Proceedings of the Society for Computation in Linguistics

Volume 1 Article 9

2018

Differentiating Phrase Structure Parsing and Memory Retrieval in the Brain

Shohini Bhattasali
Cornell University, sb2295@cornell.edu

John Hale
Cornell University, jthale@cornell.edu

Christophe Pallier
INSERM-CEA, christophe@pallier.org

Jonathan Brennan
University of Michigan - Ann Arbor, jobrenn@umich.edu

Wen-Ming Luh
Cornell University, wl358@cornell.edu

See next page for additional authors

Differentiating Phrase Structure Parsing and Memory Retrieval in the Brain

Shohini Bhattasali
Cornell University
Ithaca, NY, USA
sb2295@cornell.edu

John Hale
Cornell University
Ithaca, NY, USA
jthale@cornell.edu

Christophe Pallier
INSERM-CEA
Paris-Saclay, France
christophe@pallier.org

Jonathan R. Brennan
University of Michigan
Ann Arbour, MI, USA
jobrenn@umich.edu

Wen-Ming Luh
Cornell University
Ithaca, NY, USA
wl358@cornell.edu

R. Nathan Spreng
McGill University
Montreal, Canada
nathan.spreng@mcgill.ca

Abstract

On some level, human sentence comprehension must involve both memory retrieval and structural composition. This study differentiates these two processes using neuroimaging data collected during naturalistic listening. Retrieval is formalized in terms of "multiword expressions" while structure-building is formalized in terms of bottom-up parsing. The results most strongly implicate Anterior Temporal regions for structure-building and Precuneus Cortex for memory retrieval.

1 Introduction

This study differentiates processes of structure-building and memory retrieval in the brain, as they occur during naturalistic language comprehension. We use multiword expressions to investigate this distinction. The term itself comes from computational linguistics; roughly it means expressions that are better treated non-compositionally (Sag et al., 2002). Figure 2 on page 2 highlights several examples.

MWEs raise an important theoretical question about language processing, namely the balance between productivity and reuse (Goldberg, 2006; Jackendoff, 2002; O'Donnell, 2015). If MWEs indeed lack internal structure, then perhaps their comprehension proceeds through a single, unitary memory retrieval operation, rather than some kind of multi-step composition process. Proceeding from this hypothesis, the paper contributes a localization of these two cognitive processes in the brain through an analysis of fMRI timecourses collected during naturalistic listening.

2 Memory Retrieval vs. Structure-building

The name MWE loosely groups a wide variety of linguistic phenomena including idioms, perfunctory greetings and personal titles.

(1) *When I drew the baobabs, I was spurred on by a **sense of urgency***

(2) *"**Good morning**", said the **little prince** politely, who then turned around, but saw nothing.*

The syntactic or semantic properties of the bold-faced expressions cannot be derived just from their parts and in some way, they are conventionalized. They are plausibly stored, rather than built on the fly (Cacciari, 2014).

By contrast, other expressions are less likely to have been explicitly memorized and therefore call for some degree of structural composition, in comprehension. This sort of processing can be formalized using parsing algorithms (Hale, 2014). Figure 1 on page 2 indicates the number of reduce steps that a bottom-up parser would take, word-by-word, as it builds the depicted phrase structure. Our analysis of the neuroimaging data described in the next section takes this number as an index of structure-building effort.

3 fMRI Study

3.1 Method

We follow Brennan et al. (2012) in using a spoken narrative as a stimulus. Participants hear the story over headphones while they are in the scanner. The

Proceedings of the Society for Computation in Linguistics (SCiL) 2018, pages 74-80.
Salt Lake City, Utah, January 4-7, 2018

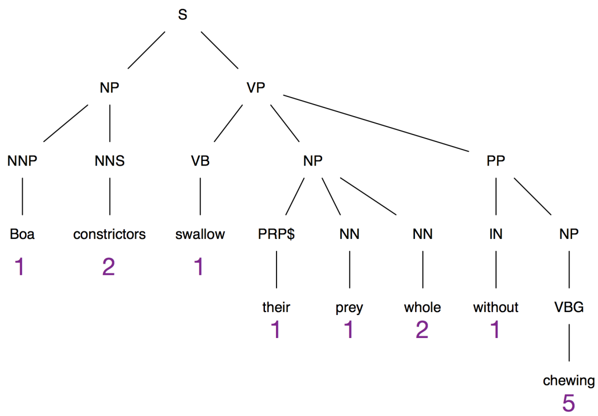

Figure 1: Phrase structure tree with bottom-up parser action counts in purple. For more on parsing algorithms see Hale (2014).

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.

3.2 Stimuli

The audio stimulus was Antoine de Saint-Exupéry's *The Little Prince*, translated by David Wilkinson and read by Nadine Eckert-Boulet.

Within this text, 1,274 MWEs were identified using a CRF tagger. This tagger was trained on examples from the English Universal Dependency treebank, in combination with external lexicons as suggested by Constant and Tellier (2012). The tagger used feature templates, as seen in Table 1 below, where w_t stands for the token at the relative position t from the current token and l_t is the label at the relative position t. The external lexicons included the Unitex lexicon (Paumier et al., 2009), SAID corpus (Kuiper et al., 2003), Cambridge International Dictionary of Idioms (White, 1998), and Dictionary of American Idioms (Makkai et al., 1995).

Among these MWEs, attestation rates for particular subtypes are given in Table 2.

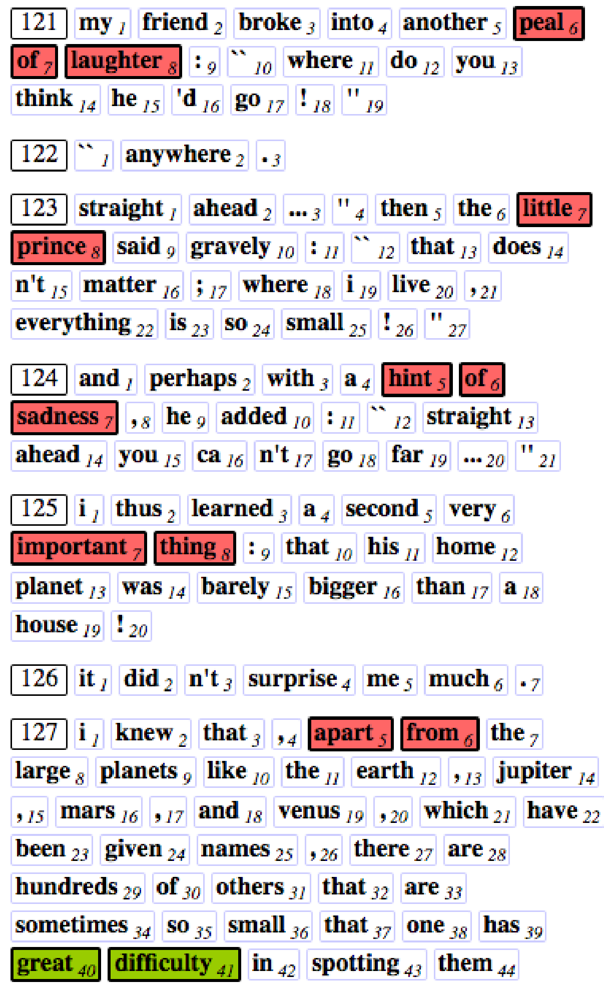

Figure 2: Samples MWEs in the English text, visualized with mwetoolkit (Ramisch et al., 2010)

$w_t = X, t \in \{2, 1, 0, 1, 2\}$	$\&l_0 = L$
Lowercase form of $w_0 = W$	$\&l_0 = L$
Prefix of $w_0 = Pwith\|P\|<5$	$\&l_0 = L$
Suffix of $w_0 = Swith\|S\|<5$	$\&l_0 = L$
w_0 contains a hyphen	$\&l_0 = L$
w_0 contains a digit	$\&l_0 = L$
w_0 is capitalized	$\&l_0 = L$
w_0 is all in capital	$\&l_0 = L$
w_0 is capitalized and BOS	$\&l_0 = L$
w_0 is part of a multiword	$\&l_0 = L$
$w_i w_j = XY, (j,k) \in \{(1,0),(0,1),(1,1)\}$	$\&l_0 = L$
$l_{-1} = L'$	$\&l_0 = L$

Table 1: Feature templates to detect MWEs

3.3 Participants

Participants were forty-two volunteers (26 women and 16 men, 18-37 years old) with no history of

MWE Category	Occurrence
Verb + Participle	145
Verb + Noun	37
Adj + Noun	285
Det + Noun	712
(Verb) + Noun + Prep + Noun	24
N-N Compounds	71

Table 2: MWE Attestation Rates

psychiatric, neurological, or other medical illness or history of drug or alcohol abuse that might compromise cognitive functions. All 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.

3.4 Presentation

After giving their informed consent, participants were familiarized 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 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 had nine chapters and at the end of each chapter 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 were used to confirm their comprehension and were viewed by the participants via a mirror attached to the head coil and they answered through a button box. The entire session lasted around 2.5 hours.

3.5 Data Collection

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

4 Data Analysis

4.1 Preprocessing

fMRI data is acquired with physical, biological constraints and preprocessing allows us to make adjustments to improve the signal to noise ratio. Primary preprocessing steps were carried out in AFNI version 16 (Cox, 1996) and include motion correction, coregistration, and normalization to standard MNI space. After the previous steps were completed, ME-ICA (Kundu et al., 2012) was used to further preprocess the data. ME-ICA is a denoising method which uses Independent Components Analysis to split the T2*-signal into BOLD and non-BOLD components. Removing the non-BOLD components mitigates noise due to motion, physiology, and scanner artifacts (Kundu et al., 2017).

4.2 Statistical Analysis

The GLM typically used in fMRI is a hierarchical model with two levels (see Poldrack et al., 2011). At the first level, the data for each subject is modelled separately to calculate subject-specific parameter estimates and within-subject variance such that for each subject, a regression model is estimated for each voxel against the time series. The second-level model takes subject-specific parameter estimates as input. It uses the between-subject variance to make statistical inferences about the larger population.

The GLM analysis was performed using SPM12 (Penny et al., 2011). The following regressors were used. One regressor formalizes structure-building using a standard bottom-up parsing algorithm (see chapter 3 of Hale, 2014). We computed the number of parser actions that would be required, word-by-word, to build the correct phrase structure tree as determined by the Stanford parser (Klein and Manning, 2003). Another regressor formalizes memory retrieval, by marking multiword expressions (MWE; see section 3.2). Each word in the text was annotated with a 0 or 1, depending on whether it was the last word of a MWE. This coding scheme expresses the idea that a different process occurs at the end of multiword expressions, and this provisionally assumes a very conservative approach to the Configuration Hypothesis (Cacciari and Tabossi, 1988; Tabossi et al., 2009). We regressed the word-by-word predictors described above against fMRI timecourses recorded during passive story-listening in a whole-brain analysis. Along with the parsing and MWE regressors of theoretical interest, we entered four "nuisance" variables or regressors of non-interest into the GLM analysis using SPM12. One regressor simply marks the offset of each spoken word in time. Another gives the log-frequency of the individual word in movie subtitles (Brysbaert and New, 2009). The last two reflect the pitch (f0) and intensity (RMS) of the talker's voice. These nuisance regressors are added to the GLM analysis to improve sensitivity, specificity and validity of activation maps (Bullmore et al., 1999; Lund et al., 2006). In particular, we sought to ensure that any conclusions about parsing and memory retrieval would be specific to those processes, as opposed to more general aspects of speech perception.

5 Results

In the second-level group analysis, bottom-up parsing and multi-word expressions were analyzed separately. Results are presented below in Tables 3 and 4 using region names from the Harvard-Oxford Cortical Structure Atlas.

5.1 Group level results for bottom-up parsing

MNI Coordinates x	y	z	Region	p-value (corrected)	k-size (cluster)	T-score (peak-level)
52	8	-22	Temporal pole	0.000	2769	12.72
54	-40	12	Supramarginal Gyrus	0.000	2212	12.69
-34	18	-12	Frontal Orbital Cortex	0.000	2380	10.40
12	20	58	Superior Frontal Gyrus	0.000	7191	9.27
42	2	48	Middle Frontal Gyrus	0.000	286	9.19
-38	26	36	Middle Frontal Gyrus	0.000	382	8.47
-40	-78	6	Lateral Occipital Cortex	0.000	693	7.42
-52	-56	32	Angular Gyrus	0.000	802	7.12
28	-52	-8	Temporal Occipital Fusiform Cortex	0.001	83	6.74
-44	46	-12	Frontal Pole	0.000	176	6.28

Table 3: Significant clusters for bottom-up parser action count after FWE voxel correction.

The largest clusters ($p < 0.05$ FWE) were observed in Anterior Temporal regions (Temporal Pole) and Frontal regions.

Figure 3, plotted with nilearn (Abraham et al., 2014) is the T-score map for the bottom-up parser action count regressor. This regressor formalizes processing effort related to structural composition.

5.2 Group level results for MWE

MNI Coordinates x	y	z	Region	p-value (corrected)	k-size (cluster)	T-score (peak-level)
6	-60	52	Precuneus Cortex	0.000	559	7.62
24	10	56	Superior Frontal Gyrus	0.000	182	7.23
-40	42	26	Frontal Pole	0.000	158	6.94
66	-38	34	Supramarginal Gyrus	0.000	103	6.77
34	38	36	Frontal Pole	0.001	58	5.83

Table 4: Significant clusters for MWEs after FWE voxel correction

The largest clusters ($p < 0.05$ FWE) were observed in Precuneus Cortex and Frontal regions.

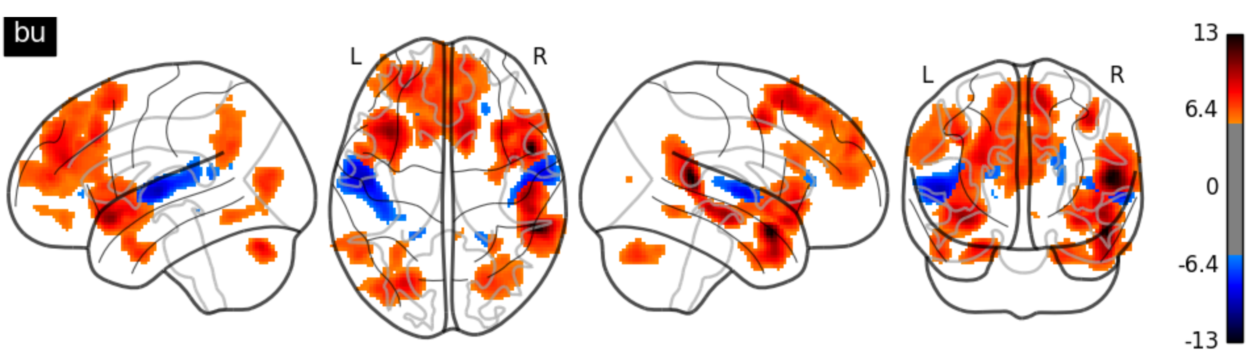

Figure 3: T-score map for the Bottom-up Parser action count regressor. Red represents the positive score while blue represents the negative score.

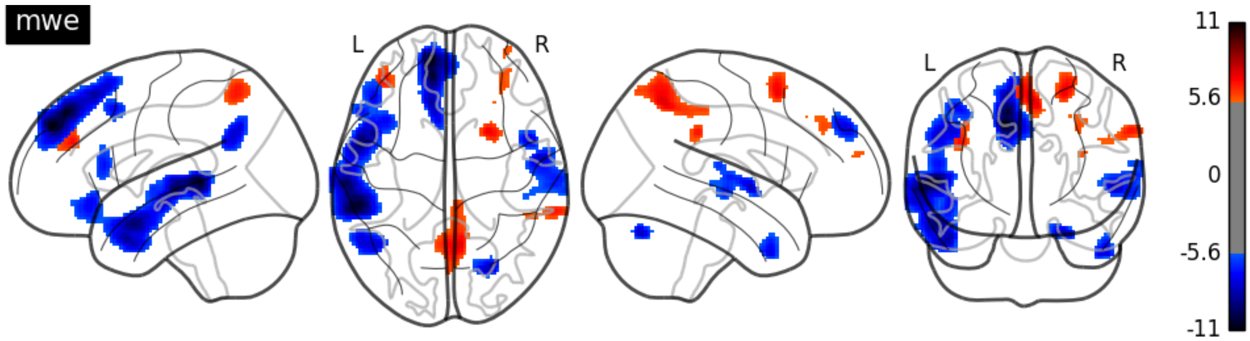

Figure 4: T-score map for the MWE status regressor. Red represents the positive score while blue represents the negative score.

Figure 4 (also plotted with nilearn) is the T-score map for the MWE status regressor, which is meant to formalize the retrieval of these noncompositional expressions.

6 Discussion

The operationalization of structure-building as bottom-up parsing highlights Anterior Temporal as well as Frontal regions. These results are consistent with earlier work including deficit-lesion data (Dronkers et al., 2004), fMRI studies of text comprehension (Ferstl et al., 2008), and magnetoencephalography studies of phrasal composition (Bemis and Pylkkänen, 2011). This literature also confirms the sensitivity of Anterior Temporal regions to parametric variation of phrase size (Pallier et al., 2011).

With respect to MWEs, significant activation was observed in the Precuneus. The Precuneus has not traditionally been viewed as part of the language network. However, it has been designated as part of the Protagonist's Perspective Interpreter Network (Mason and Just, 2006). This network in fact appears to be activated by many different sorts of story characters, not just the protagonist (Wehbe et al., 2014). Along these lines, the Precuneus activation in Figure 4 might be interpreted narrowly as an effect of reference to dramatis personae in the narrative stimulus. This restricted interpretation would be challenged by the fact that less than 25% of the MWEs in the stimulus text are references to story characters. To account for the full collection, including verbal MWEs, a more general characterization in terms of memory retrieval seems appropriate.

This more general characterization is bolstered by data that implicate the Precuneus in memory tasks:

- Verbal memory (Halsband et al., 2002)
- Spatial memory (Wallentin et al., 2008)
- Episodic memory (Andreasen et al., 1995)
- Memory-related imagery (Fletcher et al., 1995; Mashal et al., 2014)

As Spreng et al. (2009) suggest, the Precuneus could be part of a wider, task-general network that is also recruited in Theory-of-Mind, Prospection and Autobiographical memory tasks. These considerations strengthen the interpretation, based on MWEs, that the Precuneus mediates memory retrieval during naturalistic language comprehension.

The findings as a whole are broadly consistent with existing neurocognitive models of language. For example, within Hagoort's (2016) MUC model, MWE comprehension might tap memory resources, whereas bottom-up parsing might involve unification. With MUC, our analysis suggests a localization of these memory resources to the Parietal lobe. Within the Procedural/Declarative model (Ullman, 2001; 2004), rule-based linguistic knowledge would be localized to Frontal regions. This is consistent with the Frontal activations that we observe in response to the bottom-up parsing regressor.

7 Conclusion

These results point to a spatial differentiation between reuse and composition in language comprehension. Reuse, here operationalized with multiword expressions, seems to involve the Precuneus. Composition, in the sense of phrase-structure parsing, seems to call upon Anterior Temporal areas.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1607441.

References

Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. 2014. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8.

Nancy C Andreasen, Daniel SO Leary, Ted Cizadlo, Stephan Arndt, et al. 1995. Remembering the past: two facets of episodic memory explored with positron emission tomography. *The American journal of psychiatry*, 152(11):1576.

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

Jonathan Brennan, Yuval Nir, Uri Hasson, Rafael Malach, David J Heeger, and Liina Pylkkänen. 2012. Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and language*, 120(2):163–173.

Marc Brysbaert and Boris New. 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.

ET Bullmore, MJ Brammer, S Rabe-Hesketh, VA Curtis, RG Morris, SCR Williams, T Sharma, and PK McGuire. 1999. Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fmri. *Human brain mapping*, 7(1):38–48.

Cristina Cacciari and Patrizia Tabossi. 1988. The comprehension of idioms. *Journal of memory and language*, 27(6):668–683.

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

Matthieu Constant and Isabelle Tellier. 2012. Evaluating the impact of external lexical resources into a crf-based multiword segmenter and part-of-speech tagger. In *8th International Conference on Language Resources and Evaluation (LREC'12)*, pages 646–650.

Robert W. Cox. 1996. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173.

Nina F Dronkers, David P Wilkins, Robert D Van Valin, Brenda B Redfern, and Jeri J Jaeger. 2004. Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92(1):145–177.

Evelyn C Ferstl, Jane Neumann, Carsten Bogler, and D Yves Von Cramon. 2008. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Human brain mapping*, 29(5):581–593.

PC Fletcher, CD Frith, SC Baker, T Shallice, RSJ Frackowiak, and RJ Dolan. 1995. The mind's eyeprecuneus activation in memory-related imagery. *Neuroimage*, 2(3):195–200.

Adele E. Goldberg. 2006. *Constructions at work: The nature of generalization in language*. Oxford University Press.

Peter Hagoort. 2016. MUC (memory, unification, control): A model on the neurobiology of language beyond single word processing. In *Neurobiology of language*, pages 339–347. Elsever.

John T Hale. 2014. *Automaton theories of human sentence comprehension*. CSLI Publications.

U Halsband, BJ Krause, H Sipilä, M Teräs, and A Laihinen. 2002. Pet studies on the memory processing of word pairs in bilingual finnish–english subjects. *Behavioural brain research*, 132(1):47–57.

Ray Jackendoff. 2002. Foundation of language: Brain, meaning, grammar. *Evolution*.

Dan Klein and Christopher D Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 423–430. Association for Computational Linguistics.

Koenraad Kuiper, Heather McCann, Heidi Quinn, Therese Aitchison, and Kees van der Veer. 2003. Syntactically Annotated Idiom Dataset (SAID) LDC2003T10. In *Linguistic Data Consortium*, Philadelphia.

Prantik Kundu, Souheil J Inati, Jennifer W Evans, Wen-Ming Luh, and Peter A Bandettini. 2012. Differentiating bold and non-bold signals in fmri time series using multi-echo epi. *Neuroimage*, 60(3):1759–1770.

Prantik Kundu, Valerie Voon, Priti Balchandani, Michael V. Lombardo, Benedikt A. Poser, and Peter A. Bandettini. 2017. Multi-echo fmri: A review of applications in fmri denoising and analysis of bold signals. *NeuroImage*, 154:59 – 80. Cleaning up the fMRI time series: Mitigating noise with advanced acquisition and correction strategies.

Torben E Lund, Kristoffer H Madsen, Karam Sidaros, Wen-Lin Luo, and Thomas E Nichols. 2006. Non-white noise in fmri: does modelling have an impact? *Neuroimage*, 29(1):54–66.

Adam Makkai, M. T. Boatner, and J. E. Gates. 1995. *A Dictionary of American idioms*. ERIC.

Nira Mashal, Tali Vishne, and Nathaniel Laor. 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.

Robert A Mason and Marcel Adam Just. 2006. Neuroimaging contributions to the understanding of discourse processes. *Handbook of psycholinguistics*, 799.

Timothy J O'Donnell. 2015. *Productivity and reuse in language: A theory of linguistic computation and storage*. MIT Press.

Richard C Oldfield. 1971. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113.

Christophe Pallier, Anne-Dominique Devauchelle, and Stanislas Dehaene. 2011. Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, 108(6):2522–2527.

Sébastien Paumier, Takuya Nakamura, and Stavroula Voyatzi. 2009. Unitex, a corpus processing system with multi-lingual linguistic resources. *eLEX2009*, page 173.

Jonathan W Peirce. 2007. Psychopypsychophysics software in python. *Journal of neuroscience methods*, 162(1):8–13.

William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols. 2011. *Statistical parametric mapping: the analysis of functional brain images*. Academic press.

Ivan A Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 1–15. Springer.

R Nathan Spreng, Raymond A Mar, and Alice SN Kim. 2009. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of cognitive neuroscience*, 21(3):489–510.

Patrizia Tabossi, Rachele Fanari, and Kinou Wolf. 2009. Why are idioms recognized fast? *Memory & Cognition*, 37(4):529–540.

Michael T Ullman. 2001. A neurocognitive perspective on language: The declarative/procedural model. *Nature reviews. Neuroscience*, 2(10):717.

Michael T Ullman. 2004. Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1):231–270.

Mikkel Wallentin, Ethan Weed, Leif Østergaard, Kim Mouridsen, and Andreas Roepstorff. 2008. Accessing the mental spacespatial working memory processes for language and vision overlap in precuneus. *Human Brain Mapping*, 29(5):524–532.

Leila Wehbe, Brian Murphy, Partha Talukdar, Alona Fyshe, Aaditya Ramdas, and Tom Mitchell. 2014. Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. *PloS one*, 9(11):e112575.

James Gordon White. 1998. *Cambridge International Dictionary of Idioms*. Cambridge University Press, New York.