword expressions." This term (henceforth abbreviated MWE) comes from computational linguistics. Broadly, it refers to word sequences that are better-treated as a single unit rather than as a structurally-composed combination (Sag, Baldwin, Bond, Copestake, & Flickinger, 2002). MWE themselves raise a general question about the balance between productivity and reuse in language processing (Goldberg, 2006; Jackendoff, 2002; O'Donnell, 2015). If they truly lack internal structure, then their processing might proceed via a single, unitary retrieval operation rather than step-by-step composition. Even if they are amenable to some sort of internal structure analysis, existing proposals hold that the entire item is stored and retrieved as a unit.

Adopting this idea as a working hypothesis, we assume that MWE comprehension involves an additional memory retrieval, one that is not involved in non-MWE expressions. No assumptions regarding the nature of this retrieval are required, only that the stored element should correspond to the expression as a whole.

This distinction between retrieval and composition figures centrally in neurocognitive models of language processing (Bornkessel-Schlesewsky & Schlesewsky,

CONTACT John Hale jthale@cornell.edu, jthale@uga.edu
Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2018.1518533
*Co-first authors.
#Present address: Department of Linguistics, University of Georgia, Athens, GA, USA.

2009, 2013; Hagoort, 2005, 2016; Ullman, 2001, 2004, 2015). These models all suppose, to some degree or another, that these two operations in language processing dissociate at the level of neurobiology. The neurobiological implications of this claim are discussed in more detail below in Section 2.5. The present study adds to prior work by instantiating these two operations in quantitatively precise computational models, and by probing their neural localisation during a naturalistic task.[1]

2. Present study

We pursue our objective by investigating two specific research questions. The first is to localise these processes (cf. Yang, Marslen-Wilson, & Bozic, 2017): Which brain structures support cognitive processes of structural composition and retrieval of stored linguistic elements, in naturalistic language comprehension? The second research question has to do with the retrieval operation specifically: Are some MWEs better candidates for comprehension via direct retrieval than others?

We investigate these questions using naturalistic spoken story-comprehension, with a view towards to maximising ecological validity. The stimulus, introduced below in Section 2.1, is a literary text annotated with numbers that quantify processing effort associated with each operation (Sections 2.2 and 2.4). This section finishes by deriving from the existing literature a broad hypothesis regarding the organisation of these two cognitive processes in the brain (Section 2.5).

2.1. Naturalistic stimulus

Participants heard de Saint-Exupéry's *The Little Prince* (1943) as translated into English by David Wilkinson and read out loud by Nadine Eckert-Boulet. The text constitutes a fairly lengthy exposure to naturalistic language, comprising 15,388 words and lasting over an hour and a half. This follows Willems' (2015) encouragement to probe the neural bases of language comprehension with greater ecological validity.

The MWE that are identified in this text reflect contemporary usage of these expressions, and they are linguistically diverse (for details see Sections 2.2 and 3.2). To get a sense of how these MWE attestation frequencies relate to distributional patterns in English more generally, Figure 1 compares attestation counts in the stimulus text to their corresponding frequencies in the 560-million word Corpus of Contemporary American English (COCA; Davies, 2008). Over half (56.54%) of the attestations are headed by a verb, and only two items

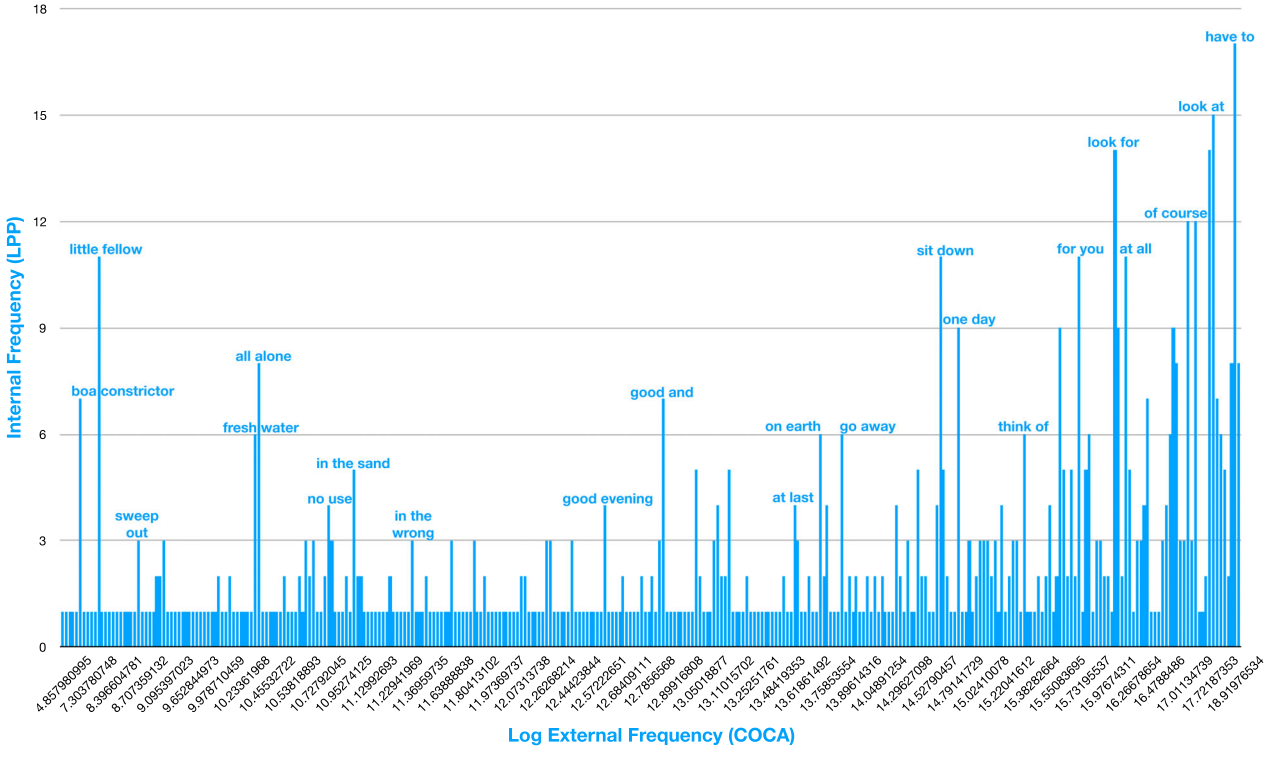

Figure 1. MWE frequency comparison. The vertical axis shows counts of multiword expression attestations within the natural spoken story stimulus text, *The Little Prince* (LPP). The horizontal axis is organised by log-frequency in the Corpus of Contemporary English (COCA; Davies, 2008).

are story-character names. These attestations are evenly distributed throughout the text with about 100 examples in each section.

2.2. MWEs as indicators of memory retrieval

2.2.1. What are MWEs?

The designation "MWE" is motivated in computational linguistics whenever a sequence of words is better treated as an indivisible whole, rather than a structure that is composed out of sub-parts.

This pragmatic criterion (see e.g. Calzolari et al., 2002) picks out a very wide class of linguistic phenomena. To get a sense of this diversity, Table 1 (after Siyanova-Chanturia, 2013) cites some cases that might plausibly be analysed as MWEs. The specific MWE identification procedure used in this study is presented further below in Section 3.2.

2.2.2. MWE and compositionality

What unifies cases of MWE is the absence of a wholly compositional linguistic analysis (Sag et al., 2002). The naturalistic story used as a stimulus in this project attests nearly all of the types cited in Table 1; a few are shown below in Example (1).

(1) So I thought **a lot** about the adventures of the jungle and **in turn**, I managed with a coloured pencil to make my first drawing.
(2) My **little fellow**, I don't know how to draw anything except **boa constrictors**, closed and open.
(3) "What are you doing there?", he said to the drinker who he found sitting **in silence** in front of **a number** of empty bottles and **a number** of full bottles.
(4) You must **see to it** that you regularly **pull out** the baobabs as soon as they can be told **apart from** the rose bushes to which they look very similar to when they are young.

The syntactic and semantic properties of the boldfaced expressions in (1) indeed resist analysis by parts; they could be considered in some way as lexicalised.

Table 1. A wide variety of linguistic phenomena are plausibly treated as MWEs.

fixed phrases	*per se, by and large*
noun compounds	*black coffee, cable car*
verb compounds	*give a presentation, come along*
binomials	*heaven and hell, safe and sound*
complex prepositions	*in spite of*
idioms	*break the ice, spill the beans*

2.2.3. MWE processing

To the extent that MWEs are not compositional, their processing may depend more on memory retrieval than composition. This idea is most clear in the case of idioms where it has been dubbed the Configuration Hypothesis (CH) by Cacciari and Tabossi (1988).

> For the CH, idioms are processed word-by-word, just like any other piece of language, until enough information has accumulated to render the sequence of words identifiable as – or highly expected to be – a memorised chunk. Only at this point the idiomatic meaning is retrieved. (Cacciari, 2014)

Here, the CH is generalised in supposing that all categories of MWE exhibit this greater dependence on memory retrieval. Section 2.5.1, below, reviews experimental evidence regarding the processing of MWEs. Of course, whether or not retrieval is actually used on a particular occasion may reflect gradations in compositionality. These gradations, as the next section details, can be quantified precisely.

2.3. Lexical cohesion of MWEs

The crisp theoretical distinction of the CH between compositional and non-compositional expressions obscures the somewhat messier reality: expressions fall along a graded spectrum of compositionality (see Table 2). To quantify this gradient of MWE *cohesion*, we use a measure called Pointwise Mutual Information (PMI; Church & Hanks, 1990). PMI is commonly used in corpus linguistics to identify collocations.[2] Intuitively, PMI is high when the word sequence under consideration occurs more often together than one would have expected, based on the frequencies of the individual words (Manning & Schütze, 1999, §5.4). More formally, PMI is a log-ratio of observed and expected

Table 2. Graded spectrum of compositionality revealed by pointwise-mutual information.

PMI	Multiword expression receiving this score
26.59	heart skipped a beat
23.80	have nothing to do with
21.26	forehead with a handkerchief
21.18	burst into tear
20.17	once upon a time
20.15	boa constrictor
18.85	peal of laughter
−2.34	be order
−2.49	do calculation
−2.72	be object
−2.98	be hundred
−3.15	a well
−3.50	drink anything
−3.63	have plan

probabilities:

$$\mathrm{PMI} = \log_2\left(\frac{O}{E}\right) \quad (1)$$

where

$$O = \frac{\text{count(whole expression)}}{\text{corpus size}}$$

and

$$E = \frac{\text{count}(w_1)*\text{count}(w_2)*\cdots*\text{count}(w_n)}{\text{corpus size}^n}$$

In Definition 1 the numerator O is a probability for the n-word collocation, $w_1 w_2 \cdots w_{n-1} w_n$. The denominator E is what one would expect if the occurrences of each word in the collocation were probabilistically independent.

MWEs that receive a higher PMI score are lexically more cohesive, suggesting less compositionality, at least at the superficial level of word co-occurrence. To the extent that their greater cohesion results in a kind of word-sequence memorisation, these highly-cohesive MWEs are excellent candidates for a processing account based on memory retrieval. Conversely, low PMI scores signal a kind of "incohesion." These expressions might be better accounted-for by a compositional analysis.

It is important to appreciate that PMI is different from the raw attestation counts that go into it. This is because PMI reflects the *ratio* of the attestations of the whole relative to the attestations of the parts. It was raw counts, rather than the ratio PMI, that were used in a fMRI study by Yang et al. (2017) that examined memory retrieval in comparison to syntactic complexity. This study found no effects of frequency in the brain beyond the single-word level. Yang et al. (2017) did observe, however, an effect of syntactic complexity by parametrically varying six types of nominal, verbal and prepositional phrases. The next section presents an alternative way to investigate this same factor using naturalistic text.

2.4. Bottom-up parsing as composition

To study composition itself, some independent characterisation of compositional processes themselves is needed. In this study, bottom-up parsing plays that role. The intermediate states of this parsing strategy quantify the amount of compositional work that an idealised system would do, in the course of processing the naturalistic stimulus text introduced above in Section 2.1. While there exists a large literature on bottom-up parsing within computational linguistics (see e.g. Hale, 2014) its essential character is easy to grasp.

Bottom-up parsing amounts to a repeated cycle of choice: whether to `shift` to the next word or `reduce` a sequence of transient elements held in memory. As shown in Figure 2, `reduce` actions are individuated by particular grammar rules. The number of parser actions required at each word defines an incremental complexity metric. We use this complexity metric to quantify composition effort in the brain, following prior work in neurolinguistics (Brennan & Pylkkänen, 2017; Brennan, Stabler, Van Wagenen, Luh, & Hale, 2016; Brennan et al., 2012; Nelson et al., 2017) and psycholinguistics (Frazier, 1985).

2.5. Hypotheses regarding retrieval of MWEs and composition

With operational definitions sketched-out for both memory retrieval (Section 2.2) and composition (Section 2.4), this section turns to what is already known about their processing in the brain. Synthesising several different literatures, a general prediction emerges that these two processes are subserved by different brain areas.

2.5.1. MWE processing is different: experimental evidence

Prior work supports the claim that MWE comprehension is distinct from other language processing such as structural composition and involves additional memory

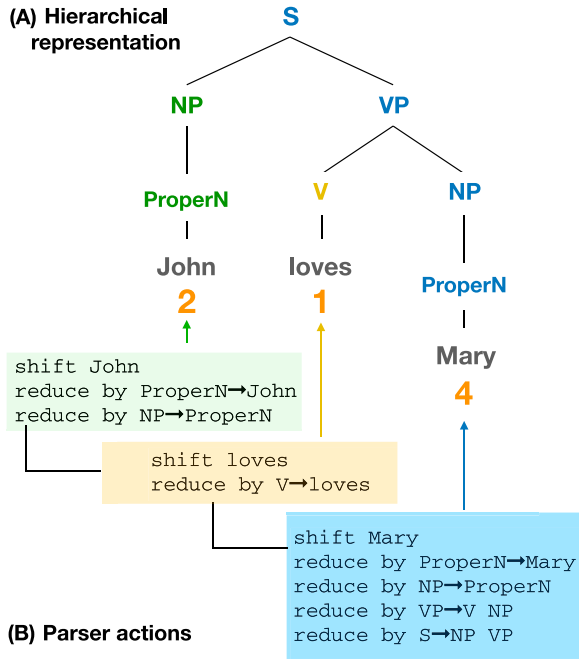

Figure 2. Panel (A) depicts hierarchical structure for *John loves Mary* to be recognised via processes of syntactic composition with the word-by-word parser action counts given in orange. Panel (B) shows the sequences of parser actions (i.e. `shift` and `reduce`) that would build the colour-coded tree nodes during bottom-up parsing.

retrieval. For instance, it is well-established at the behavioural level that MWEs are produced and understood faster due to their frequency, familiarity, and predictability (Siyanova-Chanturia & Martinez, 2014). This would follow if MWEs were remembered as chunks, in the sense of Miller (1956) that was later formalised by Laird, Rosenbloom, and Newell (1986) and Rosenbloom and Newell (1987).

Eye-tracking and EEG work further documents this processing advantage across a wide range of MWE sub-types, e.g. binomials (Siyanova-Chanturia, Conklin, & Schmitt, 2011), phrasal verbs (Yaneva, Taslimipoor, Rohanian, & Ha, 2017), complex prepositions (Molinaro, Canal, Vespignani, Pesciarelli, & Cacciari, 2013; Molinaro, Vespignani, Canal, Fonda, & Cacciari, 2008), nominal compounds (Molinaro & Carreiras, 2010; Molinaro, Carreiras, & Duñabeitia, 2012), lexical bundles (Tremblay & Baayen, 2010; Tremblay, Derwing, Libben, & Westbury, 2011), and idioms (Laurent, Denhières, Passerieux, Iakimova, & Hardy-Baylé, 2006; Rommers, Dijkstra, & Bastiaansen, 2013; Siyanova-Chanturia, Conklin, & Van Heuven, 2011; Strandburg et al., 1993; Underwood, Schmitt, & Galpin, 2004; Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010).

2.5.2. Neural basis for retrieving MWEs and composition

The hypothesis that MWEs involve additional memory retrieval reflects trends in both Construction Grammar and Connectionism (see e.g. Ellis, 2008). In these quarters, however, the relevant "memory" does not receive a precise psychological or neural characterisation. It may be that memory for MWE falls into a distinct theoretical category "between" episodic memory and general semantic memory (Renoult, Davidson, Palombo, Moscovitch, & Levine, 2012). Memories in this category are encoded on the basis of repeated personal experience, an idea that is certainly plausible for MWEs. Neuroanatomical sites for the category of memory that Renoult et al. (2012) discuss include medial temporal regions and the Precuneus. This last suggestion converges well with PET results on single word retrieval (Halsband, Krause, Sipilä, Teräs, & Laihinen, 2002).

Regarding composition operations, one prominent line of work implicates frontal regions, such as the pars triangularis and pars opercularis of the inferior frontal gyrus (Hagoort, 2016; Friederici & Gierhan, 2013; Snijders et al., 2009; Zaccarella & Friederici, 2017). Other lines of investigation implicate anterior temporal regions, based on deficit-lesion data (Dronkers et al., 2004) and text comprehension tasks (Ferstl, Neumann, Bogler, & Von Cramon, 2008). These anterior temporal areas are sensitive to parametric variation of phrase size, which would follow if they were somehow representing composed structures (Bemis & Pylkkänen, 2011a; Brennan et al., 2012; Pallier et al., 2011; Pylkkänen, 2016). These latter findings, although, conflict with patient studies showing that atrophy to anterior temporal regions does not systematically impair processing of syntactically complex sentences (Wilson et al., 2014a). This tension between functional and deficit/lesion results remains to be resolved.

This array of empirical work, in combination with the neurocognitive models of language processing discussed below (Section 2.5.3) serves to motivate the present investigation into the brain bases for MWE processing and compositional processing.

2.5.3. Neurocognitive models of language and localisation

Current neurobiological models for language suggest that retrieval and composition should be sub-served by different brain areas. These models do however differ both in their conceptualisation of these operations, and on their anatomical localisation.

The Declarative-Procedural model (Ullman, 2001) for instance, is founded upon a distinction that contrasts memory-related with non-memory-related processing. Ullman (2015) links rule-based mechanisms to frontal regions and sub-cortical structures, while memory for words is supported by medial temporal regions.

In Hagoort's Memory, Unification and Control framework (2016), composition falls under the scope of the Unification operation and is assigned to inferior frontal areas. While operating under different assumptions, this localisation is in general agreement with Zaccarella and Friederici (2017) who propose that hierarchical processing is subserved by a sub-part of the left inferior frontal gyrus. Regarding the Memory aspect of their model, Hagoort and colleagues agree partly with Ullman, associating that function (among others) to posterior temporal regions (Hagoort, 2009; Hagoort & Indefrey, 2014).

The Dual Streams Model (Hickok & Poeppel, 2007) similarly locates the Lexical Interface, where individual words would be processed, to posterior middle temporal gyrus. Syntactic phrases would be composed, part-by-part, by a combinatorial network within the anterior temporal lobes.

Another perspective is offered by the Extended Argument Dependency Model (eADM; Bornkessel-Schlesewsky & Schlesewsky, 2009, 2013), that divides up language processing in a different way. In this model sequential information (for instance about word order) is handled by a dorsal stream, while dependency information (as expressed through case-marking) is handled by a ventral stream. If MWE comprehension is sequential processing in this sense, then structures along this dorsal

stream, including the inferior parietal lobule, should be involved. On the other hand, composition should activate temporal regions along the ventral stream.

These models' localisation claims contrast in detail with each other, and with an important body of evidence from patient work (e.g. Dronkers et al., 2004; Wilson, Galantucci, Tartaglia, & Gorno-Tempini, 2012; Wilson et al., 2014b, inter alia). Despite this tension, they agree on the common proposal that retrieval and composition should each be independently localisable. It is this point of agreement that motivates the present study.

3. Methods

A spoken narrative serves as the stimulus. Participants hear this narrative over headphones while they are in the scanner (e.g. Brennan et al., 2012; Willems, Frank, Nijhof, Hagoort, & van den Bosch, 2016). The sequence of neuroimages collected during their session becomes the dependent variable in a regression against word-by-word predictors, derived from the text of the story. The overall approach to deriving time series predictions regarding the comprehension of this auditory stimulus is shown in Figure 3.

3.1. Subjects and design

Participants were fifty-one volunteers (32 women and 19 men, 18–37 years old) with no history of psychiatric, neurological, or other medical illness or history of drug or alcohol abuse that might compromise cognitive functions. All strictly qualified as right-handed on the Edinburgh handedness inventory (Oldfield, 1971). They self-identified as native English speakers and gave their written informed consent prior to participation, in accordance with Cornell University IRB guidelines.

3.2. Stimulus and MWE identification

The audio stimulus was a literary text, Antoine de Saint-Exupéry's *The Little Prince*. We applied two different approaches to MWE identification, which together yield 669 distinct MWE types for a total of 1292 attestations in the stimulus text. This section explains the two identification approaches.

The first approach uses `lgtagger` (Constant & Sigogne, 2011), a freely-available program that combines two submethods:

1 string-based matching to look up MWE in external dictionaries
2 a tagging model, based on conditional random fields and trained on hand-checked examples

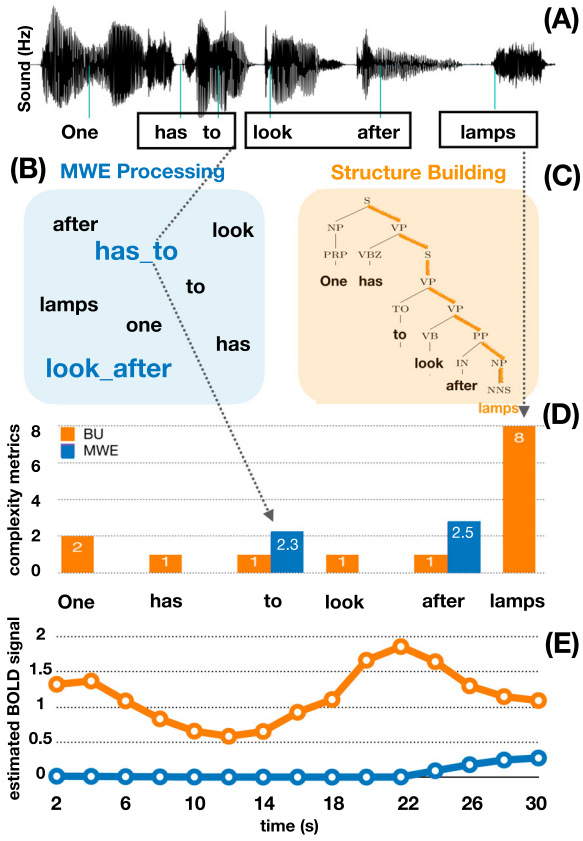

Figure 3. Deriving an expected blood-oxygen level dependent (BOLD) signal from a naturalistic text. (A) shows a segment of the spoken stimulus, with word boundaries in light blue. (B) highlights multiword expressions (MWEs), which may be retrieved directly from memory during comprehension and (C) depicts the phrase structure of the example sentence, which is composed part-by-part during comprehension. In (D) the value of these word-by-word predictors are shown together. The orange values reflect steps taken between each word by a bottom-up (BU) parser. The purple values reflect PMI scores, which quantify the cohesion of MWEs like *has to* and *look after*. Panel (E) shows the expected BOLD response, after these predictors are convolved with a haemodynamic response function (HRF).

Regarding submethod 1, the external dictionaries are: the Unitex lexicon (Paumier, Nakamura, & Voyatzi, 2009), the SAID corpus (Kuiper, McCann, Quinn, Aitchison, & van der Veer, 2003), the Cambridge International Dictionary of Idioms (White, 1998), and the Dictionary of American Idioms (Makkai, Boatner, & Gates, 1995).

Regarding submethod 2, conditional random fields are used to guess MWE marks like those shown in the final column of Table 3. Smith (2011, §3.5) offers a general introduction to the technique. The key idea here is to construe a word's MWE status as a latent variable that is related to the surface string via a collection of weak clues, the "features." Here the training examples come from MWE annotations in the English Universal

Table 3. Training data for MWE identification systems. Each word of the sentence is on a separate line with its lemma and its POS tag. + and − indicate whether the word is part of a MWE or not.

Idx	Word form	Lemma	POS	MWE
1.	I	i	PRP	−
2.	was	be	VBD	−
3.	thinking	think	VBG	−
4.	of	of	IN	−
5.	the	the	DT	+
6.	other	other	JJ	+
7.	day	day	NN	+
8.	…			

Dependencies English Web Treebank (Silveira et al., 2015). This annotated data comprises several web-related genres including weblogs, newsgroups, emails, reviews, and Yahoo! answers. The conditional random field makes its marking decisions based on feature templates that specify, for example, particular lexical items, capitalisation, or part of speech tags. A complete table of these templates is presented in Constant and Tellier (2012). The approach is limited to contiguous MWEs and identifies mainly fixed expressions such as complex function words and nominal MWEs. Crucially, this way of identifying MWEs is blind to hierarchical syntactic structure.

The second approach uses a transition based system (Al Saied, Candito, & Constant, 2017). This system is a variant of the well-known Nivre parser (Constant & Nivre, 2016), in which abstract "actions" update an abstract computational state that moves through the text, emitting MWE marks as a side-effect. The choice of which action to take is made by a classifier, here a support vector machine (for textbook introductions see e.g. Abney, 2007, §6.4 or Murphy, 2012, §14.5). There are actions that Add or Remove words from the computational state, as well as actions which Mark candidates as MWEs. This last type is actually divided into two cases, one of which composes candidate MWEs in a binary fashion, and another which marks these binary elements in a way that produces output. Figure 4 shows how this transition-based system would apply to the example "see to it."

The input to this classifier includes word form, lemma and part-of-speech information as shown in Table 3, as well as dictionary-based features and features that relate to previous actions ("history-based"). The classifier's output is a selection of one particular action. This second way of identifying MWEs is similarly denied access to any hierarchical syntactic information.

The training data for this transition-based system come from the Children's Book Test (Hill, Bordes, Chopra, & Weston, 2015), which is close to the genre of the stimulus narrative. Training examples were generated by string-matching based on the external dictionaries cited above. Compared to the conditional random field-based approach, this second approach seems better-able to find verbal MWE.

The MWEs used in the fMRI study reported here were identified by applying both of the approaches described above to the stimulus text. The union of their outputs yielded a time series of indicator variables as shown in Panel C of Figure 5. This time series has the value 1 on words that are final in a MWE, and 0 otherwise.

PMI scores for each MWE define a gradient metric as specified in Definition 1. The probabilities used in this calculation were estimated using the Corpus of Contemporary English (COCA; Davies, 2008). The observed probability O of entire MWEs was estimated by querying the 2018 version of COCA, which contains 560 million words. The expected probabilities E are based on an earlier COCA release that comprises 450 million words. These estimates are based on counts of lemmas, rather than inflected words. This use of lemmas serves to focus the analysis on the cohesion of the expression rather than its morphology.

3.3. Annotating compositional structure

Compositional processing was modelled as bottom-up parsing, as introduced above in Section 2.4. The tree structures were obtained using Stanford parser (Klein & Manning, 2003) and follow the conventions of the Penn Treebank (Marcus, Marcinkiewicz, & Santorini, 1993). They capture constituency relationships, including phrase type and embedding, but do not explicitly mark long-distance dependencies. The complexity metric is defined as the number of `reduce` actions that would be taken during bottom-up parsing of these trees. Since narrative stimulus does not readily lead to misunderstandings or "garden-path" effects in the sense of Bever (1970) we restrict ourselves to just the globally-correct parser action sequence.

Figure 5 illustrates the two retrieval estimators and the one compositional estimator for a single sentence from *The Little Prince*. Correlation matrices for all estimators are given in Appendix B.

3.4. Data presentation

After giving their informed consent, participants were familiarised with the MRI facility and assumed a supine position on the scanner gurney. The presentation script was written in PsychoPy (Peirce, 2007). Auditory stimuli were delivered through MRI-safe, high-fidelity headphones (Confon HP-VS01, MR Confon, Magdeburg, Germany) inside the head coil. The headphones were secured against the plastic frame of the coil using foam

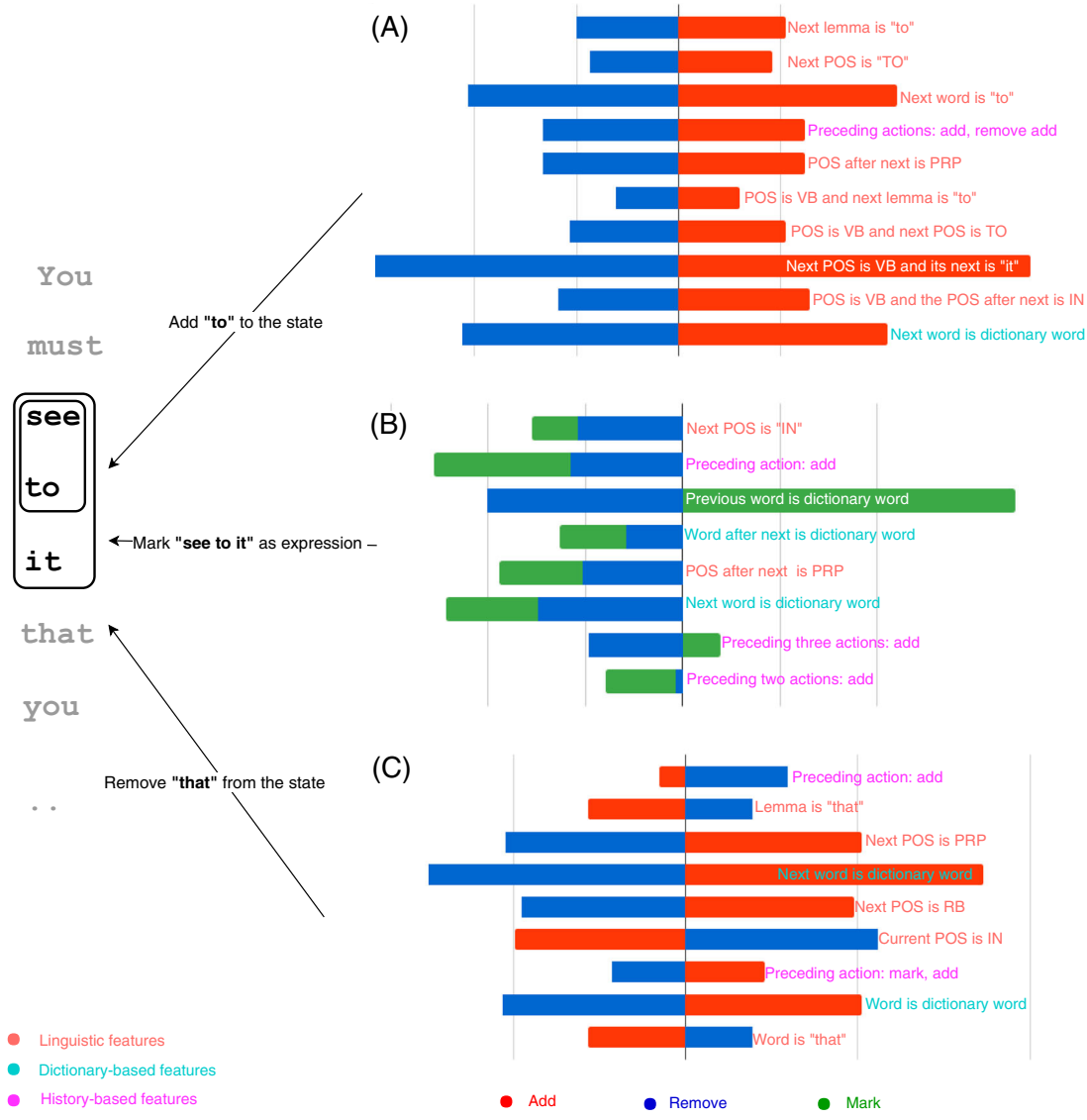

Figure 4. A transition-based approach to finding the multiword expression *see to it* from Example 1(d). Identification involves adding *see*, *to* and *it* to consecutive computational states, creating a binary element and marking this element as a multiword expression. For each state, the classifier must choose the appropriate action among the four available actions. Panel (A) shows the effect of some features on choosing the action **Add** instead of the action **Remove** for the word "to." Panel (B) shows the effect of choosing **Mark** instead of **Remove** for the composed element "see to it." Panel (C) show the effect of some other features on choosing the action **Remove** instead of **Add** for the word "that." These feature templates help the classifier choose the right action. For display purposes we show feature weights as they would be used in a simpler Logistic Classifier, rather than the actual support vector-machine (SVM) that is used in the model. In this simplified setup, decisions are made by passing the summed feature weights – the coloured bars – through a logistic function. The SVM decision rule is more complicated. For details, see Al Saied et al. (2017).

blocks. Using a spoken recitation of the US Constitution, an experimenter increased the volume until participants reported that they could hear clearly. Participants then listened passively to the audio storybook for 1 hour 38 minutes. The story was divided into nine sections and at the end of each section the participants were presented with a multiple-choice questionnaire with four questions (36 questions in total), concerning events and situations described in the story. These questions served to confirm participants' comprehension. They were viewed via a mirror attached to the head coil and answered through a button box with their right hand. The entire session lasted around 2.5 hours.

3.5. Data acquisition

Imaging was performed using a 3T MRI scanner (Discovery MR750, GE Healthcare, Milwaukee, WI) with a 32-channel head coil at the Cornell MRI Facility. Blood Oxygen Level Dependent (BOLD) signals were collected

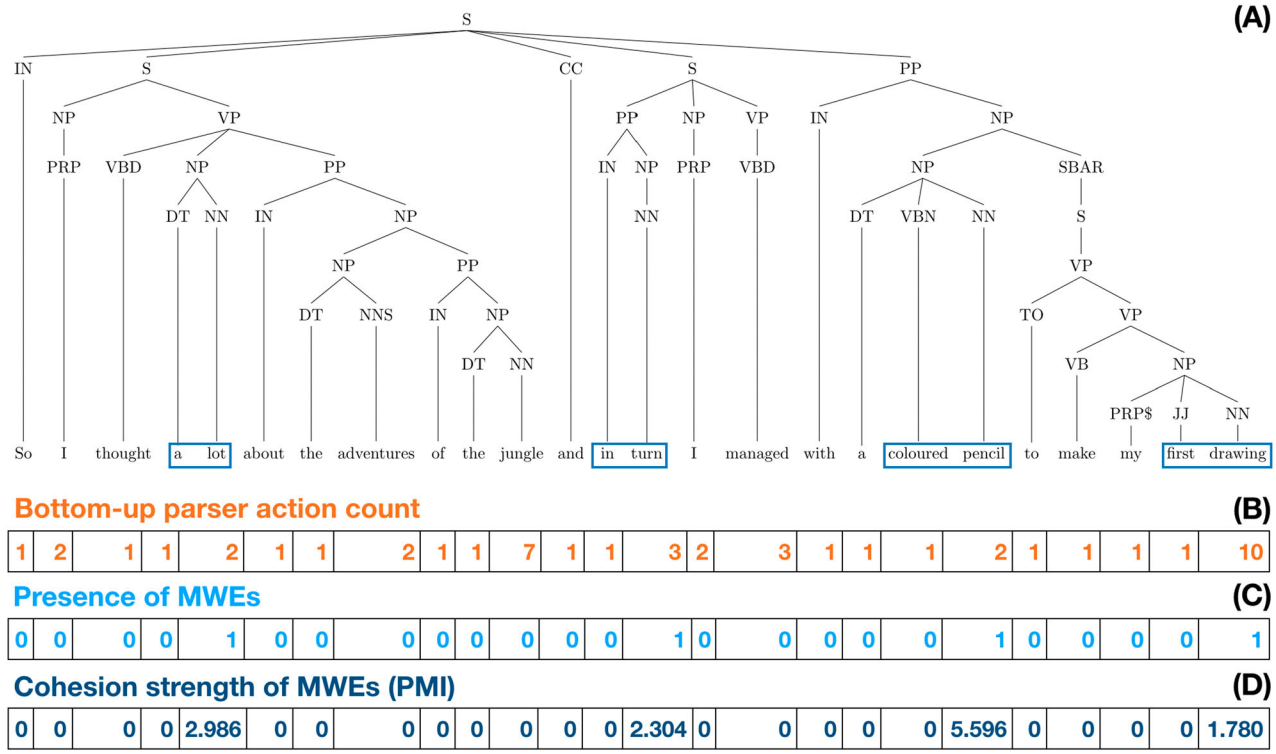

Figure 5. Comparing the word-by-word predictors on a single sentence: (A) depicts the syntactic structure of a single sentence of the naturalistic stimulus text, as recovered by the Stanford parser, (B) gives the bottom-up parser action count with respect to this tree, represented in orange and annotated at the offset of each word in the story, (C) is the categorical multiword expression predictor, represented in blue, where 0 or 1 is tagged at the offset of a word if it is the final word of a MWE, (D) is the gradient multiword expression predictor, represented in dark blue, and similar to the previous predictor the offset of the final word in a MWE is tagged with the corresponding PMI score, quantifying its lexical cohesion.

using a T2-weighted echo planar imaging (EPI) sequence (repetition time: 2000 ms, echo time: 27 ms, flip angle: 77°, image acceleration: 2X, field of view: 216 × 216 mm, matrix size 72 × 72, and 44 oblique slices, yielding 3 mm isotropic voxels). Anatomical images were collected with a high resolution T1-weighted (1 × 1 × 1 mm^3 voxel) with a Magnetisation-Prepared RApid Gradient-Echo (MP-RAGE) pulse sequence.

4. Data analyses

4.1. Preprocessing

FSL's Brain Extraction Tool (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2004) was used for skull-stripping with a fractional intensity threshold setting of 0.5. Subsequent preprocessing steps were carried out using AFNI version 16 (Cox, 1996). Anatomical and functional images were co-registered using the in-built AFNI function `3dseg`, images were normalised to the MNI–152 template, and images were resampled to 2mm isotropic voxels.

We used multi-echo independent components analysis (ME-ICA) (Kundu, Inati, Evans, Luh, & Bandettini, 2012; Kundu et al., 2013) to improve the signal-to-noise ratio in these data. ME-ICA splits the *T2** signal into BOLD-like and non BOLD-like components. Removing these non-BOLD components mitigates noise due to participants' head motion, physiology and scanner conditions such as thermal changes (Kundu et al., 2017). Indeed, there were no exclusions based on degree of head movement. Nor was any high-pass filtering or smoothing applied at this stage.

4.2. Statistical analysis

The research questions layed out above in Section 2 motivate two statistical analyses. The first analysis localises composition and memory-retrieval operations during naturalistic listening. The second analysis investigates multiword expressions along a quantitative gradient of cohesion. Both analyses employ the General Linear Model, and were carried out using SPM12 (Friston, Ashburner, Kiebel, Nichols, & Penny, 2007). The predictors were convolved using SPM's canonical HRF.

Table 4. Predictors used in Analysis 1.

Bottom-up parser action count	Number of REDUCE actions taken since the last word (§2.4)
Categorical MWE predictor	1 at the last word of a MWE, 0 otherwise (§2.2)
Word rate	Indicator for spoken word offset
Word frequency	log-frequency in movie subtitles (Brysbaert & New, 2009)
f0	fundamental frequency of the narrator's voice, reflects pitch
RMS amplitude	intensity, an acoustic correlate of volume

4.2.1. Analysis 1: with categorical predictors

We regressed the word-by-word predictors described below against fMRI timecourses recorded during passive story-listening in a whole-brain analysis. Along with the parser action count and MWE indicators of theoretical interest, four "nuisance" variables of non-interest were entered into the GLM analysis, including a unigram word frequency predictor based on attestations in movie subtitles (Brysbaert & New, 2009). This enables the effects of MWE, Parser-action count, and MWE cohesion to be assessed in way that is distinct from single-word frequency effects. We also included two variables that account for prosodic effects: the RMS amplitude and the fundamental frequency of the auditory stimulus. These control variables serve to improve the sensitivity, specificity and validity of activation maps (Bullmore et al., 1999; Lund, Madsen, Sidaros, Luo, & Nichols, 2006). The predictors entered into Analysis 1 are summarised in Table 4. Figure 5 shows the theoretical predictors' time series for a particular stimulus sentence.

4.2.2. Analysis 2: with gradient predictors

Analysis 2 uses the same predictors as in Analysis 1, except that the categorical indicator of MWE presence is replaced with a gradient predictor. As explained above in Section 2.3, this predictor rates the propensity of a MWE to be memorised or not using a standard measure of collocation strength called pointwise mutual information (PMI). The same MWEs that were annotated with a value of "1" in Analysis 1 are in Analysis 2 marked with their PMI score. This is shown in the bottom row of Figure 5.

Note that parser action count is not correlated with the categorical or the gradient MWE predictor ($r = 0.09$ in both cases). Correlation matrices for all terms entered into the regression analyses are given in Appendix B.

4.2.3. Group-level analysis

In the second-level group analysis, each contrast was analysed separately at the group-level. An 8 mm FWHM Gaussian smoothing kernel was applied on the contrast images from the first-level analysis to counteract inter-subject anatomical variation. All the group-level results reported in the next section underwent FWE

Table 5. Significant clusters of increasing activation for bottom-up parser action count after FWE voxel correction for multiple comparisons with $p < 0.05$ and cluster-extent threshold ($k > 50$). Peak activation is given in MNI Coordinates, and brain region labels come from the Harvard-Oxford Cortical Structure Atlas.

		MNI coordinates				
Regions for bottom-up parser action count	Cluster size (in voxels)	*x*	*y*	*z*	*p*-value(corrected)	*T*-score(peak level)
R Anterior Temporal	4816	52	6	−20	0.000	13.20
R Middle Temporal Gyrus		50	−20	−10	0.000	11.31
R Supramarginal Gyrus/Superior Temporal Gyrus		60	−40	−10	0.000	10.11
L Inferior Frontal Gyrus Orbitalis/Triangularis & anterior Insula (BA47)	2461	−36	18	−14	0.000	10.40
L Temporal Pole		−50	6	−26	0.000	8.30
L Putamen		−30	8	−4	0.000	6.99
R Supplementary Motor Area/Superior Frontal Gyrus (BA9)	6495	10	18	62	0.000	9.35
R Medial Superior Frontal Gyrus (BA9)		12	58	32	0.000	8.62
L Superior Frontal Gyrus		−8	18	66	0.000	8.24
L Cerebellum - Crus I	448	−24	−74	−30	0.000	8.96
R Cerebellum - Crus I	941	26	−74	−36	0.000	8.15
R Cerebellum		36	−60	−32	0.021	5.16
L Middle Occipital Gyrus/Fusiform Gyrus	1084	−34	−78	12	0.000	7.59
L Fusiform Gyrus/Temporal Occipital Cortex		−30	−58	−10	0.000	7.19
L Occipital Fusiform Gyrus		−28	−70	−14	0.010	5.85
R Precentral Gyrus	159	42	0	48	0.000	7.57
L Supramarginal Gyrus/Parietal Lobe (BA40)	665	−54	−56	30	0.000	7.35
L Parietal Lobe		−48	−66	50	0.032	5.45
L Supramarginal Gyrus		−52	−58	50	0.036	5.41
R Temporal Occipital Cortex/Fusiform Gyrus (BA19)	164	30	−50	−10	0.001	6.75
L Inferior Frontal Gyrus Orbitalis/Frontal Pole (BA11)	252	−44	46	−12	0.001	6.61
L Frontal Pole		−36	60	−6	0.013	5.75
L Middle Frontal Gyrus (BA9)	252	−42	24	44	0.001	6.49
L Precuneus	154	−10	−52	38	0.003	6.25
R Middle Occipital Gyrus	160	28	−72	22	0.003	6.19
L Caudate	54	−14	16	10	0.005	6.07

voxel correction for multiple comparisons which resulted in *T*-scores > 5.3.

5. Results

Behavioural results of the comprehension task showed attentive listening to the auditory story presentation. Across 51 participants, average accuracy on the comprehension questions was 90% (SD = 3.7%). All whole-brain effects reported survived a $p < 0.05$ Family-Wise-Error threshold at the voxel level. Tables introduced below use brain region labels from the Harvard-Oxford Cortical Structure Atlas.

5.1. Analysis 1: results with categorical predictors

5.1.1. Results for composition

Bottom-up parser action count shows a broad activation pattern both in right and left hemisphere. The peak activation is right lateralised in the anterior temporal lobe within a main cluster of activation which extends through the middle and superior temporal gyri. While anterior temporal activation is bilateral, both middle temporal gyrus and posterior superior temporal gyrus are only right lateralised. The second strongest cluster of increased activation is observed in the left inferior frontal gyrus stretching over pars orbitalis and triangularis and extending to the anterior insula and the putamen. A similar increased activation is observed in the right inferior frontal gyrus.

Additional activation clusters were seen in areas including the prefrontal cortex, supplementary motor areas, the cerebellum, the left temporal-parietal junction, as well as inferior temporal lobes. These are all detailed in Table 5 and Figure 6.

5.1.2. Results for MWE presence

The categorical MWE predictor gives rise to two clusters of activation both in the right precuneus cortex, as presented below in Figure 6 and Table 6.

5.2. Analysis 2: results with gradient predictors

Analysis 2 investigates memory retrieval further by rating how cohesive each MWE is. This is done using pointwise mutual information, introduced above in Section 2.3. Increasing MWE cohesion, as seen through the positive correlation with PMI, yields a single cluster in the right precuneus.

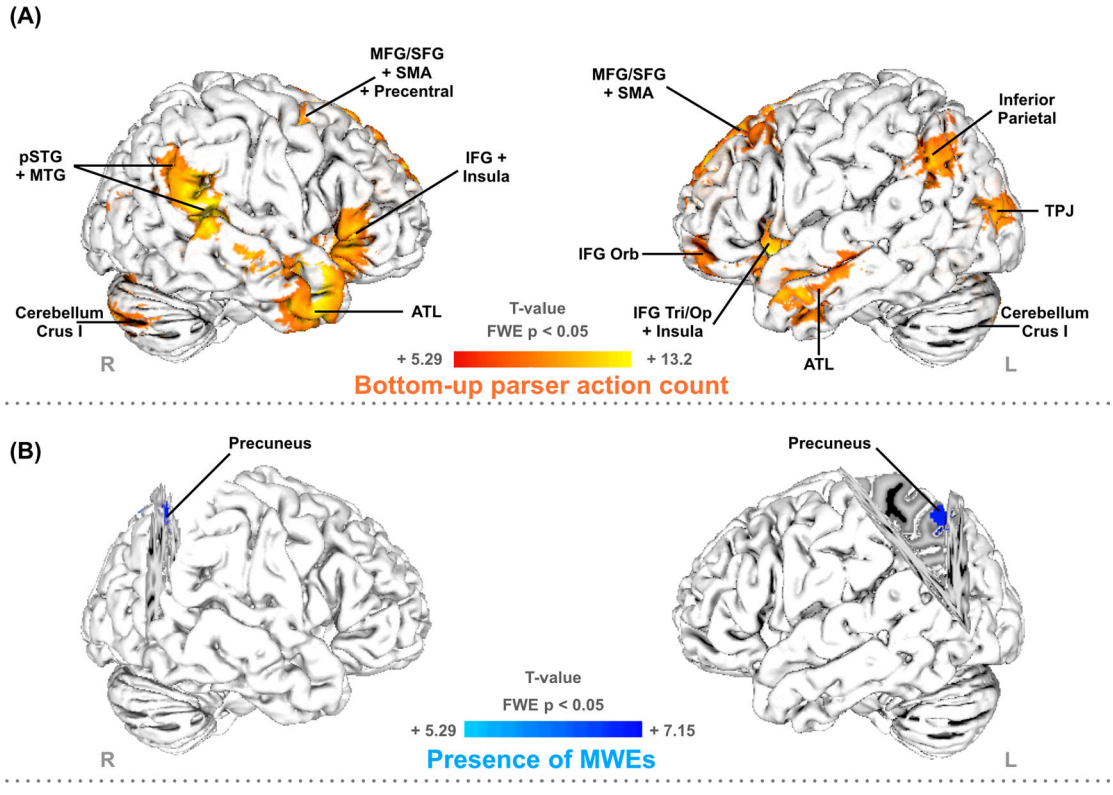

Figure 6. Whole brain contrast images with significant clusters are projected onto a template brain (Holmes et al., 1998): Panel (A) shows us the significant clusters Bottom-up parser action count in orange; Panel (B) shows the significant clusters for Multiword expressions in blue. All images are underwent FWE voxel correction for multiple comparisons with $p < 0.05$. A cluster-extent threshold $k > 50$ is applied for display purposes.

Table 6. Significant clusters of increasing activation for multiword expressions after FWE voxel correction for multiple comparisons with $p < 0.05$. Peak activation is given in MNI Coordinates.

Regions for multiword expression	Clustersize (in voxels)	MNI coordinates			*p*-value (corrected)	*T*-score (peak level)
		x	*y*	*z*		
R Precuneus	209	6	−70	56	0.000	7.15
R Precuneus	18	6	−48	50	0.019	5.63

Table 7. Significant clusters for the increasing and decreasing cohesion measure of MWEs after FWE voxel correction for multiple comparisons with $p<0.05$. Peak activation is given in MNI Coordinates.

Regions for PMI	Cluster size (in voxels)	MNI coordinates			*p*-value (corrected)	*T*-score (peak level)
		x	*y*	*z*		
Correlated with increasing MWE cohesion						
R Precuneus	244	6	−68	56	0.000	7.33
Correlated with decreasing MWE cohesion						
L Superior Frontal Gyrus	2039	−18	32	52	0.000	8.39
L Precentral Gyrus (BA9)		−44	8	40	0.000	7.26
L Middle Frontal Gyrus		−38	22	46	0.000	6.89
L Angular Gyrus	688	−42	−58	34	0.000	7.27
L Inferior Parietal Lobule		−48	−46	50	0.000	5.76
L Inferior Temporal Gyrus	320	−60	−44	−4	0.000	7.25
L Inferior Frontal Gyrus Pars Triangularis	211	−46	30	18	0.000	6.41
L Middle Temporal Gyrus	152	−56	0	−32	0.000	6.49
L Frontal Pole (BA10)	50	−6	64	2	0.001	5.88
R Superior Frontal Gyrus	35	14	52	2	0.001	6.15
L Inferior Frontal Gyrus orbitalis	33	−38	48	18	0.001	5.55
R Inferior Temporal Gyrus	21	58	−10	−32	0.001	5.78
R Superior Frontal Gyrus/ SMA (BA6)	15	12	24	58	0.001	5.94

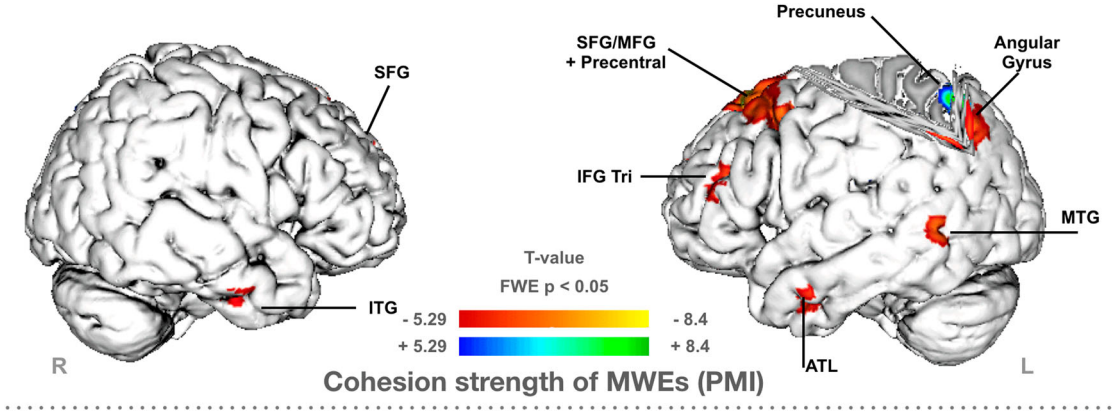

Figure 7. Significant clusters for the increasing and decreasing cohesion measure of MWEs after FWE voxel correction for multiple comparisons with $p < 0.05$. Increasing cohesion measures are represented in blue and decreasing cohesion measures are represented in orange.

Left-lateralised activity in superior frontal gyrus, angular gyrus, pars triangularis, posterior middle temporal gyrus, and frontal pole was observed in proportion to decreasing lexical cohesion, as seen through the negative correlation with PMI scores. These are detailed in Table 7 and in Figure 7.

6. Discussion

This study investigated the neural substrates of two processes in language comprehension: retrieval and composition of stored linguistic elements. Departing from prior work, we distinguish these operations using computational methods and a naturalistic stimulus. Multiword expressions identified in *The Little Prince* using the methods of Al Saied et al. (2017) and Constant and Sigogne (2011) serve as a hypothesis about points in the narrative where extra memory retrievals would occur.

6.1. Lexical retrieval and MWEs

Significant activation in the precuneus for MWEs in both Analysis 1 and Analysis 2 offers initial support for the idea

that this region subserves memory retrieval of these stored elements.

6.1.1. Precuneus and retrieval of MWEs

Both the categorical Analysis 1 and the gradient Analysis 2 indicate the centrality of the precuneus for MWE-related memory processes. This area features a graded effect such that it is more active for increasingly cohesive MWEs, as estimated by the PMI metric. As discussed above in Section 2.3, highly-cohesive MWEs are excellent candidates for a processing account based on direct retrieval from memory.

The involvement of the precuneus in the sort of memory retrieval prompted by MWEs is consistent with studies attesting its participation in larger memory networks, e.g. for verbal material (Halsband et al., 2002). The functional characterisation of the precuneus as part of a network sub-serving memory tasks has been reported for different memory-based processes, such as verbal memory (Halsband et al., 2002), spatial memory (Wallentin, Weed, Østergaard, Mouridsen, & Roepstorff, 2008), episodic memory (Andreasen et al., 1995), memory-related imagery (Fletcher et al., 1995; Mashal, Vishne, & Laor, 2014). Notably, some studies found that the Precuneus is active in non-imagery related episodic memory for musical sequences or abstract words (e.g. Platel et al., 1997).

The precuneus has also been identified as a part of the Default Network, perhaps playing a role in a Dorsal-Medial subsystem of that network (Andrews-Hanna, Smallwood, & Spreng, 2014). This subsystem is said to support story comprehension among other aspects of self-generated thought. Although the precuneus has been designated as part of the Protagonist's Perspective Interpreter Network (Mason & Just, 2006) and attested in naturalistic reading by Wehbe et al. (2014), an interpretation of the results of this particular study in terms of reference to story characters is implausible. This is because less than 2% of the MWEs in the stimulus narrative are names of story characters.

More broadly, the postero-medial portion of the parietal lobe has been widely linked to processing of complex lexical information by previous studies. Its sensitivity to the number of complements of a verb was reported by Shetreet, Palti, Friedmann, and Hadar (2007) in the medial precuneus and the anterior cingulate cortex (see also den Ouden, Fix, Parrish, & Thompson, 2009; Shetreet, Friedmann, & Hadar, 2010). The precuneus has also been linked to lexical processing of information of a relatively high complexity (Shetreet, Friedmann, & Hadar, 2009). These results converge with a strand of the fMRI literature over the past decade that has characterised precuneus as a language-relevant region.

6.1.2. Incohesive MWEs and Broca's area

While highly cohesive expressions evoke a focal pattern of activation that is distinct from classical left-hemisphere language areas, the cortical sites that showed a sensitivity to a decrease in MWEs' cohesion strength include the left superior frontal gyrus and also encompass several areas of the perisylvian language network. Notably the left inferior frontal gyrus, or Broca's area (pars triangularis and pars orbitalis), and also anterior and posterior regions of the left temporal gyrus. These perisylvian areas have been connected with compositional processes (e.g. Pallier et al., 2011), and this is consistent with the leading idea that less-cohesive expressions are more likely to require compositional processes. Indeed a similar set of regions are activated in response to the Bottom-up parsing predictor, as discussed further in the next sub-section.

6.2. Parser action count as composition

Word-by-word composition effort, quantified in terms of bottom-up parser actions, correlates in Analysis 1 with a highly bilateral pattern across several areas in the language network. Notably, both anterior-frontal and anterior-temporal regions are involved bilaterally.

6.2.1. Anterior frontal and anterior temporal regions

Inferior frontal regions encompassing different sub-parts of the inferior frontal gyrus and anterior insula are commonly attributed a role in compositional processes (Friederici & Gierhan, 2013; Hagoort, 2016; Snijders et al., 2009; Zaccarella & Friederici, 2015), and our findings are consistent with that work. Interestingly, prior work using methods most similar to ours (bottom-up parsing, a naturalistic stimulus) have *not* highlighted the inferior frontal gyrus (Brennan et al., 2012, 2016). We hesitate to draw strong inferences from comparing "significant" to "non-significant" results, but this discrepancy may reflect on the increase in statistical power afforded by our current study, which uses a stimulus that is over seven times longer that used by Brennan and colleagues, and also reports data from almost twice as many participants.

The balance of activation that we observe in response to increasing parser action count is consistent with Ullman's Declarative-Procedural model. It shows a predominately anterior distribution in left hemisphere, alongside smaller activation clusters in the inferior parietal lobe and temporo-occipital-parietal junction. Further, the observed activation of putamen within the large bilateral cluster in inferior frontal gyrus and anterior

insula conforms to Ullman's prediction that rule-governed language use is instantiated in a Fronto-basal ganglia network (Ullman, 2004, 2015).

The composition effect observed bilaterally in anterior temporal lobe is consistent with previous work on compositional in naturalistic narrative (Brennan et al., 2012). It confirms and extends a broad range of prior findings, including those from simple two-word phrases (Bemis & Pylkkänen, 2011b), from parametric variation of constituent size (Pallier et al., 2011), and comparisons of simple sentences to unstructured lists of words (Humphries, Binder, Medler, & Liebenthal, 2006; Rogalsky & Hickok, 2009; Stowe et al., 1998; Vandenberghe, Nobre, & Price, 2002; Xu, Kemeny, Park, Frattali, & Braun, 2005). Our result also conforms with evidence that anterior temporal lesions impair "basic levels of constituent-structure processing" (Dronkers et al., 2004, p. 161).

Recent neuroimaging work correlating brain activity with processing-complexity metrics across grammar types and parsing strategies has convergently pointed to the anterior temporal complex (Brennan et al., 2016; Li & Hale, 2018; for a review see Brennan, 2016). The precision of this localisation has also been supported by results using intracranial recording (Nelson et al., 2017).

The results reported here, along with previous findings, underline the involvement of anterior temporal lobe in basic composition processes (Friederici & Gierhan, 2013; Hickok & Poeppel, 2007). Yet, alongside many of the findings just reviewed, our results stand in contrast to the apparent stability of compositional processes following anterior temporal atrophy (e.g. Wilson et al., 2014a, 2016). The present study does not resolve the tension between these two literatures. We speculate that such a resolution might involve more detailed consideration of the specific mechanisms that contribute to observed neural activation patterns, and perhaps how compensatory processes respond to neuronal atrophy.

The unexpected right lateralisation of posterior temporal activation may suggest that the way the human brain processes linguistic stimuli within a contextually rich setting – one more similar to the everyday language environment – shows a strongly bilateral involvement of the language network. Future work that parametrically varies contextual richness, from more isolated to more naturalistic stimuli, may shed light on this speculation.

6.2.2. Left posterior temporal regions and parser action count

The operationalisation of compositional processing in terms of bottom-up parser actions highlights the language network to a large extent. However, we did not find a significant correlation between this metric and activity in left posterior temporal areas. Our result contrasts with posterior temporal effects of composition that are reported in some prior imaging studies (e.g. Brennan et al., 2016; Pallier et al., 2011; Snijders et al., 2009; Vandenberghe et al., 2002) and are also highlighted in studies of primary progressive aphasia (Wilson et al., 2012). This absence of evidence should be treated carefully; caution is especially warranted as different studies take different perspectives on composition itself.

The phrase-structures used in our analysis only capture some aspects of composition, and so a more comprehensive view of this aspect of language processing may shed light on discrepancies between studies. For example, the LPTL composition effect reported by Brennan et al. (2016) did not appear when analyzing the same types of tree structures used in the present analysis, but only from tree structures that explicitly encode long-distance dependencies and other more abstract aspects of grammar. Furthermore, the bottom-up parsing metric is unlexicalised. It counts distances, for instance along the spine of a tree as in Figure 3(C), but it treats all nodes as equal, ignoring lexical information except at the preterminal level. Subjects and direct objects receive the same score if they close off the same number of constituents. This conception of composition leaves out many classical distinctions that are known to affect cerebral activity, such as semantic selection, argument structure (Frankland & Greene, 2015; Thompson et al., 2007), long-distance dependencies (Ben-Shachar, Palti, & Grodzinsky, 2004; Santi & Grodzinsky, 2010; Shetreet & Friedmann, 2014), and case-checking (Nieuwland, Martin, & Carreiras, 2012).

Such a more comprehensive perspective may also shed light on another aspect of the present results. While parser action count does not correlate with left posterior temporal activation, we do see such a correlation for *decreasing* MWE cohesion (see Section 6.1.2). Such a correlation is consistent with the hypothesis that less coherent expressions load more heavily onto certain compositional operations, and such compositional operations evoke, at least in some cases, posterior temporal activity.

Overall the results support the idea that composition and retrieval of stored linguistic elements like MWEs involve different parts of the language network. They also confirm that bilateral anterior frontal regions play an important role in linguistic composition during naturalistic spoken story comprehension.

6.2.3. Composition and the cerebellum

An often reported but less-discussed area in language processing is the Cerebellum. In this study its relationship to parser action count in Crus I suggests its involvement

in compositional processes bilaterally. A growing number of studies have provided evidence for cerebellar involvement in non-motor language functions (see Stoodley & Schmahmann, 2009 for a meta-analysis), revealing interesting patterns of co-activation with classical left prefrontal language areas and right cerebellar hemisphere during language tasks and comprehension (Fedorenko, Duncan, & Kanwisher, 2012), or during the processing of complex syntactic structures (Christensen, 2008; Fabre, 2017; Shetreet & Friedmann, 2014). Specifically, Crus I and Lobule VII are typically proposed as members of Prefrontal-cerebellar connectivity loops (Stoodley & Schmahmann, 2009).

Moreover, an emerging view of perceptual and cognitive processing in the Cerebellum puts special emphasis on sequencing and predictive processes (Leggio et al., 2008; Molinari, Chiricozzi, et al., 2008), which are presumably also required in syntactic sentence parsing operations. Although no special modelling of predictive processing was undertaken in this study, bottom-up approaches such as the LR parser do include predictive information in their control state (see e.g. Shieber & Johnson, 1993; Stabler, 1991).

6.2.4. Composition of form and of meaning

The foregoing discussion has remained neutral as to whether activation within the anterior temporal lobe and other brain areas that correlate with bottom-up parsing reflects syntactic processing, or compositional semantic processing. Syntactic structure-building is transparently associated with bottom-up parser action count (panel B of Figure 5) that is itself defined in terms of syntactic phrases such as NP, VP and S. However, our approach is not able to rule out a semantic explanation. A semantic role would follow within approaches to natural language grammar that adopt the "rule-to-rule" assumption (Bach, 1976). Following the seminal work of Montague (1974) these theories of linguistic competence suppose that syntactic rules generally are paired with corresponding semantic rules. This organisation of grammar, which is introduced briefly in Barker AND Jacobson (2007), typifies Combinatorial Category Grammar (Steedman, 2000), Generalized/Head-driven Phrase-Structure Grammars (Klein & Sag, 1985) and certain forms of Minimalism (see e.g. Kobele, 2006, §2.2). A consensus from these conceptions of grammar is that aspects of linguistic performance that correlate with syntactic processing will necessarily also correlate with semantic processing.

Of course, the particular rules being processed do matter, and in this case the rules are plainly syntactic in nature. To tease apart semantic from syntactic explanations, future work could develop an alternative grammar, unlike the Penn Treebank phrase structures used here, and evaluate whether this alternative semantically-based theory offers a better explanation of the observed fMRI timecourses. Such an approach may be particularly fruitful in constructions where an additional semantic rule of "coercion" applies, unaccompanied by a corresponding syntactic rule (Pylkkänen, 2008). These considerations, regarding the kinds of rules that are used during comprehension, are in any event quite orthogonal to questions about their order of application (Hale, 2014)

7. Conclusion

Analysis of MWEs and parser action counts in naturalistic spoken story comprehension supports the localisation of memory retrieval to the precuneus in a way that trades-off with the other well-known language regions such as inferior frontal gyrus and anterior temporal lobe. This trade-off seems to reflect the degree to which particular MWE form a cohesive unit. These findings are broadly consistent with several contemporary neurocognitive models of language processing, such as the proposals of Hagoort (2016), Friederici and Gierhan (2013) and Ullman (2015).

Apart from the light that they shed on the neural bases of these two language-related cognitive processes, these results also demonstrate the benefits of computational methods for the automatic annotation of narrative texts. Such methods can be leveraged to tease out distinct sub-processes of complex cognitive processes, like language understanding, using experimental setups with greater ecological validity. As such they pave the way for increasing synergy between computational linguistics and the cognitive neuroscience of language.

Notes

1. While the experience of listening to an audiobook is quite a natural one e.g. for daily commuters, it is not quite *conversational*. We put aside the question of which form of language is most basic, contenting ourselves with a comparative sense of the term "naturalistic." The implicit comparison is between story-listening and lists of unrelated sentence stimuli, devoid of literary content.
2. PMI is one of several cohesion metrics that have been advanced in both computational linguistics and corpus linguistics; Evert (2008) offers a tutorial introduction.

Acknowledgements

The authors would like to thank Adam Mahar, Skyler Yeatman, Jaelyn Moore, Jacob Collard, Katey Huddleston, and Marissa Holl at Cornell University for assisting us with data collection.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This material is based upon work supported by the National Science Foundation under Grant Number 1607441 (USA). We also gratefully acknowledge support from the French National Research Agency (ANR) under grants ANR-14-CERA-0001 and ANR-16-NEUC-0005-01.

ORCID

Shohini Bhattasali http://orcid.org/0000-0002-6767-6529
Jonathan R. Brennan http://orcid.org/0000-0002-3639-350X

References

Abney, S. (2007). *Semisupervised learning for computational linguistics*. Boca Raton, FL: CRC Press.

Al Saied, H., Candito, M., & Constant, M. (2017). *The ATILF-LLF system for the PARSEME Shared ask: A transition-based verbal multiword expression tagger*. Proceedings of the 13th workshop on Multiword Expressions (MWE 2017), Valencia, Spain (pp. 127–132). Association for Computational Linguistics. Retrieved from https://github.com/hazemalsaied/ATILF-LLF-MWE-Analyser

Andreasen, N. C., O'Leary, D. S., Cizadlo, T., Arndt, S., Rezai, K., Watkins, L., … Hichwa, R. (1995). Remembering the past: Two facets of episodic memory explored with positron emission tomography. *The American Journal of Psychiatry*, *152*(11), 1576–1585.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, *1316*(1), 29–52.

Bach, E. (1976). An extension of classical transformational grammar. In *Problems in linguistic metatheory: Proceedings of the 1976 conference at Michigan State University* (pp. 183–224). Lansing: Michigan State University.

Barker, C., & Jacobson, P. (2007). Introduction: Direct compositionality. In C. Barker & P. Jacobson (Eds.), *Direct compositionality* (pp. 1–22). Oxford: Oxford University Press.

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

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

Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). Neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *NeuroImage*, *21*(4), 1320–1336.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York, NY: Wiley.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, *3*(1), 19–58.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: A new dorsal–ventral stream model of sentence comprehension. *Brain and Language*, *125*(1), 60–76.

Brennan, J. R. (2016). Naturalistic sentence comprehension in the brain. *Language and Linguistics Compass*, *10*(7), 299–313.

Brennan, J. R., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, *120*(2), 163–173.

Brennan, J. R., & Pylkkänen, L. (2017). MEG evidence for incremental sentence composition in the anterior temporal lobe. *Cognitive Science*, *41*(S6), 1515–1531.

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, *157–158*, 81–94.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Curtis, V. A., Morris, R. G., Williams, S. C. R., … McGuire, P. K. (1999). Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. *Human Brain Mapping*, *7*(1), 38–48.

Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one? *The Mental Lexicon*, *9*(2), 267–293.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*(6), 668–683.

Calzolari, N., Fillmore, C. J., Grishman, R., Ide, N., Lenci, A., MacLeod, C., & Zampolli, A. (2002). *Towards best practice for multiword expressions in computational lexicons*. Proceedings of the 3rd international conference on Language Resources and Evaluation (LREC 2002), Las Palmas, Canary Islands (pp. 1934–1940). Retrieved from http://www.lrec-conf.org/proceedings/lrec2002/pdf/259.pdf

Christensen, K. R. (2008). Interfaces, syntactic movement, and neural activation: A new perspective on the implementation of language in the brain. *Journal of Neurolinguistics*, *21*(2), 73–103.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29.

Constant, M., & Nivre, J. (2016). *A transition-based system for joint lexical and syntactic analysis*. Proceedings of the 54th annual meeting of the Association for Computational Linguistics (Long papers) (Vol. 1, pp. 161–171), Berlin.

Constant, M., & Sigogne, A. (2011). *MWU-aware part-of-speech tagging with a CRF model and lexical resources*. Proceedings of the workshop on Multiword Expressions: From Parsing and Generation to the Real World (pp. 49–56). Association for Computational Linguistics. Retrieved from http://igm.univ-mlv.fr/~mconstan/index.php?section=software

Constant, M., & Tellier, I. (2012). *Evaluating the impact of external lexical resources into a CRF-based multiword segmenter and part-of-speech tagger*. 8th international conference on Language Resources and Evaluation (LREC'12), Istanbul (pp. 646–650). Retrieved from http://www.lrec-conf.org/lrec2012/

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 560 million words, 1990–present*. Salt Lake City, UT: Brigham Young University.

den Ouden, D.-B., Fix, S., Parrish, T. B., & Thompson, C. K. (2009). Argument structure effects in action verb naming in static and dynamic conditions. *Journal of Neurolinguistics*, *22*(2), 196–215.

de Saint-Exupéry, A. (1943). *Le petit prince*. New York, NY: Harcourt, Brace and World.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1), 145–177.

Ellis, N. C. (2008). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Malden, MA: Blackwell Publishing.

Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 1212–1248). Berlin: W. de Gruyter.

Fabre, M. (2017). *The sentence as cognitive object – The neural underpinnings of syntactic complexity in Chinese and French* (Unpublished doctoral dissertation). INALCO Paris.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*(21), 2059–2062.

Fedorenko, E., Scott, T. L., Brunner, P., Coon, W. G., Pritchett, B., Schalk, G., & Kanwisher, N. (2016). Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences*, *113*(41), E6256–E6262.

Ferstl, E. C., Neumann, J., Bogler, C., & Von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*(5), 581–593.

Fletcher, P., Frith, C., Baker, S., Shallice, T., Frackowiak, R., & Dolan, R. (1995). The mind's eye-precuneus activation in memory-related imagery. *NeuroImage*, *2*(3), 195–200.

Frankland, S. M., & Greene, J. D. (2015). An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proceedings of the National Academy of Sciences*, *112* (37), 11732–11737.

Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 129–187). Cambridge: Cambridge University Press. Retrieved from http://www.cambridge.org/US/academic/subjects/languages-linguistics/computational-linguistics/natural-language-parsing-psychological-computational-and-theoretical-perspectives?format=PB

Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250–254.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., & Penny, W. (Eds.). (2007). *Statistical parametric mapping: The analysis of functional brain images*. London: Academic Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423.

Hagoort, P. (2009). Reflections on the neurobiology of syntax. In D. Bickerton & E. Szathmáry (Eds.), *Biological foundations and origin of syntax* (pp. 279–296). Cambridge, MA: MIT Press. Retrieved from https://mitpress.mit.edu/books/biological-foundations-and-origin-syntax

Hagoort, P. (2016). MUC (Memory, Unification, Control): A model on the neurobiology of language beyond single word processing. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (Chap. 28, pp. 339–347). London: Elsevier.

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, *37*, 347–362.

Hale, J. T. (2014). *Automaton theories of human sentence comprehension*. Stanford, CA: CSLI Publications.

Halsband, U., Krause, B. J., Sipilä, H., Teräs, M., & Laihinen, A. (2002). PET studies on the memory processing of word pairs in bilingual Finnish–English subjects. *Behavioural Brain Research*, *132*(1), 47–57.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.

Hill, F., Bordes, A., Chopra, S., & Weston, J. (2015). *The Goldilocks principle: Reading children's books with explicit memory representations*. arXiv preprint arXiv:1511.02301.

Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., & Evans, A. C. (1998). Enhancement of mr images using registration for signal averaging. *Journal of Computer Assisted Tomography*, *22*(2), 324–333.

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.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press. Retrieved from https://global.oup.com/academic/product/foundations-of-language-9780199264377

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790.

Klein, D., & Manning, C. D. (2003). *Accurate unlexicalized parsing*. Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, Sapporo, Japan (pp. 423–430).

Klein, E., & Sag, I. A. (1985). Type-driven translation. *Linguistics and Philosophy*, *8*(2), 163–201.

Kobele, G. M. (2006). *Generating copies: An investigation into structural identity in language and grammar* (Unpublished doctoral dissertation). UCLA.

Kuiper, K., McCann, H., Quinn, H., Aitchison, T., & van der Veer, K. (2003). *Syntactically Annotated Idiom Dataset (SAID)* (LDC2003T10). Philadelphia, PA: Linguistic Data Consortium.

Kundu, P., Brenowitz, N. D., Voon, V., Worbe, Y., Vértes, P. E., Inati, S. J., … Bullmore, E. T. (2013). Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proceedings of the National Academy of Sciences*, *110*(40), 16187–16192.

Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage*, *60*(3), 1759–1770.

Kundu, P., Voon, V., Balchandani, P., Lombardo, M. V., Poser, B. A., & Bandettini, P. A. (2017). Multi-echo fMRI: A review of

applications in fMRI denoising and analysis of BOLD signals. *NeuroImage*, *154*, 59–80.
Laird, J., Rosenbloom, P., & Newell, A. (1986). Chunking in Soar, anatomy of a general learning mechanism. *Machine Learning*, *1*, 11–46.
Laurent, J.-P., Denhières, G., Passerieux, C., Iakimova, G., & Hardy-Baylé, M.-C. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, *1068*(1), 151–160.
Leggio, M. G., Tedesco, A. M., Chiricozzi, F. R., Clausi, S., Orsini, A., & Molinari, M. (2008). Cognitive sequencing impairment in patients with focal or atrophic cerebellar damage. *Brain*, *131*(5), 1332–1343.
Li, J., & Hale, J. (2018). Grammatical predictors for fMRI imecourses. In R. C. Berwick & E. P. Stabler Jr. (Eds.), *Minimalist parsing*. Oxford: Oxford University Press.
Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W.-L., & Nichols, T. E. (2006). Non-white noise in fMRI: Does modelling have an impact? *NeuroImage*, *29*(1), 54–66.
Makkai, A., Boatner, M. T., & Gates, J. E. (1995). *A dictionary of American idioms*. Hauppage, NY: Barron's Educational Series.
Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
Marcus, M., Marcinkiewicz, M., & Santorini, B. (1993). Building a large annotated corpus of English: The penn treebank. *Computational Linguistics*, *19*(2), 313–330.
Mashal, N., Vishne, T., & Laor, N. (2014). The role of the precuneus in metaphor comprehension: Evidence from an fMRI study in people with schizophrenia and healthy participants. *Frontiers in Human Neuroscience*, *8*, 818.
Mason, R. A., & Just, M. A. (2006). Neuroimaging contributions to the understanding of discourse processes. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 765–799). Amsterdam: Elsevier.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.
Molinari, M., Chiricozzi, F. R., Clausi, S., Tedesco, A. M., De Lisa, M., & Leggio, M. G. (2008). Cerebellum and detection of sequences, from perception to cognition. *The Cerebellum*, *7* (4), 611–615.
Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes*, *28*(6), 762–788.
Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, *83*(3), 176–190.
Molinaro, N., Carreiras, M., & Duñabeitia, J. A. (2012). Semantic combinatorial processing of non-anomalous expressions. *NeuroImage*, *59*(4), 3488–3501.
Molinaro, N., Vespignani, F., Canal, P., Fonda, S., & Cacciari, C. (2008). Cloze probability does not only affect N400 amplitude: The case of complex prepositions. *Psychophysiology*, *45*(6), 1008–1012.
Montague, R. (1974). *Formal philosophy: Selected paper of Richard Montague* (Edited and with an introduction by R. H. Thomason). New Haven, CT: Yale University Press.
Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.
Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., … Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, *114*(18), E3669–E3678.
Nieuwland, M. S., Martin, A. E., & Carreiras, M. (2012). Brain regions that process case: Evidence from Basque. *Human Brain Mapping*, *33*(11), 2509–2520.
O'Donnell, T. J. (2015). *Productivity and reuse in language: A theory of linguistic computation and storage*. Cambridge, MA: MIT Press.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.
Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, *108*(6), 2522–2527.
Paumier, S., Nakamura, T., & Voyatzi, S. (2009). Unitex, a corpus processing system with multi-lingual linguistic resources. *eLEX2009*, 173.
Peirce, J. W. (2007). Psychopy: Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.
Platel, H., Price, C., Baron, J.-C., Wise, R., Lambert, J., Frackowiak, R. S., … Eustache, F. (1997). The structural components of music perception. A functional anatomical study. *Brain*, *120* (2), 229–243.
Pylkkänen, L. (2008). Mismatching meanings in brain and behavior. *Language and Linguistics Compass*, *2*(4), 712–738.
Pylkkänen, L. (2016). Composition of complex meaning: Interdisciplinary perspectives on the left anterior temporal lobe. In G. Hickok & S. Small (Eds.), *Neurobiology of language*. London: Academic Press.
Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: At the crossroads of semantic and episodic memory. *Trends in Cognitive Sciences*, *16*(11), 550–558.
Rogalsky, C., & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex*, *19*(4), 786–796.
Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762–776.
Rosenbloom, P. S., & Newell, A. (1987). Learning by chunking: A production-system model of practice. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development* (pp. 221–286). Cambridge, MA: MIT Press.
Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). *Multiword expressions: A pain in the neck for NLP*. International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2002), Mexico City, Mexico (pp. 1–15).
Santi, A., & Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *NeuroImage*, *51*(4), 1285–1293.
Shetreet, E., & Friedmann, N. (2014). The processing of different syntactic structures: fMRI investigation of the linguistic

distinction between wh-movement and verb movement. *Journal of Neurolinguistics*, *27*(1), 1–17.
Shetreet, E., Friedmann, N., & Hadar, U. (2009). An fMRI study of syntactic layers: Sentential and lexical aspects of embedding. *NeuroImage*, *48*(4), 707–716.
Shetreet, E., Friedmann, N., & Hadar, U. (2010). Cortical representation of verbs with optional complements: The theoretical contribution of fMRI. *Human Brain Mapping*, *31*(5), 770–785.
Shetreet, E., Palti, D., Friedmann, N., & Hadar, U. (2007). Cortical representation of verb processing in sentence comprehension: Number of complements, subcategorization, and thematic frames. *Cerebral Cortex*, *17*(8), 1958–1969.
Shieber, S., & Johnson, M. (1993). Variations on incremental interpretation. *Journal of Psycholinguistic Research*, *22*(2), 287–318.
Silveira, N., Dozat, T., Schuster, S., Connor, M., de Marneffe, M.-C., Schneider, N., … Bauer, J. (2015). *May, Universal dependencies English web treebank*. Retrieved from http://universaldependencies.org/treebanks/en_ewt/index.html
Siyanova-Chanturia, A. (2013). Eye-tracking and ERPs in multiword expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon*, *8*(2), 245–268.
Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*(2), 251–272.
Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. B. (2011). Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 776–784.
Siyanova-Chanturia, A., & Martinez, R. (2014). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549–569.
Smith, N. (2011). *Linguistic structure prediction*. San Rafael, CA: Morgan & Claypool.
Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., … Flitney, D. E. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219.
Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, *19*(7), 1493–1503.
Stabler, E. (1991). Avoid the pedestrian's paradox. In R. C. Berwick, S. P. Abney, and C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics* (pp. 199–237). Dordrecht: Kluwer.
Steedman, M. (2000). *The syntactic process*. Cambridge, MA: MIT Press.
Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *NeuroImage*, *44*(2), 489–501.
Stowe, L. A., Broere, C. A. J., Paans, A. M. J., Wijers, A. A., Mulder, G., Vaalburg, W., & Zwarts, F. (1998). Localizing components of a complex task: Sentence processing and working memory. *Neuroreport*, *9*(13), 2995–2999.
Strandburg, R. J., Marsh, J. T., Brown, W. S., Asarnow, R. F., Guthrie, D., & Higa, J. (1993). Event-related potentials in high-functioning adult autistics: Linguistic and nonlinguistic visual information processing tasks. *Neuropsychologia*, *31*(5), 413–434.
Thompson, C. K., Bonakdarpour, B., Fix, S. C., Blumenfeld, H. K., Parrish, T. B., Gitelman, D. R., & Mesulam, M.-M. (2007). Neural correlates of verb argument structure processing. *Journal of Cognitive Neuroscience*, *19*(11), 1753–1767.
Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London: The Continuum International Publishing Group.
Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569–613.
Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*(10), 717–726.
Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1–2), 231–270.
Ullman, M. T. (2015). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (Chap. 76, pp. 953–968). London: Elsevier.
Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it. *Formulaic Sequences: Acquisition, Processing, and Use*, *9*, 153.
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.
Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*(8), 1682–1700.
Wallentin, M., Weed, E., Østergaard, L., Mouridsen, K., & Roepstorff, A. (2008). Accessing the mental space-spatial working memory processes for language and vision overlap in precuneus. *Human Brain Mapping*, *29*(5), 524–532.
Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A., & Mitchell, T. (2014). Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. *PLoS One*, *9*(11), e112575.
White, J. G. (1998). *Cambridge international dictionary of idioms*. New York, NY: Cambridge University Press.
Willems, R. M. (Ed.). (2015). *Cognitive neuroscience of natural language use*. Cambridge: Cambridge University Press.
Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2016). Prediction during natural language comprehension. *Cerebral Cortex*, *26*, 2506–2516.
Wilson, S. M., DeMarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Mandelli, M. L., … Gorno-Tempini, M. L. (2014a). What role does the anterior temporal lobe play in sentence-level processing? Neural correlates of syntactic processing in semantic variant primary progressive aphasia. *Journal of Cognitive Neuroscience*, *26*(5), 970–985.
Wilson, S. M., DeMarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Mandelli, M. L., … Gorno-Tempini, M. L. (2014b). What role does the anterior temporal lobe play in sentence-level processing? Neural correlates of syntactic processing in semantic variant primary progressive aphasia. *Journal of Cognitive Neuroscience*, *26*(5), 970–985.

Wilson, S. M., DeMarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Miller, B. L., & Gorno-Tempini, M. L. (2016). Variable disruption of a syntactic processing network in primary progressive aphasia. *Brain*, *139*(11), 2994–3006.

Wilson, S. M., Galantucci, S., Tartaglia, M. C., & Gorno-Tempini, M. L. (2012). The neural basis of syntactic deficits in primary progressive aphasia. *Brain and Language*, *122*(3), 190–198.

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

Yaneva, V., Taslimipoor, S., Rohanian, O., & Ha, L. A. (2017). Cognitive processing of multiword expressions in native and non-native speakers of English: Evidence from gaze data. In R. Mitkov (Ed.), *International conference on computational and corpus-based phraseology* (pp. 363–379). Cham: Springer.

Yang, Y.-H., Marslen-Wilson, W. D., & Bozic, M. (2017). Syntactic complexity and frequency in the neurocognitive language system. *Journal of Cognitive Neuroscience*, *29*(9), 1605–1620.

Zaccarella, E., & Friederici, A. D. (2015). Merge in the human brain: A sub-region based functional investigation in the left Pars Opercularis. *Frontiers in Psychology*, *6*, 1818.

Zaccarella, E., & Friederici, A. D. (2017). The neurobiological nature of syntactic hierarchies. *Neuroscience and Biobehavioral Reviews*, *81*, 205–212.