

Modeling incremental language comprehension in the brain with Combinatory Categorical Grammar

Miloš Stanojević^{1*} Shohini Bhattachali³ Donald Dunagan²

Luca Campanelli^{2,5} Mark Steedman¹ Jonathan R. Brennan⁴ John Hale²

¹University of Edinburgh, UK ²University of Georgia, USA

³University of Maryland, USA ⁴University of Michigan, USA ⁵Haskins Laboratories, USA

Abstract

Hierarchical sentence structure plays a role in word-by-word human sentence comprehension, but it remains unclear how best to characterize this structure and unknown how exactly it would be recognized in a step-by-step process model. With a view towards sharpening this picture, we model the time course of hemodynamic activity within the brain during an extended episode of naturalistic language comprehension using Combinatory Categorical Grammar (CCG). CCG has well-defined incremental parsing algorithms, surface compositional semantics, and can explain long-range dependencies as well as complicated cases of coordination. We find that CCG-derived predictors improve a regression model of fMRI time course in six language-relevant brain regions, over and above predictors derived from context-free phrase structure. Adding a special Revealing operator to CCG parsing, one designed to handle right-adjunction, improves the fit in three of these regions. This evidence for CCG from neuroimaging bolsters the more general case for mildly context-sensitive grammars in the cognitive science of language.

1 Introduction

The mechanism of human sentence comprehension remains elusive; the scientific community has not come to an agreement about the sorts of abstract steps or cognitive operations that would best explain people’s evident ability to understand sentences as they are spoken word-by-word. One way of approaching this question begins with a competence grammar that is well-supported on linguistic grounds, then adds other theoretical claims about how that grammar is deployed in real-time processing. The combined theory is then evaluated against observations from actual human language processing. This approach has been successful in accounting for eye-tracking data, for instance

starting from Tree-Adjoining Grammar and adding a special Verification operation (Demberg et al., 2013).

In this spirit, the current paper models the hemodynamics of language comprehension in the brain using complexity metrics from psychologically-plausible parsing algorithms. We start from a mildly context-sensitive grammar that supports incremental interpretation,¹ Combinatory Categorical Grammar (CCG; for a review see Steedman and Baldridge, 2011). We find that CCG offers an improved account of fMRI blood-oxygen level dependent time courses in “language network” brain regions, and that a special Revealing parser operation, which allows CCG to handle optional postmodifiers in a more human-like way, improves fit yet further (Stanojević and Steedman, 2019; Stanojević et al., 2020). These results underline the consensus that an expressive grammar, one that goes a little beyond context-free power, will indeed be required in an adequate model of human comprehension (Joshi, 1985; Stabler, 2013).

2 A focus on the algorithmic level

A step-by-step process model for human sentence parsing would be a proposal at Marr’s (1982) middle level, the algorithmic level (for a textbook introduction to these levels, see Bermúdez, 2020, §2.3). While this is a widely shared research goal, a large proportion of prior work linking behavioral and neural data with parsing models has relied upon

¹This work presupposes that sentence interpretation for the most part reflects compositional semantics, and that comprehension proceeds by and large incrementally. This perspective does not exclude the possibility that highly frequent or idiosyncratic patterns might map directly to interpretations in a noncompositional way (see Ferreira and Patson, 2007; Blache, 2018 as well as Slattery et al., 2013; Paolazzi et al., 2019 and discussion of Bever’s classic 1970 proposal by Phillips 2013). de Lhoneux et al. (2019) shows how to accommodate these cases as multi-word expressions in a CCG parser. Bhattachali et al. (2019) maps brain regions implicated in these two theorized routes of human sentence processing.

*Correspondence to m.stanojevic@ed.ac.uk

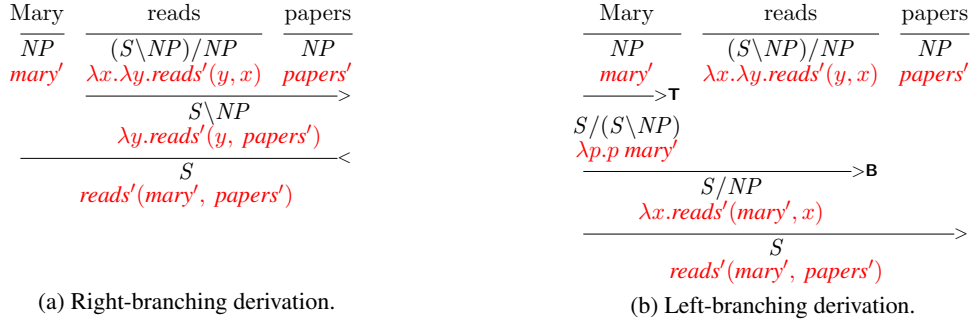


Figure 1: Semantically equivalent CCG derivations.

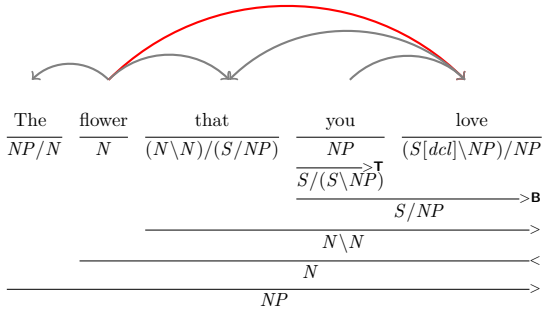


Figure 2: CCG derivation and extracted semantic dependencies for a relative clause from the Little Prince. Red highlighting indicates filler-gap relationship.

the surprisal linking hypothesis, which is not an algorithm. In fact surprisal wraps an abstraction barrier around an algorithmic model, deriving predictions solely from the probability distribution on that model’s outputs (for a review see Hale, 2016). This abstraction is useful because it allows for the evenhanded comparison of sequence-oriented models such as ngrams or recurrent neural networks against hierarchical, syntax-aware models. And indeed in eye-tracking, this approach confirms that some sort of hierarchical structure is needed (see e.g. Fossum and Levy, 2012; van Schijndel and Schuler, 2015). This same conclusion seems to be borne out by fMRI data (Henderson et al., 2016; Brennan et al., 2016; Willems et al., 2016; Shain et al., 2020).

But precisely because of the abstraction barrier that it sets up, surprisal is ill-suited to the task of distinguishing ordered steps in a processing mechanism. We therefore put surprisal aside in this paper, focusing instead on complexity metrics that are nearer to algorithms; the ones introduced below in §5.3 all map directly on to tree traversals. By counting derivation tree nodes, these metrics track work that the parser does, rather than the rar-

ity of particular words or ambiguity of particular constructions.²

Previous research at the algorithmic level has been limited in various ways. Brennan et al. (2016) used an expressive grammar, but it was not broad coverage and the step counts were based on derived X-bar trees rather than the derivation trees that would need to be handled by a provably correct parsing algorithm (Stanojević and Stabler, 2018). Brennan et al. (2020) used a full-throated parser but employed the Penn Treebank phrase structure without explicit regard for long-distance dependency. Figure 2 shows an example of one of these dependencies.

3 Why CCG?

CCG presents an opportunity to remedy the limitations identified above in section 2. As already mentioned, CCG is appropriately expressive (Vijay-Shanker and Weir, 1994). And it has special characteristics that are particularly attractive for incremental parsing. CCG can extract filler-gap dependencies such as those in the object relative clause in Figure 2, synchronously and incrementally building surface compositional semantics (cf. Demberg 2012).³ CCG also affords many different ways of

²Counting derivation-tree nodes dissociates from surprisal. Brennan et al. (2020) addresses the choice of linking hypothesis empirically by deriving both step-counting and surprisal predictors from the same parser. The former but not the latter predictor significantly improves a regression model of fMRI timecourse in posterior temporal lobe, even in the presence of a co-predictor derived from a sequence-oriented language model.

³The derivations in Figure 1 and 2 use type-raising as a parser operation. In the definition of CCG from Steedman (2000) type-raising is not a syntactic, but a lexical operation. The reason why we use it as a parsing operation is because that is the way it was defined in the CCGbank (Hockenmaier and Steedman, 2007) and because it is implemented as such in all broad coverage parsers. Type-raising contributes to the complexity metric described in Section 5.3

long as the attachment respects the node’s syntactic type. In Figure 3a the *right spine* is highlighted in blue. However, none of the constituents on the right spine can be modified by “daily” because the constituent that needs to be modified, “reads papers” was never built; it is not part of the left-branching derivation. To address this, the [Stanojević and Steedman](#) parser includes a second innovation: it applies a special *tree-rotation* operation that transforms left-branching derivations into semantically equivalent right-branching ones. In Figure 3b this operation produces a new right spine, *revealing* a node of type `S\NP`, which is the type assigned to English verb phrases in CCG. In Figure 3c the adjunct “daily” is properly attached to this boxed node via Application, a CCG rule that is used quite generally across many different constructions.

The idea of attaching right-adjuncts to a node of an already-built tree has appeared several times before ([Pareschi and Steedman, 1987](#); [Niv, 1994](#); [Ambati et al., 2015](#); [Stanojević and Steedman, 2019](#)) and in all cases it crucially leverages CCG’s flexible constituency as shown in Figure 1. See [Stanojević et al. \(2020\)](#) for more detailed treatment of Sturt and Lombardo’s construction using predictive completion. The present study examines whether or not the addition of the Revealing operation increases the fidelity of CCG-derived parsing predictions to human fMRI time course data.

5 Methods

We follow [Brennan et al. \(2012\)](#) and [Willems et al. \(2016\)](#) in using a spoken narrative as a stimulus in the fMRI study. Participants hear the story over headphones while they are in the scanner. The neuroimages collected during their session serve as data for regression modeling with word-by-word predictors derived from the text of the story.

5.1 The Little Prince fMRI Dataset

The English audio stimulus was Antoine de Saint-Exupéry’s *The Little Prince*, translated by David Wilkinson and read by Karen Savage. It constitutes a fairly lengthy exposure to naturalistic language, comprising 19,171 tokens, 15,388 words and 1,388 sentences, and lasting over an hour and a half. This is the fMRI version of the EEG corpus described in [Stehwien et al. \(2020\)](#). It has been used before to investigate a variety of brain-language questions unrelated to CCG parsing ([Bhattasali et al., 2019](#); [Bhattasali and Hale, 2019](#); [Li et al., 2018](#)). Prior to

parsing, number expressions were spelled out i.e. 42 as “forty two” and all punctuation was removed.

5.1.1 Participants

Participants comprised 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](#)). All self-identified as native English speakers and gave their written informed consent prior to participation, in accordance with the university’s IRB guidelines. Participants were compensated for their time, consistent with typical practice for studies of this kind. They were paid \$65 at the end of the session. Data from three out of the 51 participants was excluded because they did not complete the entire session or moved their head excessively.

5.1.2 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. 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 served to assess participants’ comprehension. The entire session lasted around 2.5 hours.

5.1.3 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 sequence (repetition time: 2000 ms, echo time: 27 ms, flip angle: 77deg, image acceleration: 2X, field of view: 216×216 mm, matrix size 72×72, and 44 oblique slices, yielding 3 mm

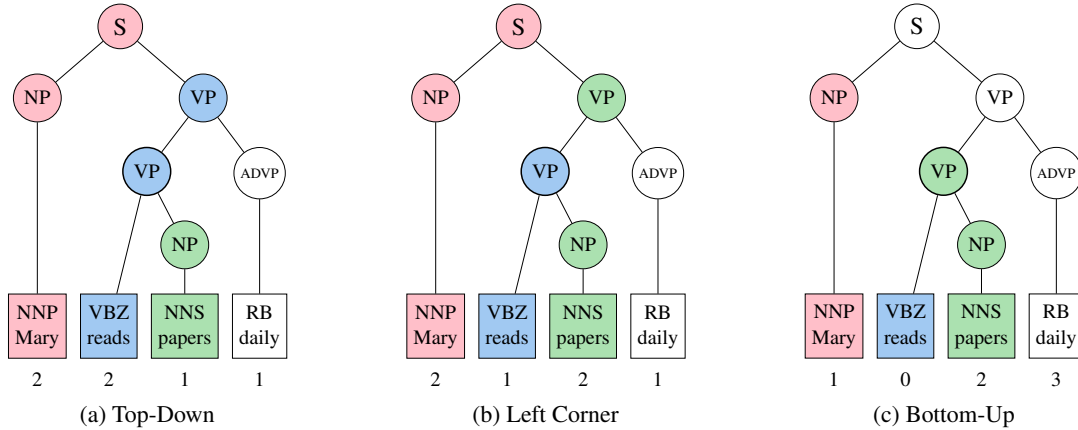


Figure 4: Different parsing strategies for constituency trees. Below each word is a complexity measure associated with that word. It is equivalent to the number of round nodes visited by the parser when the word is being integrated.

isotropic voxels). Anatomical images were collected with a high resolution T1-weighted ($1 \times 1 \times 1$ mm³ voxel) with a Magnetization-Prepared Rapid Gradient-Echo (MP-RAGE) pulse sequence.

5.1.4 Preprocessing

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).

5.2 Grammatical annotations

We annotated each sentence in *The Little Prince* with phrase structure parses from the *benepar* constituency parser (Kitaev and Klein, 2018). Previous studies have used the Stanford CoreNLP parser, but *benepar* is much closer to the current state-of-the-art in constituency parsing. To find CCG derivations we used *RotatingCCG* by Stanojević and Steedman (2019; 2020).

5.3 Complexity Metric

The complexity metric used in this study is the number of nodes visited in between leaf nodes, on a given traversal of a derivation tree. This corresponds to the number of parsing actions that would

be taken, per word, in a mechanistic model of human comprehension (see e.g. Kaplan, 1972; Frazier, 1985). These numbers (i.e. written below the leaves of the trees in Figure 4) are intended as predictions about sentence processing effort, which may be reflected in the fMRI BOLD signal (see discussion of convolution with hemodynamic response function in §6.2).

For constituency parses we examine bottom-up (aka shift-reduce parsing), top-down, and left-corner parsing. Figure 4 shows all these parsing strategies on an example constituency tree. This Figure highlights three points: (a) that the complexity metrics correspond to visited nodes of the tree (b) that they are incremental metrics, computed word by word and (c) that alternative parsing strategies lead to different predictions.

In CCG all natural parsing strategies are bottom-up. The main difference among them is what kind of derivation they deliver. We evaluate right-branching derivations, left-branching derivations and revealing derivations; the latter are simply left-branching derivations with the addition of the Revealing operation. To compute this we get the best derivation from a CCG parser and then convert it to the three different kinds of semantically equivalent derivations using the *tree-rotation* operation (Niv, 1994; Stanojević and Steedman, 2019).

In the case of revealing derivations we count only the nodes that are constructed with reduce and right-adjunction operations, but we do not count the nodes constructed with tree-rotation. This is because tree-rotation is not an operation that introduces anything new in the interpretation — tree-rotation only helps the right-adjunction operation

reveal the constituent that needs to be modified.

All parsing strategies have the same total number of nodes, but only differ in the abstract timing of those nodes' construction. In general, left-branching derivations construct nodes earlier than do the corresponding right-branching derivations. However, in the case of right-adjunction both left- and right-branching derivations delay construction of many nodes until the right-adjunct is consumed. This is not the case with the revealing derivations that are specifically designed to allow flexibility with right-adjuncts.

5.4 Hypotheses

Using the formalism-specific and parsing strategy-specific complexity metrics defined above in §5.3, we evaluate three hypotheses.

Hypothesis 1 (H1): *CCG improves a model of fMRI BOLD time courses above and beyond context-free phrase structure grammar.*

Mildly context-sensitive grammars like CCG capture properties of sentence structure that are only very inelegantly covered by context-free phrase structure grammars. For instance, the recovery of filler-gap dependency in Figure 2 follows directly from the definition of the combinators. This hypothesis supposes that the brain indeed does work to recover these dependencies, and that that work shows up in the BOLD signal.

Hypothesis 2 (H2): *The Revealing parser operation explains unique variability in the BOLD signal, variability not accounted for by other CCG derivational steps.*

As described above in §4, Revealing allows a CCG parser to handle right-adjunction gracefully. This hypothesis in effect proposes that this enhanced psychological realism should extend to fMRI.

Hypothesis 3 (H3): *Left-branching CCG derivations improve BOLD activity prediction over right-branching.*

Left-branching derivations provide maximally incremental CCG analyses. If human processing is maximally incremental, and if this incrementality is manifested in fMRI time courses, then complexity metrics based on left-branching CCG derivations should improve model fit over and above right-branching.

6 Data Analysis

6.1 Regions of Interest

We consider six regions of interest in the left hemisphere: the pars opercularis (IFG_oper), the pars triangularis (IFG_tri), the pars orbitalis (IFG_orb), the superior temporal gyrus (STG), the superior temporal pole (sATL) and the middle temporal pole (mATL). These regions are implicated in current neurocognitive models of language (Haagoort, 2016; Friederici, 2017; Matchin and Hickok, 2020). However evidence suggests that particular sentence-processing operations could be localized to different specific regions within this set (Lopopolo et al., 2021; Li and Hale, 2019; Brennan et al., 2020). We use the parcellation provided by the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) for SPM12. For each subject, extracting the average blood-oxygenation level-dependent (BOLD) signal from each region yields 2,816 data points for each region of interest (ROI). These data served as dependent measures in the statistical analyses described below in §6.3.

6.2 Predictors

The predictors of theoretical interest are the parser-derived complexity metrics described above in section 5.3. To these we add additional covariates that are known to influence human sentence processing. The first of these is Word Rate, which has the value 1 at the offset of each word and zero elsewhere. The second is (unigram) word Frequency. This is a log-transformed attestation count of the given word type in a corpus of movie subtitles (Brybaert and New, 2009). The third is the root-mean-squared (RMS) intensity of the audio. Finally we include the fundamental frequency f_0 of the narrator's voice as recovered by the RAPT pitch tracker (Talkin, 1995). These control predictors serve to rule out effects that could be explained by general properties of speech perception (cf. Goodkind and Bicknell 2021; Bullmore et al. 1999; Lund et al. 2006). The word-by-word complexity metrics are given timestamps according to the offsets of the words with which they correspond.

In order to use these predictors to model the BOLD signal, we convolve the time-aligned vectors with the SPM canonical hemodynamic response function which consists of a linear combination of two gamma functions and links neural activity and the estimated BOLD signal (see e.g.

Henson and Friston, 2007). After convolution, each of the word-by-word metrics of interest is orthogonalized against convolved word rate to remove correlations attributable to their common timing. Figure 7 in the Appendix reports correlations between these predictors.

6.3 Statistical Analyses

Data were analyzed using linear mixed-effects regression.⁴ All models included random intercepts for subjects. Random slopes for the predictors were not retained either because of convergence failures or because they did not alter the pattern of results.

A theory-guided, model comparison framework was used to contrast alternative hypotheses (articulated in §5.4). The Likelihood Ratio test was used to compare the fit of competing regression models (for an introduction, see Bliese and Ployhart, 2002). Effects were considered statistically significant with $\alpha = 0.008$ (0.05/6 regions, following the Bonferroni procedure).⁵

As a quantitative comparison between ROIs was not directly relevant for the research questions at issue, statistical analyses were carried out by region. This approach, as compared to examining the effects of, and the interactions between, all ROIs and predictors in the same analysis, reduced the complexity of the models and facilitated parameter estimation.

Hypothesis H1 was tested by examining the overall predictive power of the CCG-derived predictors over and above a baseline model that included word rate, word frequency, sound power, fundamental frequency, and word-by-word node counts derived from all three phrase structure parsing strategies:

$$(I) \text{ BOLD} \sim \text{word_rate} + \text{word_freq} + \text{RMS} + f_0 + \text{bottom-up} + \text{top-down} + \text{left-corner} \{ \text{CCG-left} + \text{CCG-right} + \text{CCG-revealing} \}$$

To test H2, we examined whether node counts incorporating the Reveal operation explained BOLD signal variability over and above a baseline model that included, in addition to the variables in (I), node counts from left branching and right branching CCG derivations:

$$(II) \text{ BOLD} \sim \text{word_rate} + \text{word_freq} + \text{RMS} + f_0 + \text{bottom-up} + \text{top-down} + \text{left-corner} + \text{CCG-left} + \text{CCG-right} \{ \text{CCG-revealing} \}$$

Last, for H3 in section 5.4, we tested whether word-by-word traversals of left branching CCG derivations accounted for any significant amount of BOLD signal variability over and above right branching. This amounts to asking whether CCG processing is maximally eager or maximally delayed.

$$(III) \text{ BOLD} \sim \text{word_rate} + \text{word_freq} + \text{RMS} + f_0 + \text{bottom-up} + \text{top-down} + \text{left-corner} + \text{CCG-right} \{ \text{CCG-left} \}$$

7 Results

Behavioral results on the comprehension task showed attentive listening to the spoken narrative with average response accuracy of 90% (SD = 3.7%).

7.1 H1: CCG-specific effects

The first question that we investigated was whether CCG derivations would account for any significant amount of BOLD activity over and above bottom-up, top-down, and left-corner phrase structure parsing strategies in addition to baseline covariates (i.e. as introduced above in §5.3 and depicted in Figure 4). The overall predictive power of the three CCG derivations emerged to significantly improve the models fit in all six regions examined, thus providing strong support for H1. For all analyses, the complete tables of results are provided in the Appendix (Tables 1 to 6).

To better understand the source of those effects, we followed-up with an additional set of analyses in which we contrasted one CCG parsing strategy at a time against the same baseline model. These CCG parsing strategies exhibit a region-specific pattern of fits which is summarized in Figure 5.⁶

7.2 H2: The Revealing parser operation

The second hypothesis, H2 in section 5.4, is about hemodynamic effects of the Revealing operation. The results summarized in Figure 6 supported this hypothesis: the CCG-revealing predictor significantly improved model fit to the BOLD signal in three of six ROIs examined (IFG_tri, IFG_oper,

⁴Regression analyses used the lme4 R package (version 1.1-26; Bates et al., 2015).

⁵A Bonferroni correction of 0.05/6 reflects the fact each of the three hypotheses was tested with a single Likelihood Ratio test per ROI, irrespective of the number of variables in the models compared.

⁶The direction of the effects for the analyses in both Figure 5 and 6 was not affected by the correlation among variables (Figure 7 in the Appendix).

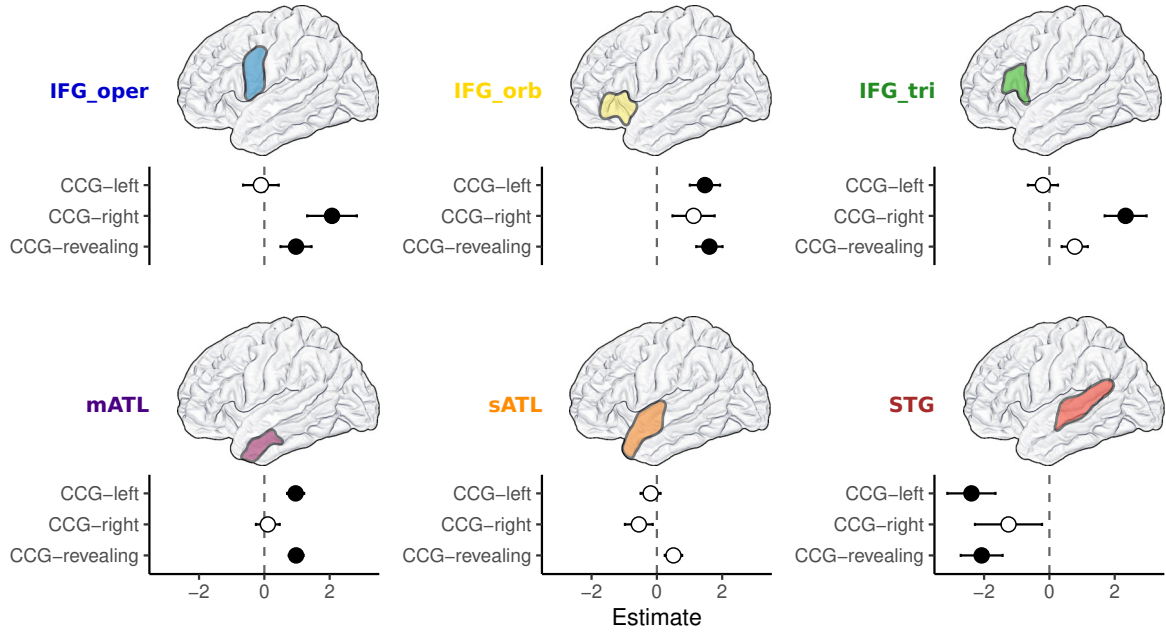


Figure 5: CCG derivation effects by ROI. Coefficient point estimates \pm SE. Filled points indicate that the predictor significantly improved model fit. Note that for IFG_oper the CCG-revealing predictor is only marginally significant after Bonferroni correction across ROIs.

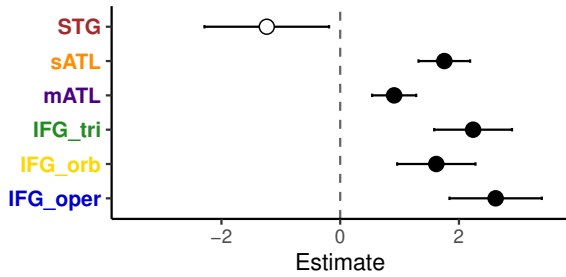


Figure 6: Effects of the CCG-revealing predictor by ROI. Coefficient point estimates \pm SE. Filled points indicate that the predictor significantly improved model fit. Note that for IFG_orb and mATL, the effects became only marginally significant after Bonferroni correction.

sATL) and marginally significant in two others after Bonferroni correction (IFG_orb and mATL). The positive sign of the statistically significant coefficients in Figure 6 indicates that, as expected, increased processing cost, as derived from the CCG-revealing parser, was associated with increased BOLD activity.

7.3 H3: Left- versus Right-branching

In the last set of analyses, we investigated whether left-branching CCG derivations improve BOLD activity predictions over right-branching derivations

(H3 in section 5.4).

It emerged that the CCG-left predictor significantly improved model fit in IFG_tri, IFG_orb, STG, mATL, and, but only marginally significant after Bonferroni correction, IFG_oper. These findings, overall, indicate the ability of left branching CCG derivations to account for a unique amount of BOLD activity during language processing.

8 Discussion

The improvement that CCG brings to modeling fMRI time courses — over and above predictors derived from well-known context-free parsing strategies — confirms that mildly context-sensitive grammars capture real aspects of human sentence processing, as suggested earlier by Brennan et al. (2016). We interpret the additional improvement due to the Revealing operation as neurolinguistic evidence in support of that particular way of achieving heightened incrementality in a parser. While it is possible that other incremental parsing techniques might adequately address the challenge of right adjunction (see §4 above) we are at present unaware of any that are supported by evidence from human neural signals. The patterning of fits across regions aligns with the suggestion that different kinds of processing, some more eager and others less so, may be happening across

the brain (cf. [Just and Varma 2007](#)). For instance the explanatory success of predictors derived from left-branching and Revealing derivations in the middle temporal pole (mATL) supports the idea that this region tracks tightly time-locked, incremental language combinatorics⁷ while other regions such as the inferior frontal gyrus (IFG) hang back, waiting to process linguistic relationships until the word at which they would be integrated into a right-branching CCG derivation (roughly consistent with [Friederici, 2017](#); [Pylkkänen, 2019](#)).

In the superior temporal gyrus (STG) the sign of the effect changes for CCG-derived predictors. This is the unique region where [Lopopolo et al. \(2021\)](#) observe an effect of phrase structure processing, as opposed to dependency grammar processing. This could be because our CCG is lexicalized. Of course, the CCGbank grammar captures many other aspects of sentence structure besides lexical dependencies (see [Hockenmaier and Steedman 2007](#)).

[Shain et al. \(2020\)](#) use a different, non-combinatory categorial grammar to model fMRI time courses. Whereas this earlier publication employs the surprisal linking hypothesis to study predictive processing, the present study considers instead the parsing steps that would be needed to recover grammatical descriptions assigned by CCG. This distinction can be cast as the difference between Marr’s computational and algorithmic levels of analysis, as suggested above in §2. But besides the choice of vantage point, there are conceptual differences that lead to different modeling at both levels. For instance, the generalized categorial grammar of [Shain et al. \(2020\)](#) is quite expressive and may go far beyond context-free power. But in that study it was first flattened into a probabilistic context-free grammar (PCFG) to derive surprisal predictions. The present study avoids this step by deriving processing complexity predictions directly from CCG derivations using node count. This directness is important when reasoning from human data, such as neural signals, to mathematical properties of formal systems, such as grammars (see discussion of Competence hypotheses in [Steedman, 1989](#)).

⁷This predictive relationship between left-branching derivations in middle temporal pole timecourses is observed in (the brains of) native speakers of English, a head-initial language. An exciting direction for future work concerns the possibility that the brain bases of language processing might covary with typological distinctions like head direction (cf. [Bornkessel-Schlesewsky and Schlewsky, 2016](#)).

This prior work by [Shain et al. \(2020\)](#) includes a telling observation: that surprisal from a 5-gram language model improves fit to brain data, over and above a PCFG. [Shain et al.](#) hypothesize that this additional contribution is possible expressly because of PCFGs’ context-freeness, and that a (mildly) context-sensitive grammar would do better. The results reported here are consistent with this suggestion.

9 Conclusion and Future Work

CCG, a mildly context-sensitive grammar, helps explain the time course of word-by-word language comprehension in the brain over and above Penn Treebank-style context-free phrase structure grammars regardless of whether they are parsed left-corner, top-down or bottom-up. This special contribution from CCG is likely attributable to its more realistic analysis of “movement” constructions (e.g. [Figure 2](#)) which would not be assigned by naive context-free grammars. CCG’s flexible approach to constituency may offer a way to understand both immediate and delayed subprocesses of language comprehension from the perspective of a single grammar. The Revealing operation, designed to facilitate more human-like CCG parsing, indeed leads to increased neurolinguistic fidelity in a subset of brain regions that have been previously implicated in language comprehension.

We look ahead in future work to quantifying the effect of individual complexity metrics across brain regions using alternative metrics related to surprise and memory (e.g. [Graf et al., 2017](#)). This future work also includes investigation of syntactic ambiguity, for instance via beam search along the lines of [Crabbé et al. \(2019\)](#) using the incremental neural CCG model of [Stanojević and Steedman \(2020\)](#).

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant numbers 1903783 and 1607251. The work was supported by an ERC H2020 Advanced Fellowship GA 742137 SEMANTAX grant.

Ethical Considerations

The fMRI study described in section 5.1 was approved by Cornell University’s Institutional Review Board under protocol ID #1310004157

References

- Steven P. Abney and Mark Johnson. 1991. [Memory requirements and local ambiguities of parsing strategies](#). *Journal of Psycholinguistic Research*, 20:233–249.
- Gerry Altmann and Mark Steedman. 1988. [Interaction with context during human sentence processing](#). *Cognition*, 30:191–238.
- Bharat Ram Ambati, Tejaswini Deoskar, Mark Johnson, and Mark Steedman. 2015. [An Incremental Algorithm for Transition-based CCG Parsing](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 53–63. Association for Computational Linguistics.
- Douglas Bates, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. [Fitting linear mixed-effects models using lme4](#). *Journal of Statistical Software*, 67(1):1–48.
- José Luis Bermúdez. 2020. *Cognitive science: an introduction to the science of the mind*. Cambridge University Press.
- Shohini Bhattasali, Murielle Fabre, Wen-Ming Luh, Hazem Al Saied, Mathieu Constant, Christophe Pallier, Jonathan R. Brennan, R. Nathan Spreng, and John T. Hale. 2019. [Localising memory retrieval and syntactic composition: An fMRI study of naturalistic language comprehension](#). *Language, Cognition and Neuroscience*, 34(4):491–510.
- Shohini Bhattasali and John Hale. 2019. [Diathesis alternations and selectional restrictions: A fMRI study](#). *Papers from the Annual Meeting of the Chicago Linguistic Society*, 55:33–43.
- Philippe Blache. 2018. [Light-and-deep parsing: A cognitive model of sentence processing](#). In Thierry Poibeau and Aline Villavicencio, editors, *Language, Cognition and Computational Models*, pages 27–52. Cambridge University Press, Cambridge, U.K.
- Paul D. Bliese and Robert E. Ployhart. 2002. [Growth modeling using random coefficient models: Model building, testing, and illustrations](#). *Organizational Research Methods*, 5(4):362–387.
- Ina Bornkessel-Schlesewsky and Matthias Schlewsky. 2016. [The importance of linguistic typology for the neurobiology of language](#). *Linguistic Typology*, 20(3):303.
- 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.
- Jonathan R. Brennan, Chris Dyer, Adhiguna Kuncoro, and John T. Hale. 2020. [Localizing syntactic predictions using recurrent neural network grammars](#). *Neuropsychologia*, 146:107479.
- Jonathan R. Brennan and Liina Pylkkänen. 2017. [MEG evidence for incremental sentence composition in the anterior temporal lobe](#). *Cognitive Science*, 41(S6):1515–1531.
- Jonathan R. Brennan, Edward P. Stabler, Sarah E. Van Wagenen, Wen-Ming Luh, and John T. Hale. 2016. [Abstract linguistic structure correlates with temporal activity during naturalistic comprehension](#). *Brain and Language*, 157-158:81–94.
- 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.
- Edward T. Bullmore, Michael J. Brammer, Sophia Rabe-Hesketh, Vivienne A. Curtis, Robin G. Morris, Steve C.R. Williams, Tonmoy Sharma, and Philip K. 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.
- Robert W. Cox. 1996. [AFNI: software for analysis and visualization of functional magnetic resonance neuroimages](#). *Computers and Biomedical research*, 29(3):162–173.
- Benoit Crabbé, Murielle Fabre, and Christophe Pallier. 2019. [Variable beam search for generative neural parsing and its relevance for the analysis of neuroimaging signal](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1150–1160, Hong Kong, China. Association for Computational Linguistics.
- Miryam de Lhoneux, Omri Abend, and Mark Steedman. 2019. [Investigating the effect of automatic MWE recognition on CCG parsing](#). In Yannick Parmentier and Jakub Waszczuk, editors, *Representation and parsing of multiword expressions: current trends*, pages 183–215. Language Science Press.
- Vera Demberg. 2012. [Incremental derivations in CCG](#). In *Proceedings of the 11th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+11)*, pages 198–206, Paris, France.
- Vera Demberg, Frank Keller, and Alexander Koller. 2013. [Incremental, predictive parsing with psycholinguistically motivated Tree-Adjoining Grammar](#). *Computational Linguistics*, 39(4):1025–1066.
- Fernanda Ferreira and Nikole D. Patson. 2007. [The ‘good enough’ approach to language comprehension](#). *Language and Linguistics Compass*, 1(1-2):71–83.
- Victoria Fossum and Roger Levy. 2012. [Sequential vs. hierarchical syntactic models of human incremental sentence processing](#). In *Proceedings of the*

- 3rd Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2012), pages 61–69, Montréal, Canada. Association for Computational Linguistics.
- 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. 2017. *Language in Our Brain: The Origins of a Uniquely Human Capacity*. MIT Press.
- Adam Goodkind and Klinton Bicknell. 2021. Local word statistics affect reading times independently of surprisal. *arXiv preprint arXiv:2103.04469*.
- Thomas Graf, James Monette, and Chong Zhang. 2017. Relative clauses as a benchmark for Minimalist parsing. *Journal of Language Modelling*, 5:57–106.
- Peter Hagoort. 2016. MUC (Memory, Unification, Control): a model on the neurobiology of language beyond single word processing. In Gregory Hickok and Steven L. Small, editors, *Neurobiology of language*, chapter 28. Elsevier.
- John Hale. 2016. Information-theoretical complexity metrics. *Language and Linguistics Compass*, 10(9):397–412.
- John T. Hale. 2014. *Automaton Theories of Human Sentence Comprehension*. CSLI, Stanford.
- John M Henderson, Wonil Choi, Matthew W Lowder, and Fernanda Ferreira. 2016. Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. *Neuroimage*, 132:293–300.
- Rik Henson and Karl Friston. 2007. Convolution models for fMRI. In Karl J. Friston, John T. Ashburner, Stefan J. Kiebel, Thomas E. Nichols, and William D. Penny, editors, *Statistical parametric mapping: the analysis of functional brain images*, chapter 14. Academic Press.
- Julia Hockenmaier and Mark Steedman. 2007. CCG-bank: A corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics*, 33(3):355–396.
- Aravind K. Joshi. 1985. Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In David R. Dowty, Lauri Karttunen, and Arnold M. Zwicky, editors, *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, page 206–250. Cambridge University Press.
- 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.
- Ronald M. Kaplan. 1972. Augmented transition networks as psychological models of sentence comprehension. *Artificial Intelligence*, 3:77–100.
- Nikita Kitaev and Dan Klein. 2018. Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2676–2686, Melbourne, Australia. Association for Computational Linguistics.
- 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.
- Jixing Li, Murielle Fabre, Wen-Ming Luh, and John Hale. 2018. Modeling brain activity associated with pronoun resolution in English and Chinese. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*.
- Jixing Li and John Hale. 2019. Grammatical predictors for fMRI time-courses. In Robert C. Berwick and Edward P. Stabler, editors, *Minimalist parsing*, pages 159–173. Oxford University Press.
- Alessandro Lopopolo, Antal van den Bosch, Karl-Magnus Petersson, and Roel M. Willems. 2021. Distinguishing syntactic operations in the brain: Dependency and phrase-structure parsing. *Neurobiology of Language*, 2(1):152–175.
- 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.
- David Marr. 1982. *Vision: A computational investigation into the human representation and processing of visual information*. Freeman.
- William Marslen-Wilson. 1973. Linguistic structure and speech shadowing at very short latencies. *Nature*, 244:522–523.
- William Matchin and Gregory Hickok. 2020. The cortical organization of syntax. *Cerebral Cortex*, 30(3):1481–1498.
- Matthew J. Nelson, Imen El Karoui, Kristof Giber, Xiaofang Yang, Laurent Cohen, Hilda Koopman, Sydney S. Cash, Lionel Naccache, John T. Hale, Christophe Pallier, and Stanislas Dehaene. 2017. Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academic of Sciences*, 114(18):E3669–E3678.

- Michael Niv. 1994. [A psycholinguistically motivated parser for CCG](#). In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 125–132. Association for Computational Linguistics.
- Richard C. Oldfield. 1971. [The assessment and analysis of handedness: the Edinburgh inventory](#). *Neuropsychologia*, 9(1):97–113.
- Caterina Laura Paolazzi, Nino Grillo, Artemis Alexiadou, and Andrea Santi. 2019. [Passives are not hard to interpret but hard to remember: evidence from online and offline studies](#). *Language, Cognition and Neuroscience*, 34(8):991–1015.
- Remo Pareschi and Mark Steedman. 1987. [A Lazy Way to Chart-parse with Categorical Grammars](#). In *Proceedings of the 25th Annual Meeting on Association for Computational Linguistics, ACL '87*, pages 81–88, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jonathan W. Peirce. 2007. [Psychopy—Psychophysics software in Python](#). *Journal of Neuroscience Methods*, 162(1):8–13.
- Colin Phillips. 2013. [Parser-grammar relations: We don't understand everything twice](#). In Montserrat Sanz, Itziar Laka, and Michael K. Tanenhaus, editors, *Language Down the Garden Path: The Cognitive and Biological Basis of Linguistic Structures*, pages 294–315. Oxford University Press.
- Liina Pykkänen. 2019. [The neural basis of combinatory syntax and semantics](#). *Science*, 366(6461):62–66.
- Philip Resnik. 1992. [Left-corner parsing and psychological plausibility](#). In *Proceedings of the 14th International Conference on Computational Linguistics, COLING 92*, pages 191–197.
- Cory Shain, Idan Asher Blank, Marten van Schijndel, William Schuler, and Evelina Fedorenko. 2020. [fMRI reveals language-specific predictive coding during naturalistic sentence comprehension](#). *Neuropsychologia*, 138:107307.
- Timothy J. Slattery, Patrick Sturt, Kiel Christianson, Masaya Yoshida, and Fernanda Ferreira. 2013. [Lingering misinterpretations of garden path sentences arise from competing syntactic representations](#). *Journal of Memory and Language*, 69(2):104–120.
- Edward P. Stabler. 2013. [The epicenter of linguistic behavior](#). In Montserrat Sanz, Itziar Laka, and Michael K. Tanenhaus, editors, *Language Down the Garden Path: The Cognitive and Biological Basis for Linguistic Structures*, chapter 17, pages 316–323. Oxford University Press.
- Miloš Stanojević and Edward Stabler. 2018. [A Sound and Complete Left-Corner Parsing for Minimalist Grammars](#). In *Proceedings of the Eighth Workshop on Cognitive Aspects of Computational Language Learning and Processing*, pages 65–74. Association for Computational Linguistics.
- Miloš Stanojević and Mark Steedman. 2019. [CCG Parsing Algorithm with Incremental Tree Rotation](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 228–239, Minneapolis, Minnesota. Association for Computational Linguistics.
- Miloš Stanojević and Mark Steedman. 2020. [Max-Margin Incremental CCG Parsing](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4111–4122, Online. Association for Computational Linguistics.
- Miloš Stanojević, John Hale, and Mark Steedman. 2020. [Predictive Processing of Coordination in CCG](#). In *Proceedings of the 33rd Annual CUNY Conference on Human Sentence Processing*, Amherst, Massachusetts. University of Massachusetts.
- Mark Steedman. 1989. [Grammar, Interpretation, and Processing from the Lexicon](#). In *Lexical Representation and Process*. MIT Press.
- Mark Steedman. 2000. *The Syntactic Process*. MIT Press, Cambridge, MA.
- Mark Steedman and Jason Baldrige. 2011. [Combinatory categorial grammar](#). In Robert D. Borsley and Kersti Börjars, editors, *Non-transformational syntax: formal and experimental models of grammar*, chapter 5. Wiley-Blackwell.
- Sabrina Stehwen, Lena Henke, John Hale, Jonathan Brennan, and Lars Meyer. 2020. [The Little Prince in 26 languages: Towards a multilingual neuro-cognitive corpus](#). In *Proceedings of the Second Workshop on Linguistic and Neurocognitive Resources*, pages 43–49.
- Patrick Sturt and Vincenzo Lombardo. 2005. [Processing coordinated structures: Incrementality and connectedness](#). *Cognitive Science*, 29(2):291–305.
- David Talkin. 1995. [A robust algorithm for pitch tracking \(RAPT\)](#). In W. B. Kleijn and K. K. Paliwal, editors, *Speech coding and synthesis*, pages 495–518. Elsevier.
- Michael Tanenhaus, Michael Spivey-Knowlton, Kathleen Eberhard, and Julie Sedivy. 1995. [Integration of visual and linguistic information in spoken language comprehension](#). *Science*, 268:1632–1634.
- Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. 2002. [Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain](#). *Neuroimage*, 15(1):273–289.

- Marten van Schijndel and William Schuler. 2015. [Hierarchic syntax improves reading time prediction](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1597–1605, Denver, Colorado. Association for Computational Linguistics.
- K. Vijay-Shanker and David J. Weir. 1994. [The equivalence of four extensions of context-free grammars](#). *Mathematical Systems Theory*, 27:27–511.
- Roel M Willems, Stefan L. Frank, Annabel D Nijhof, Peter Hagoort, and Antal van den Bosch. 2016. [Prediction during natural language comprehension](#). *Cerebral Cortex*, 26(6):2506–2516.

Appendix

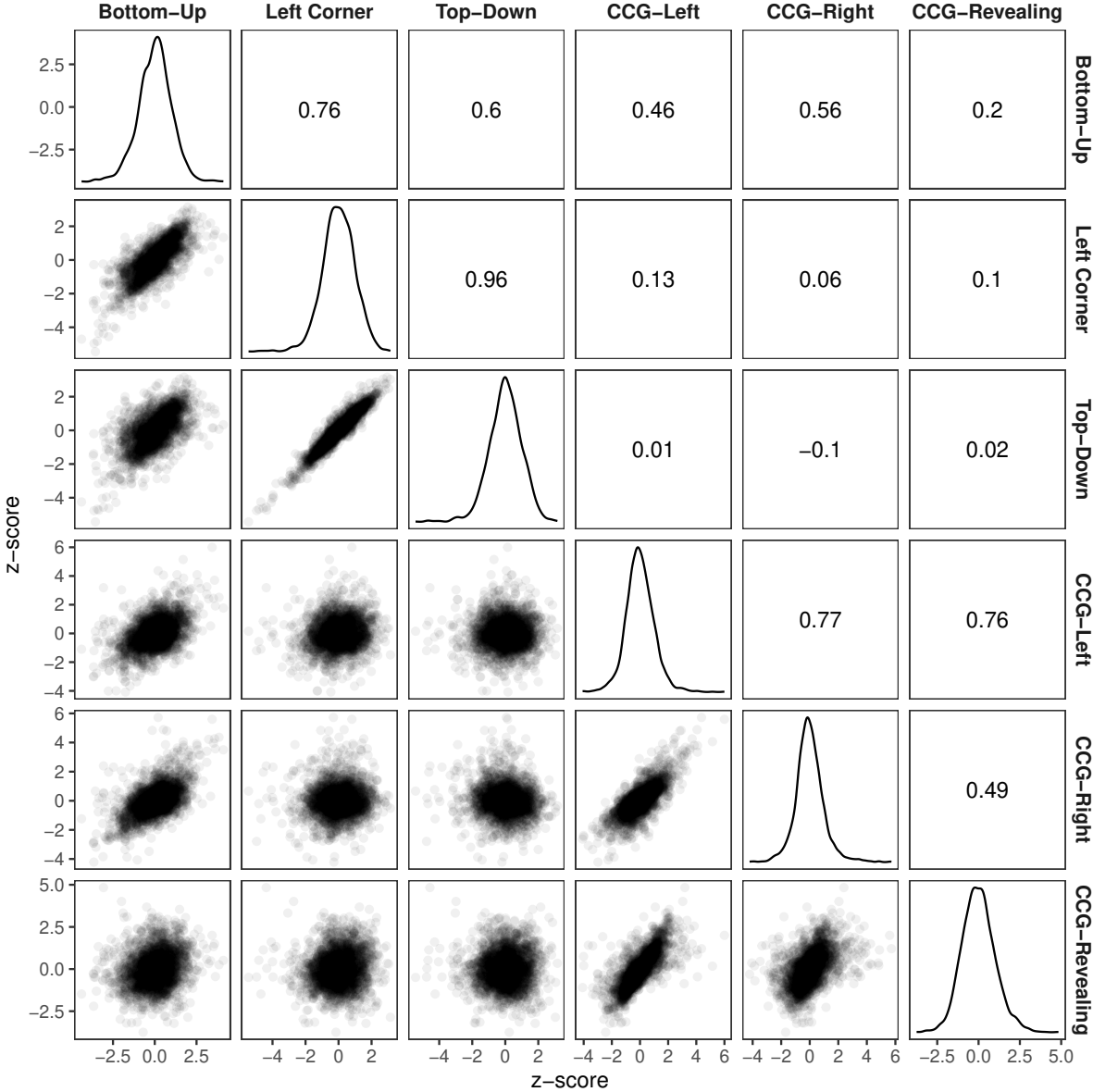


Figure 7: Scatterplot matrix with density plots and Pearson correlation coefficients for phrase structure and CCG derivations.

Hypothesis 1

Model 1: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown$

Model 2: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown + CCG_left + CCG_right + CCG_revealing$

Region	AIC_1^a	AIC_2^b	ΔAIC^c	$\chi^2(3)$	p
IFG_oper	1762175	1762156	-19.26	25.26	<0.001
IFG_orb	1717590	1717579	-10.31	16.31	0.001
IFG_tri	1715945	1715913	-32.10	38.1	<0.001
mATL	1562113	1562092	-20.92	26.92	<0.001
sATL	1604738	1604726	-12.09	18.09	<0.001
STG	1843201	1843194	-7.20	13.2	0.004

Table 1: Hypothesis 1, CCG-specific effects: CCG_left + CCG_right + CCG_revealing. ^aAkaike Information Criterion for the baseline model (model 1). ^bAkaike Information Criterion for model 2. ^c $AIC_2 - AIC_1$. Bonferroni adjusted significance threshold: $0.05/6 = 0.008$.

Hypothesis 1: Follow-up analyses

Model 1: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown$

Model 2: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown + CCG_right$

Region	AIC_1^a	AIC_2^b	ΔAIC^c	$\chi^2(1)$	p
IFG_oper	1762175	1762170	-5.42	7.42	0.007
IFG_orb	1717590	1717589	-1.03	3.03	0.082
IFG_tri	1715945	1715934	-11.25	13.25	<0.001
mATL	1562113	1562115	1.91	0.09	0.770
sATL	1604738	1604739	0.31	1.69	0.193
STG	1843201	1843201	0.51	1.49	0.223

Table 2: Hypothesis 1, CCG-specific effects: CCG_right. ^aAkaike Information Criterion for the baseline model (model 1). ^bAkaike Information Criterion for model 2. ^c $AIC_2 - AIC_1$. Bonferroni adjusted significance threshold: $0.05/6 = 0.008$.

Model 1: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown$

Model 2: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown + CCG_left$

Region	AIC_1^a	AIC_2^b	ΔAIC^c	$\chi^2(1)$	p
IFG_oper	1762175	1762177	1.96	0.04	0.846
IFG_orb	1717590	1717582	-8.12	10.12	0.002
IFG_tri	1715945	1715947	1.81	0.19	0.666
mATL	1562113	1562101	-11.53	13.53	<0.001
sATL	1604738	1604740	1.60	0.4	0.527
STG	1843201	1843192	-8.52	10.52	0.001

Table 3: Hypothesis 1, CCG-specific effects: CCG_left. ^aAkaike Information Criterion for the baseline model (model 1). ^bAkaike Information Criterion for model 2. ^c $AIC_2 - AIC_1$. Bonferroni adjusted significance threshold: $0.05/6 = 0.008$.

Model 1: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown$
 Model 2: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown + CCG_revealing$

Region	AIC_1^a	AIC_2^b	ΔAIC^c	$\chi^2(1)$	p
IFG_oper	1762175	1762173	-2.18	4.18	0.041
IFG_orb	1717590	1717576	-14.00	16	<0.001
IFG_tri	1715945	1715943	-1.77	3.77	0.052
mATL	1562113	1562096	-16.49	18.49	<0.001
sATL	1604738	1604737	-1.67	3.67	0.055
STG	1843201	1843192	-8.51	10.51	0.001

Table 4: Hypothesis 1, CCG-specific effects: CCG_revealing. ^aAkaike Information Criterion for the baseline model (model 1). ^bAkaike Information Criterion for model 2. ^c $AIC_2 - AIC_1$. Bonferroni adjusted significance threshold: $0.05/6 = 0.008$.

Hypothesis 2

Model 1: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown + CCG_left + CCG_right$

Model 2: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown + CCG_left + CCG_right + CCG_revealing$

Region	AIC_1^a	AIC_2^b	ΔAIC^c	$\chi^2(1)$	p
IFG_oper	1762165	1762156	-9.34	11.34	0.001
IFG_orb	1717583	1717579	-4.03	6.03	0.014
IFG_tri	1715922	1715913	-9.68	11.68	0.001
mATL	1562096	1562092	-4.01	6.01	0.014
sATL	1604741	1604726	-14.32	16.32	<0.001
STG	1843193	1843194	0.61	1.39	0.239

Table 5: Hypothesis 2, CCG Revealing operation. ^aAkaike Information Criterion for the baseline model (model 1). ^bAkaike Information Criterion for model 2. ^c $AIC_2 - AIC_1$. Bonferroni adjusted significance threshold: $0.05/6 = 0.008$.

Hypothesis 3

Model 1: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown + CCG_right$

Model 2: $BOLD \sim word_freq + word_rate + RMS + f0 + bottomup + leftcorner + topdown + CCG_right + CCG_left$

Region	AIC_1^a	AIC_2^b	ΔAIC^c	$\chi^2(1)$	p
IFG_oper	1762170	1762165	-4.49	6.49	0.011
IFG_orb	1717589	1717583	-5.25	7.25	0.007
IFG_tri	1715934	1715922	-11.18	13.18	<0.001
mATL	1562115	1562096	-18.82	20.82	<0.001
sATL	1604739	1604741	1.93	0.07	0.788
STG	1843201	1843193	-8.32	10.32	0.001

Table 6: Hypothesis 3, Left- versus Right-CCG parsing. ^aAkaike Information Criterion for the baseline model (model 1). ^bAkaike Information Criterion for model 2. ^c $AIC_2 - AIC_1$. Bonferroni adjusted significance threshold: $0.05/6 = 0.008$.