Modeling fMRI time courses with linguistic structure at various grain sizes

John T. Hale and **David E. Lutz** and **Wen-Ming Luh**
Cornell University
Ithaca, NY 14853 USA
{jthale,del82,wenmingluh}@cornell.edu

Jonathan R. Brennan
The University of Michigan
Ann Arbor, MI 48109 USA
jobrenn@umich.edu

Abstract

Neuroimaging while participants listen to audiobooks provides a rich data source for theories of incremental parsing. We compare nested regression models of these data. These mixed-effects models incorporate linguistic predictors at various grain sizes ranging from part-of-speech bigrams, through surprisal on context-free treebank grammars, to incremental node counts in trees that are derived by Minimalist Grammars. The fine-grained structures make an independent contribution over and above coarser predictors. However, this result only obtains with time courses from anterior temporal lobe (aTL). In analogous time courses from inferior frontal gyrus, only n-grams improve upon a non-syntactic baseline. These results support the idea that aTL does combinatoric processing during naturalistic story comprehension, processing that bears a systematic relationship to linguistic structure.

1 Introduction

The cognitive science of language confronts two different notions of its own subject matter. One notion is rooted in the psychology of an individual: what states of mind does *this* person go through as he or she uses language? The other notion starts from languages themselves. As a structural system, how does *this* language differ from another? There is a tension between these two views. Classically, this tension is resolved by adopting the Competence Hypothesis (Chomsky, 1965, page 9). It suggests that the best description of the language system should also figure as a "basic component" in the best description of the language-user. This Hypothesis is programmatic enough to have received several different interpretations over the years (Bresnan and Kaplan, 1982; Steedman, 1989; Stabler, 1991). Can a refined version of it be accepted or rejected in light of experimental data?

Recent work with eye-tracking has wrestled with just this question (Frank and Bod, 2011; Fossum and Levy, 2012; van Schijndel and Schuler, 2015). The argument concerns the strength of the fitted coefficients for different types of grammatical predictors. These "language model" predictors contribute to varying degrees in regression models of the eye-fixation record. In certain cases, it appears that higher-order structure — for instance, phrase structure — is unhelpful. On the other hand, other cases suggest that higher-order structure does shine through in the eye-movement record. In this debate, fitted coefficients on the more linguistically-sophisticated predictors have been taken to quantify the veridicality of the Competence Hypothesis. The linguistic predictors that researchers examine qualify as "basic components" to the extent that they improve the regression model that they are part of.

Of course, psycholinguists have known for a long time that low-level factors such as word frequency and bigram probability are useful in explaining eye-fixation times (Thibadeau et al., 1982; McDonald and Shillcock, 2003). These are not relevant to the Competence Hypothesis. Rather, the action is with higher-order factors: predictors based on larger domains of locality, as defined by grammars that could plausibly play a role in the best descrip-

Proceedings of CMCL 2015, pages 89–97,
Denver, Colorado, June 4, 2015.

tion of language as a structural system.

The research reported in this paper adopts the same model-comparison methodology as Frank, Fossum, van Schijndel and their co-authors. But it applies this method to spatially localized neural time courses obtained using fMRI. Using grammatical predictors at six different levels of "richness" we compare a family of nested regression models. We find that phrase structure in the style of the Penn Treebank (Marcus et al., 1993) improves a regression, over and above various n-gram baselines. X-bar structures generated by Minimalist Grammars (Stabler, 1997, 2011) improve yet further over that. This holds for time courses taken from anterior temporal lobe (aTL), an area that has been implicated in "basic syntactic processing" (Friederici and Gierhan, 2013). But only the n-gram predictors are useful in modeling time courses from inferior frontal gyrus (IFG), a traditional syntax area (Grodzinsky and Friederici, 2006). Section 6 discusses this pattern of results in light of other work on naturalistic language comprehension.

2 Methods and Materials

The methodology follows Brennan et al. (2012) in the use of spoken narrative as a stimulus. Participants listen to an audiobook while in the scanner. The sequence of images collected during the spoken presentation becomes the dependent variable in a regression against a times series of linguistic predictors derived from the text of the story. In contrast to the work of Frank, Fossum and van Schijndel, we used auditory rather than visual presentation.

Participants

Thirteen college-age volunteers (6 women) participated for pay, but we excluded two individuals whose inferred head-movements exceeded 0.6mm or had eight or more movements $\geq$0.1mm. All qualified as right-handed on the Edinburgh handedness inventory (Oldfield, 1971). They self-identified as native English speakers and gave their informed consent.

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: 77 deg, 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.

Presentation

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.

Stimuli

The audio stimulus was Kristin McQuillan's reading of the first chapter of Lewis Carroll's Alice in Wonderland from `librivox.org`. We chose this text both because of its use in prior imaging work (Brennan et al., 2012) and because fine-grained syntactic annotations are available for it. We used Praat to normalize the spoken-language audio signal to 70dB and dilate it by 20%. This slowed speech improved comprehension in the scanner. The audio presentation lasted 12.4 minutes. Upon emerging from the scanner, participants completed a twelve-question, multiple-choice quiz concerning events and situations described in the story.

3 Data analysis

Preprocessing

We used SPM8 (Friston et al., 2007) to spatially-realign functional images (EPI) and co-register them with participants' structural images (MP-RAGE). Smoothing was 3mm isotropic, and SPM8's ICBM template was used to put the data into MNI stereotaxic coordinates.

Linking hypotheses

We linked linguistic structures (e.g. POS tag sequences, Penn-style trees, X-bar trees) to predictions about BOLD signal in two different ways.

Link #1: Surprisal

For probabilistic language models, we linked the probability of a word in its left-context to BOLD signal using the log-reciprocal of the probability of the next word. This is "surprisal" in the sense of Hale (2001).

Link #2: Node Count

With non-probabilistic grammars, we linked the syntactic structure of a sentence to the BOLD signal it evokes by counting the number of tree nodes between successive words, including "empty" nodes such as the traces of movement. This link expresses the basic claim that more grammatical structure implies greater comprehension effort. While intuitive, the precise formulation of this idea has been tricky; see Frazier (1985, section 4.4) for critical discussion. Our two node count hypotheses were based, respectively, on top-down and bottom-up parsing (see e.g. Hale, 2014, chapter 3). The top-down traversal that we used enumerates nodes in a depth-first, left to right order analogous to an LL parser. The bottom-up traversal that we used enumerates daughters before mothers in the manner of a shift-reduce LR parser. Taking the stimulus text to be largely unambiguous for native English-speaking listeners, we assume a perfect oracle that enumerates nodes of just the correct tree.

Hemodynamic Response

Via these linking hypotheses, we derived time series of predictions about the effortfulness of comprehending each word in the text. Following Just & Varma (2007) we convolved these time series with SPM8's canonical hemodynamic response function (HRF) to arrive at an expected BOLD signal. This HRF is a difference of Gamma functions (Friston et al., 2007, chapter 14). Figure 1 summarizes this methodology graphically.

Regions of interest

By defining an impulse at the offset of each spoken word, an atheoretical predictor call Rate local-

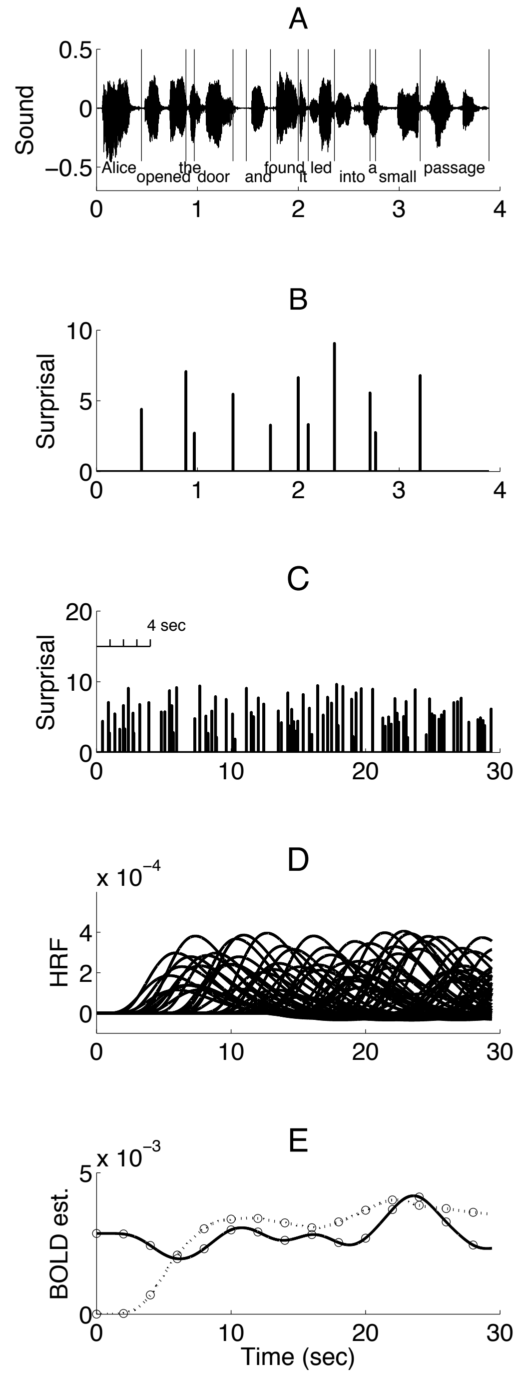


Figure 1: Deriving an expected BOLD signal from a sequence of linguistic structures (analogous to Fig. 11 of Just & Varma (2007)): (A) Segmentation of a spoken narrative (B) Complexity metric, such as surprisal, defines intensity of point event (C) Points shown over a longer interval (D) Points convolved with canonical HRF (E) Summed HRFs yield estimate of BOLD response (dotted) made orthogonal to low-level covariates (solid).

izes brain regions whose BOLD signal varies in time with the speech stimulus. We localized spherical regions of interest, each with radius 10mm, centered on the maxima of the Rate predictor within three anatomically-constrained regions that were defined based on prior work on the neural bases of syntax.

Maxima for this predictor that fell bilaterally in temporal lobe anterior to Heschl's Gyrus served to define the center of left and right anterior temporal region. Anterior temporal lobe has shown sensitivity to the presence vs. absence of constituent structure (Stowe et al., 1998; Vandenberghe et al., 2002; Humphries et al., 2006; Snijders et al., 2009; Bemis and Pylkkänen, 2011; Pallier et al., 2011) and brain damage to this region correlates with deficits in morphosyntax (Dronkers et al., 2004). These data, and others, have led to the proposal that the anterior temporal lobe is involved in "basic syntactic processes." (Friederici and Gierhan, 2013, p. 252).

Maxima of the Rate predictor that fell in the left frontal lobe and were listed as "inferior frontal gyrus" in the Harvard-Oxford Brain Atlas defined the center of our left inferior frontal gyrus region. Numerous findings from lesion-induced syntactic deficits (Caramazza and Zurif, 1976; Grodzinsky, 2000) and neuroimaging of brain activations for simple and syntactically complex sentences (Just et al., 1996; Stromswold et al., 1996; Stowe et al., 1998; Snijders et al., 2009; Santi and Grodzinsky, 2007) have implicated this region in various aspects of grammatical processing.

Statistical Analysis

Statistical analysis was conducted in two stages. In the first stage, we constructed a family of mixed-effects regression models using non-syntactic predictors together with one syntax predictor drawn from each model. Parameters more than two standard errors away from zero were taken to "significant" in this analysis (Gelman and Hill, 2007).

In the second stage we conducted a set of stepwise model comparisons to evaluate the unique contribution of each "grain size" which showed a significant contribution in stage one. Comparisons were evaluated using likelihood ratio tests. Models were nested in order of smallest to largest grain size (i.e. amount of hierarchy), and least to most predictive for measures of node count. The list of fixed effects for each model entered in to this comparison is given in Table 1.

All models included fixed effects for word Rate (see above), log unigram frequency, and three principle components representing head-movements, heart rate, and lung action. A fixed effect for prosodic breaks was also included to control for correlations between acoustic variance and syntactic structure. This predictor is a perceptual judgment of break index strength made in light of ToBI annotation guidelines by two independent raters. Subjects were treated as a random intercept in all models.

4 Structure at various grain sizes

Markov Models

2gram.l, 2gram.p, 3gram.l, 3gram.p – We used OpenGRM to fit Markov Models of various orders (Allauzen et al., 2007). These models were trained on the version of Alice in Wonderland that is distributed by Project Gutenberg, etext # 11. As a preprocessing step, chapter headings were removed and all words converted to lowercase. Lexicalized (*.l*) and unlexicalized POS (*.p*) models were created.

Penn-style Phrase Structure

cfg.surp, cfg.bu, cfg.td – We used the EarleyX implementation of Stolcke's probabilistic Earley parser to compute surprisal values from phrase structure grammars (Luong et al., 2013; Stolcke, 1995). We used a grammar whose rules came from Stanford parser output, when applied to the entire Alice in Wonderland book (Klein and Manning, 2003). Punctuation was removed. This renders the training data comparable across *cfg.surp* and $\{2,3\}$*gram*. The node count predictors were based on the same Penn-style structures as in Brennan et al. (2012).

X-bar Trees

mg.bu, mg.td – We used Minimalist Grammars to define more detailed analyses for each sentence. These grammars extend and reorganize the analyses discussed in Hale (2003, chapter 4) in a way that is guided by Sportiche, Koopman & Stabler (2013). They derive "X-bar" structural descriptions that integrate constituency, dependency and movement information. Figure 3 highlights a case where the X-bar predictor includes additional de-

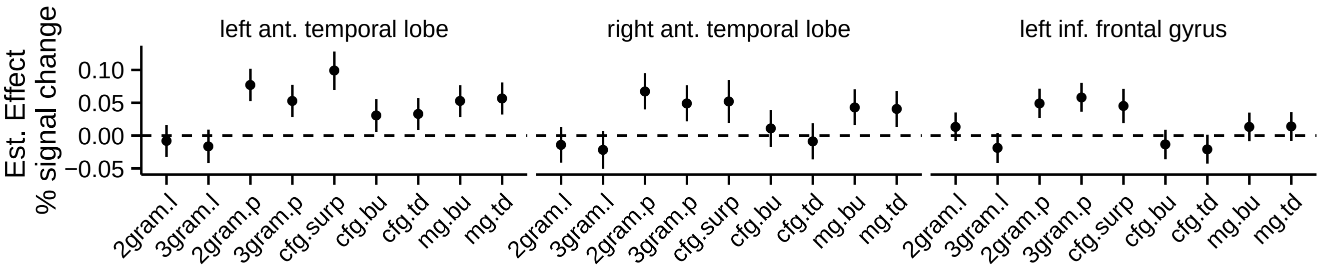


Figure 2: Magnitude of fitted coefficients across all syntax predictors, considered individually. Fine-grained linguistic structure from Penn-style *cfg.surp* and X-bar structures $mg.\{bu, td\}$ are positive predictors of the neural time-course in anterior temporal lobe. See Table 2 for stepwise model comparison.

tail, namely about movement. Of course, these trees encode many other aspects of sentence structure that are treated in Minimalist theories of syntax (see e.g. Adger (2003) or Hornstein, Nunes and Grohmann (2005) for an introduction). Traversing these trees in either Bottom-Up or Top-Down order, we obtain node counts analogous to those used in Brennan et al (2012).

model	*description*
Ø	lexical, prosodic, and physiological but no syntactic predictors
A	add POS tag 2-gram
B	add POS tag 3-gram
C	add CFG surprisal
D	add bottom-up CFG node count
E	add bottom-up X-bar node count
F	add top-down X-bar node count

Table 1: Nested models

left aTL	*predictor*	$\chi^2(1)$	*p*
Ø to A	2gram.p	38.9	$< .001$
B to C	cfg.surp	22.3	$< .001$
D to E	mg.bu	5.5	< 0.05
right aTL			
Ø to A	2gram.p	23.7	$< .001$
D to E	mg.bu	4.3	< 0.05
left IFG			
Ø to A	2gram.p	19.6	$< .001$
A to B	3gram.p	8.4	$< .01$

Table 2: Statistically-significant steps in a model comparison

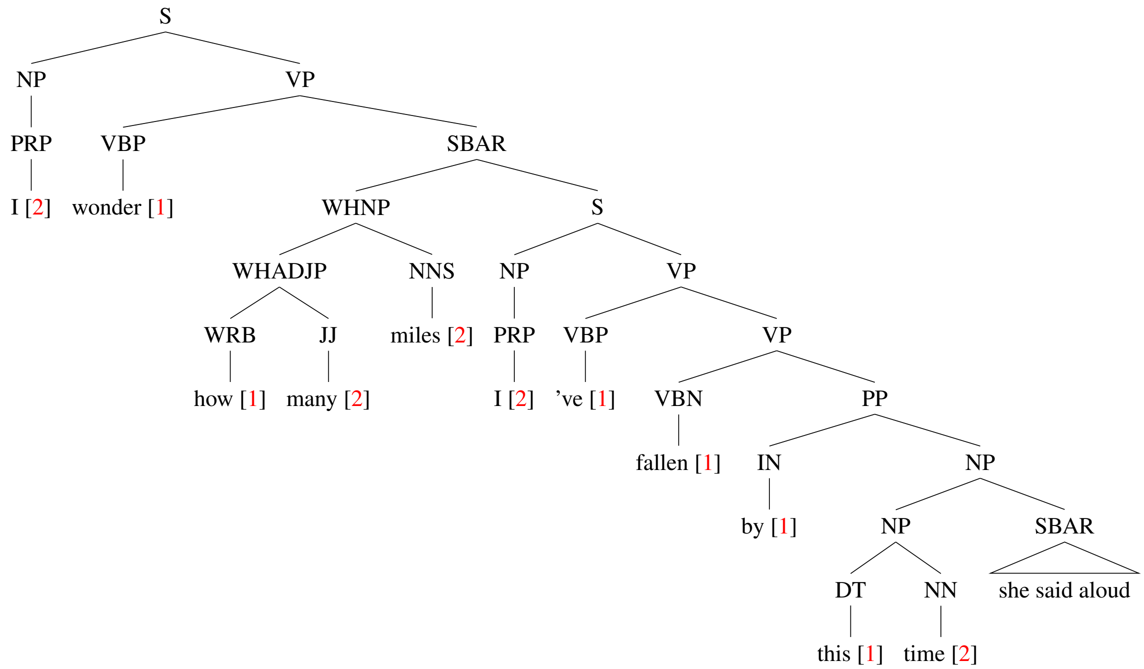
5 Results

Fine-grained predictors based on X-bar trees and Penn-style phrase structure each improved a mixed-effects model of the neural time course in anterior temporal lobe during naturalistic story comprehension. This did not obtain in the time courses from inferior frontal gyrus. Figure 2 shows the estimated coefficients (± 2 standard errors) for each of the syntax predictors when included alone in a model with only word-level and physiological “nuisance” predictors. Table 2 reports a model comparison that tests which predictors contribute independently of other, coarser-grained predictors. It lists the steps that reached statistical significance at the $p < 0.05$ level for each ROI. Table 2 uses the letters A–F to identify increasingly refined models along a progression that is described in Table 1.

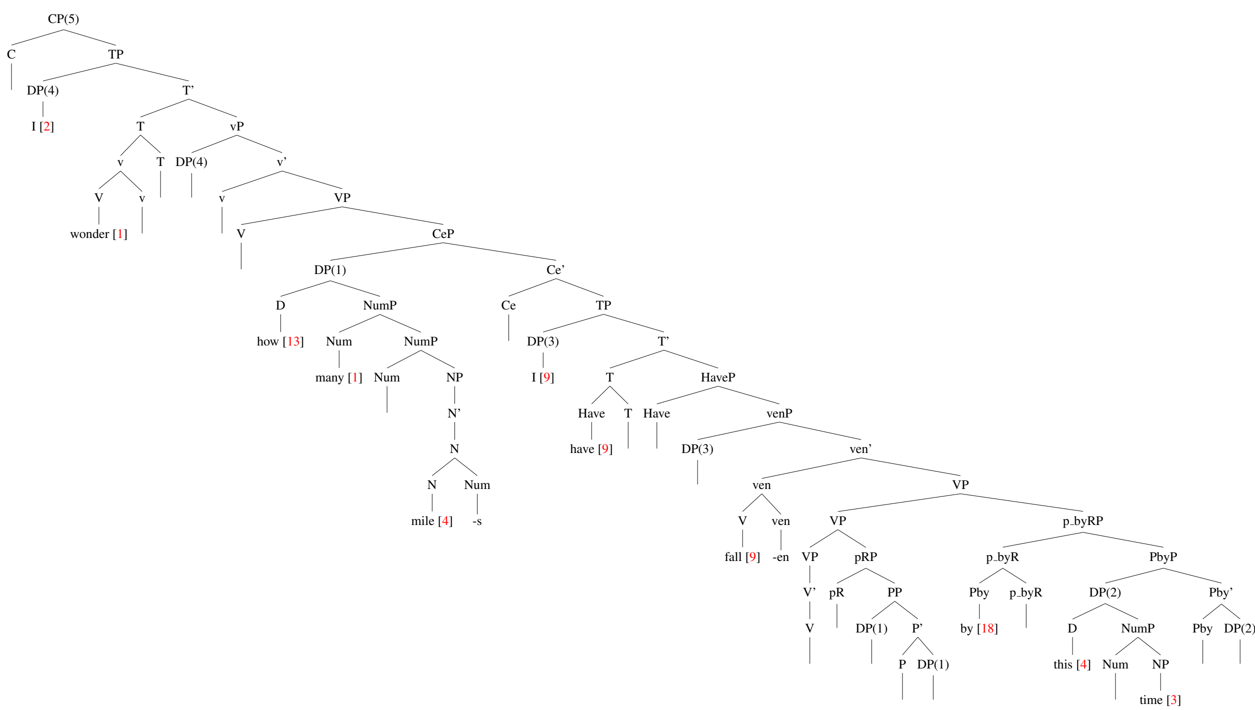
Performance on the post-scan was substantially higher (median=11) than chance (3 out of 12). This confirms that participants were indeed attending to the story.

6 Discussion

As Figure 2 indicates, a variety of grammatical predictors turned out to be helpful in explaining BOLD signals. Even the most abstruse predictors that we considered, ones based on X-bar structures generated by Minimalist Grammars, led to reliable improvements over baseline models. This suggests



(a) Coarse-grained Penn-style structure. Bracketted numbers show the value of *cfg.bu* at each word.



(b) Fine-grained X-bar structure. Bracketted numbers show the value of *mg.bu* at each word.

Figure 3: Coarse-grained versus fine-grained syntactic analyses of the same sentence. Numbers in square brackets are BOLD signal predictors. They reflect the presence of empty nodes in 3(b) but not 3(a). This aspect of the structure impacts the prediction at the word "by".

that indeed, the human sentence processing system is sensitive to hierarchical structure at least in the anterior temporal lobe. It is possible that the null result reported by Frank et al. (2011; 2015) reflects the difficulty of measuring nuanced syntactic processing activity with behavioral and ERP measures.

While n-gram predictors were helpful throughout, Penn phrase structures and X-bar structures did not gain purchase in inferior frontal gyrus the way they did in anterior temporal regions. This "aTL-specificity" corroborates earlier findings that used node count but not surprisal (Brennan et al., 2012).

The bilateral character of the aTL results aligns well with related work with written stimuli by Wehbe et al (2014). Using word features related to labelled dependency arcs (i.e. noun modifier, verbal complement) and part of speech tags, Wehbe and colleagues found a cluster of voxels in right anterior temporal lobe where syntactic information contributed to high performance in a classification-by-prediction task. This points to a temporal lobe language network whose normal mode of operation employs both hemispheres.

7 Conclusion

If we take the model comparison approach — as applied to neural time courses — to be an empirical test of the Competence Hypothesis, then the Hypothesis survives. These data support the view that humans use linguistic structure to comprehend spoken narratives. This finding re-prompts the question that occupied Bresnan, Kaplan, Steedman & Stabler: which linguistic theory is most helpful in understanding that comprehension? Answering this question is more than one lab can manage. We therefore plan to release this time course data so that the broader cognitive science community can try out alternative models based on a wider variety of parsing theories.

Acknowledgments

With sadness we acknowledge our deceased collaborator, Sarah Van Wagenen, who passed away before this work was finished. We are grateful to Edward P. Stabler, Jr. for conceptual and practical help with this project, and to Rob Mason for valuable discussions. We also thank the students of LING7710: Frederick Callaway, Elana Feldman, Jaclyn Jeffrey-Wilensky and Adam Mahar. This project was supported in part by NIH grant 1S10RR025145 and NSF CAREER award number 0741666.

References

David Adger. 2003. *Core Syntax: a Minimalist Approach*. Oxford University Press.

Cyril Allauzen, Michael Riley, Johan Schalkwyk, Wojciech Skut, and Mehryar Mohri. 2007. OpenFst: A general and efficient weighted finite-state transducer library. In *Proceedings of the* 12^{th} *International Conference on Implementation and Application of Automata*, pages 11–23. Springer.

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 Pylkkäen. 2012. Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, 120(2):163 – 173.

Joan Bresnan and Ronald M. Kaplan. 1982. Introduction: Grammars as mental representations of language. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages xvii,lii. MIT Press, Cambridge, MA.

Alfonso Caramazza and Edgard B. Zurif. 1976. Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3(4):572–582.

Noam Chomsky. 1965. *Aspects of the Theory of Syntax*. MIT Press.

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: Towards a new functional anatomy of language. *Cognition*, 92(1-2):145–177.

Victoria Fossum and Roger Levy. 2012. Sequential vs. hierarchical syntactic models of human incremental sentence processing. In *Proceedings of the 3rd Annual Workshop on Cognitive Modeling and Computational Linguistics*, pages 61–69, Montreal, Quebec.

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

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

Lyn Frazier. 1985. Syntactic complexity. In David R. Dowty, Lauri Karttunen, and Arnold M. Zwicky, editors, *Natural language parsing: Psychological, computational, and theoretical perspectives*, chapter 4, pages 129–189. Cambridge University Press.

Angela D. Friederici and Sarah M.E. Gierhan. 2013. The language network. *Current Opinion in Neurobiology*, 23(2):250 – 254.

Karl J. Friston, John Ashburner, Stefan J. Kiebel, Thomas E. Nichols, and Wiliam D. Penny, editors. 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.

Andrew Gelman and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Yosef Grodzinsky and Angela D. Friederici. 2006. Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology*, 16(2):240–246.

Yosef Grodzinsky. 2000. The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, 23(01):1–21.

John Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

John Hale. 2003. *Grammar, uncertainty and sentence processing*. Ph.D. thesis, Johns Hopkins University, Baltimore, Maryland.

John Hale. 2014. Automaton theories of human sentence comprehension. CSLI Publications, September.

Norbert Hornstein, Jairo Nunes, and Kleanthes K. Grohmann. 2005. *Understanding Minimalism*. Cambridge University Press.

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

Marcel Just and Sashank Varma. 2007. The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, & Behavioral Neuroscience*, 7:153–191.

Marcel A. Just, Patricia A. Carpenter, Timothy A. Keller, William F. Eddy, and Keith R. Thulborn. 1996. Brain activation modulated by sentence comprehension. *Science*, 274(5284):114–116.

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

Minh-Thang Luong, Michael C. Frank, and Mark Johnson. 2013. Parsing entire discourses as very long strings: Capturing topic continuity in grounded language learning. *Transactions of the Association for Computational Linguistics*, 1(3):315–323.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, 19:313–330.

Scott A. McDonald and Richard C. Shillcock. 2003. Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research*, 43(16):1735–1751.

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, 02.

Andrea Santi and Yosef Grodzinsky. 2007. Working memory and syntax interact in Broca's area. *NeuroImage*, 37(1):8–17.

Tineke M. Snijders, Theo Vosse, Gerard Kempen, Jos J.A. Van Berkum, Karl Magnus Petersson, and Peter Hagoort. 2009. Retrieval and unification of syntactic structure in sentence comprehension: an fmri study using word-category ambiguity. *Cerebral Cortex*, 19(7):1493–1503.

Dominique Sportiche, Hilda Koopman, and Edward Stabler. 2013. *An Introduction to Syntactic Analysis and Theory*. Wiley-Blackwell.

Edward Stabler. 1991. Avoid the pedestrian's paradox. In Robert C. Berwick, Steven P. Abney, and Carol Tenny, editors, *Principle-Based Parsing: computation and psycholinguistics*, Studies in Linguistics and Philosophy, pages 199–237. Kluwer, Dordrecht.

Edward Stabler. 1997. Derivational minimalism. In Christian Retoré, editor, *Logical Aspects of Computational Linguistics*, pages 68–95. Springer.

Edward P. Stabler. 2011. Computational perspectives on minimalism. In *The Oxford Handbook of Linguistic Minimalism*, chapter 27. Oxford University Press.

Mark Steedman. 1989. Grammar, interpretation and processing from the lexicon. In William Marslen-Wilson, editor, *Lexical Representation and Process*, chapter 16, pages 463–504. MIT Press, Cambridge, MA.

Andreas Stolcke. 1995. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, 21(2):165–201.

Laurie A. Stowe, Cees A. J. Broere, Anne M. J. Paans, Albertus A. Wijers, Gijsbertus Mulder, Wim Vaalburg, and Frans Zwarts. 1998. Localizing components of a complex task: Sentence processing and working memory. *Neuroreport*, 9(13):2995–2999.

Karin Stromswold, David Caplan, Nathaniel Alpert, and Scott Rauch. 1996. Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, 52:452–473.

Robert Thibadeau, Marcel A. Just, and Patricia Carpenter. 1982. A model of the time course and content of reading. *Cognitive Science*, 6:157–203.

Marten van Schijndel and William Schuler. 2015. Hierarchic syntax improves reading time prediction. In *Proceedings of NAACL 2015*, Denver, Colorado, USA, June. Association for Computational Linguistics.

Rik R.C. Vandenberghe, Anna C. Nobre, and Cathy J. Price. 2002. The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, 14(4):550–560, 05.

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, November.