Sentence Comprehension as a Cognitive Process: A computational modeling approach Day 1: An introduction to sentence comprehension

Shravan Vasishth and Felix Engelmann

Universität Potsdam, Germany
University of Manchester, UK
vasishth@uni-potsdam.de,felix.engelmann@manchester.ac.uk
http://bit.ly/sentcomp

August 15, 2016

What is sentence processing

Two central goals in this field are to understand

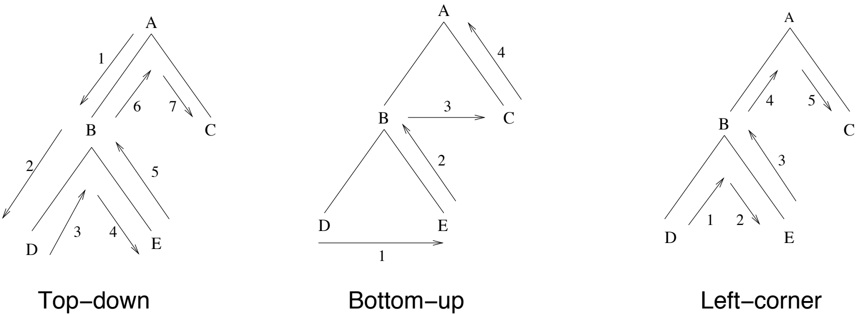
- online **parsing mechanisms** in human sentence comprehension
  - left-corner parsing, top-down, bottom-up? lookahead?
  - probabilistic parsing?
  - serial vs parallel vs ranked parallel?
  - deterministic vs non-deterministic parsing?
  - what kind of information is used to make parsing decisions (syntactic only, syntactic+semantic+...?)
- constraints on **dependency completion**
  - a general preference to attach co-dependents locally
  - constraints on retrieval processes
  - the consequences of probabilistic predictive parsing (expectation effects)
  - "good-enough" processing, underspecification, tracking only local n-grams ("local coherence")

Introduction

- In this course, we will give a fairly narrow perspective on processing sentences out of context.
- We provide an extensive reading list on the course website for further details on the topics we mention.
- These slides also have references at the end.
- Please consult the references on the website and the ones cited in these slides for a fuller picture.

Introduction

Left-corner parsing, probabilistic parsing



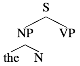
Introduction: parsing mechanisms

Left-corner parsing [1], probabilistic parsing

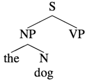
Left-Corner Parsing

S → *NP VP*	*Det* → *a,the*	*NP* → *Det N*
N → *man, dog*	*V* → *ran, saw*	*VP* → *V*
	VP → *V NP*	

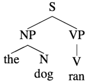
INPUT: *the*
GOAL CATEGORY STACK: [S]
ACTIONS: If *the* is the left corner of any phrase structure rule then replace the stack content with the LHS of that rule. Repeat this left-corner rule until no further steps are possible. Wait for next input word. These actions yield the structure to the right:



INPUT: *dog*
GOAL CATEGORY STACK: [N NP VP S]
ACTIONS: Use the left-corner rule to expand *dog* to N. Since N is predicted in the incremental structure built so far (Step 1), integrate the N built up bottom-up into the tree. Since no further applications of the left-corner rule are possible, wait for the next input.



INPUT: *ran*
GOAL CATEGORY STACK: [VP S]
ACTIONS: Use the left-corner rule to expand *ran* to V, and apply this rule once again to expand to VP. Since a VP is predicted in the structure, integrate this with the tree.



Introduction

Left-corner parsing, probabilistic parsing

Purely top-down or purely bottom-up strategies turn out to be inappropriate models for human parsing [2, 3, 4] since they are unable to capture the observation [5, 468-470] that left-branching and right-branching structures are relatively easy to process compared to center embeddings:

(1)
a. Bill's book's cover is dirty.
b. Bill has the book that has the cover that is dirty.
c. The rat the cat the dog chased killed ate the malt.

Introduction

Left-corner parsing, probabilistic parsing

More frequent attachments are preferred over rare attachments [6].

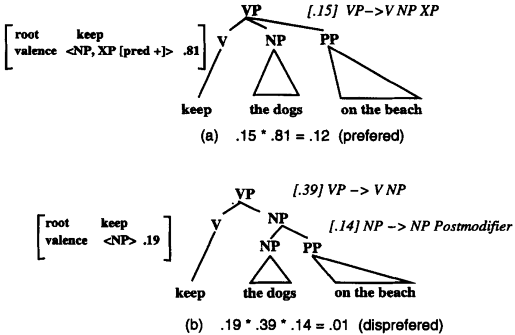
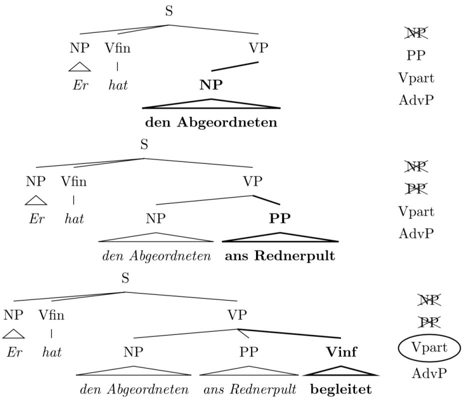


Figure 8. Annotated Parse Trees for Two Interpretations of *keep the dogs on the beach*

Introduction

Left-corner parsing, probabilistic parsing

Expectations for an upcoming verb phrase are sharpened if the verb's appearance is delayed [7].



Introduction: parsing mechanisms

serial / parallel / ranked parallel

A general assumption in most work today is that parse choices are strictly serial. But theoretically, other options are possible, and there is some evidence for ranked parallelism [8].

- Serial: compute a single analysis, and if that fails, backtrack and compute new analysis (most classical theories, e.g., [9, 10, 11]).
- Parallel:
  - Ranked: Compute all analyses in parallel, but rank them (e.g. by likelihood).
  - Prune: using, e.g., beam search.
  - Don't prune at all—generate all possible structures and then compute a function over them (e.g. entropy reduction, or surprisal) to find the optimal one [12, 13].

Introduction

deterministic / non-deterministic

- A common early assumption was that parsing was essentially deterministic.
- A heuristic is to always prefer to attach locally [11]. Example:

(2)
a. (low attachment)
The car of the **driver** *that had the moustache* was pretty cool.

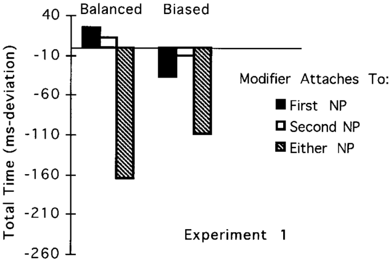
b. (high attachment)
The **driver** of the car *that had the moustache* was pretty cool.

c. (globally ambiguous)
The **son** of the **driver** *that had the moustache* was pretty cool.

- Prediction: 2a,c easier to process than 2b.

Introduction

deterministic / non-deterministic

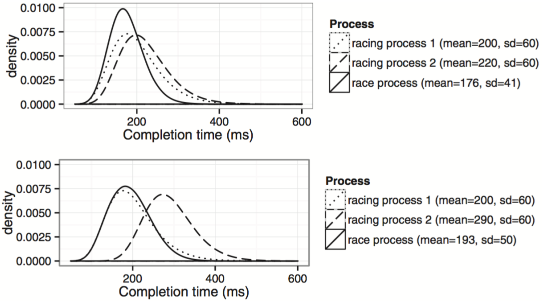


[14] found/claimed that the word *moustache* was read fastest in the globally ambiguous sentence: the **ambiguity advantage**.

Introduction

deterministic / non-deterministic

One explanation [15] for this is to assume a non-deterministic race process (also see [16]):



Introduction: parsing mechanisms

information sources: syntax only / all sources of information

[17] found evidence against syntax-first proposals, but [18] found evidence for syntax-first. (*A too-common example of how prior beliefs of researchers are, uncannily, always magically confirmed*.)

(3)
a. The defendant examined by the lawyer turned out to be unreliable.
b. The evidence examined by the lawyer turned out to be unreliable.

Introduction: constraints on dependency completion

A local attachment preference

Non-local dependency completion tends to be more difficult than local dependency completion [19, 20].

$$\text{The nurse}\left\{\begin{array}{c}\emptyset\\ \text{from the clinic}\\ \text{who was from the clinic}\end{array}\right\}\text{supervised the administrator}\ldots$$

Introduction: constraints on dependency completion

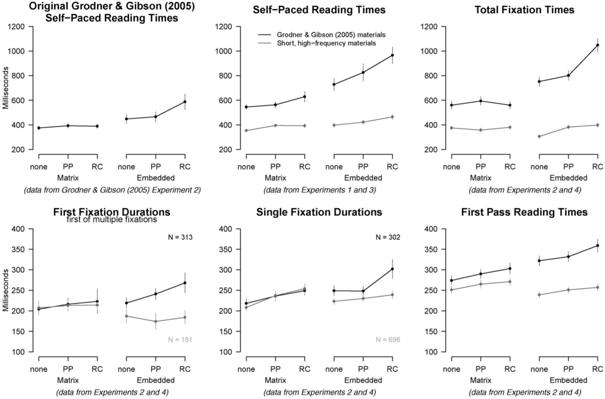
A local attachment preference

(4)
a. The administrator who the **nurse supervised** scolded the medic while ...
b. The administrator who the **nurse** from the clinic **supervised** scolded the medic while ...
c. The administrator who the **nurse** who was from the clinic **supervised** scolded the medic while ...

Introduction: constraints on dependency completion

A local attachment preference

Source: [20].



Introduction: constraints on dependency completion

Good-Enough processing / underspecification / local coherence

Source: [21]

(5) a. The coach smiled at the player who was tossed a frisbee

b. The coach smiled at the player ~~who was~~ tossed a frisbee

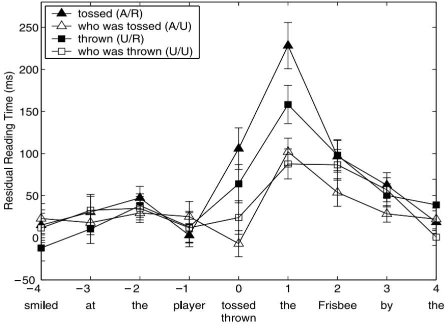
Subjects seem to treat

(6) "the player tossed a frisbee"

as a main clause.

Introduction: constraints on dependency completion

Good-Enough processing / underspecification / local coherence



Introduction: constraints on dependency completion

Uncertainty increases with argument-verb distance (Safavi et al 2016)

(4) a. Strong predictability, short distance (PP)

Ali a:rezouyee bara:ye man kard va…
Ali wish-INDEF for 1.S do-PST and…

'Ali made a wish for me and…'

b. Strong predictability, long distance (RC+PP)

Ali a:rezouyee ke besya:r doost-da:sht-am
Ali wish-INDEF that a lot like-1.S-PST
bara:ye man kard va…
for 1.S do-PST and…

'Ali made a wish that I liked a lot for me and…'

c. Weak predictability, short distance (PP)

Ali shokola:ti bara:ye man xarid va…
Ali chocolate-INDEF for 1.S buy-PST and…

'Ali bought a chocolate for me and …'

d. Weak predictability, long distance (RC+PP)

Ali shokola:ti ke besya:r doost-da:sht-am
Ali chocolate-INDEF that a lot like-1.S-PST
bara:ye man xarid va…
for 1.S buy-PST and…

'Ali bought a chocolate that I liked a lot for me and…'

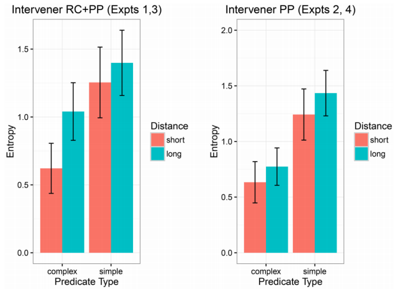


FIGURE 6 | The estimated entropy (with 95% confidence intervals), computed using the sentence completion data, for the two experiment designs.

Introduction: constraints on dependency completion

Uncertainty increases with argument-verb distance (Safavi et al 2016)

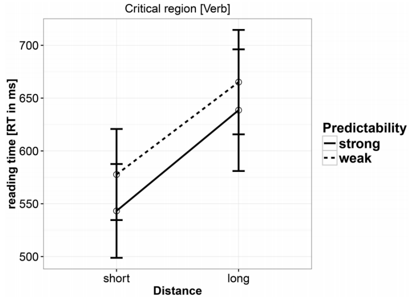


FIGURE 2 | Reading times at the critical verb in Experiment 2.

Introduction: constraints on dependency completion

Constraints on retrieval

Similarity-based interference has been implicated as a cause for difficulty in completing subject-verb dependencies.
The essential idea is that retrieving an item (e.g., a noun) is harder (e.g., at a verb) if there are other competing items present that are similar on some dimension.

An implementation of this idea is Lewis and Vasishth (2005) (henceforth LV05), which is the subject of this course.

The model assumptions

This is often called "the" cue based model, but there are many cue-based models (Van Dyke's, McElree's conceptions are different from the LV05 model).

1. **Grammatical knowledge and left-corner parsing algorithm**:
   Note that a parser can do nothing without a grammar. So even asking a question like "is it the grammar or the parser?" technically doesn't even mean anything.
   - If-then production rules drive structure building
   - Rules are hand-crafted in toy models, but scaling up has been done (Boston, Hale, Kliegl, Vasishth, Lang Cog Proc 2011).
2. **Constraints on memory processes affecting retrieval**:
   allows us to model individual differences in attention and working memory capacity

Retrieval at any dependency completion point is a key (but not only) determinant of processing difficulty or facilitation.

Introduction and background

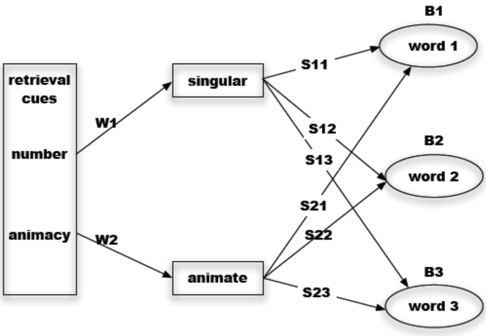
The memory constraints in the model

Code:
https://github.com/felixengelmann/act-r-sentence-parser-em

Latency factor F (:lf) **—> Speed**	$RT = Fe^{-(f * A_i)}$
Decay parameter d (:bll) **—> Speed, forgetting**	$B_i = \ln(\sum_{j=1}^{n} t_j^{-d}) + \beta_i$
Source activation W_k of buffer k (e.g., goalbuffer :ga) This activation is distributed among goal-related chunks. **—> Accuracy (goal-relevant), speed**	$A_i = B_i + S_i + P_i + \varepsilon_i$ $S_i = \sum_k \sum_j W_{kj} S_{ji}$
Mismatch penalty P (:mp) **—> Error sensitivity**	$P_i = \sum_k PM_{ki}$
Similarity M_{ki} between the value k in the retrieval specification and the value in the corresponding slot of chunk i **—> Association between cue and target**	

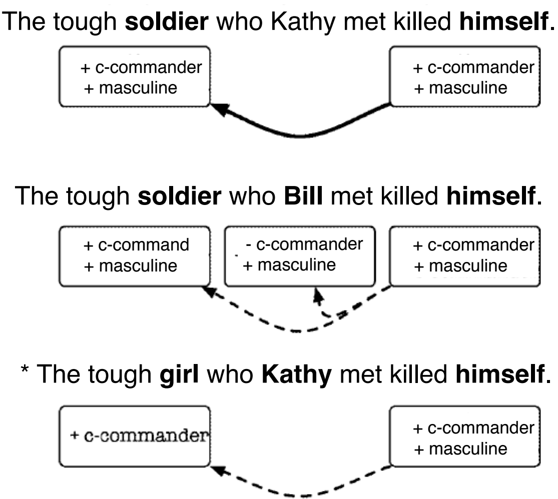
Introduction and background

The memory constraints in the model: Similarity based interference



Introduction and background

The memory constraints in the model: Partial Matching



Introduction and background

Possible evidence for partial matching: Processing polarity ([23] cf. [24, 25, 22])

Source: [22]

(7) a. No diplomats that a congressman would trust have ever supported a drone strike.
b. *The diplomats that no congressman could trust have ever supported a drone strike
c. *The diplomats that a congressman would trust have ever supported a drone strike.

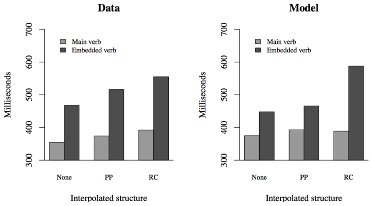
Condition	Data	Model
(7a) Accessible licensor	85	96
(7b) Inaccessible licensor	**70**	**61**
(7c) No licensor	83	86

Introduction: constraints on dependency completion

Constraints on retrieval

Consider again the Grodner and Gibson 05 results and our model [1] results:

$$\text{The nurse} \left\{ \begin{array}{c} \emptyset \\ \text{from the clinic} \\ \text{who was from the clinic} \end{array} \right\} \text{supervised the administrator} \ldots$$



Introduction: constraints on dependency completion

Lewis & Vasishth 2005, Engelmann, Jäger, Vasishth 2016

(8) a. *Target-match; distractor-mismatch*
The surgeon$^{+masc}_{+c\text{-}com}$ who treated Jennifer$^{-masc}_{-c\text{-}com}$ had pricked himself$\{^{masc}_{c\text{-}com}\}$...

b. *Target-match; distractor-match*
The surgeon$^{+masc}_{+c\text{-}com}$ who treated Jonathan$^{+masc}_{-c\text{-}com}$ had pricked himself$\{^{masc}_{c\text{-}com}\}$...

Modeling retrieval processes in sentence comprehension

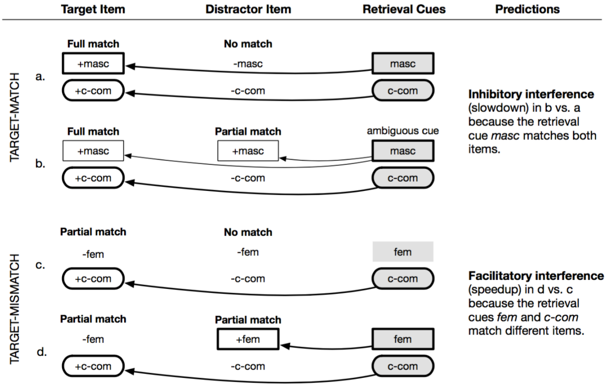
Lewis & Vasishth 2005, Engelmann, Jäger, Vasishth 2016

(9) a. *Target-mismatch; distractor-mismatch*
The surgeon$^{-fem}_{+\,c\text{-}com}$ who treated Jonathan$^{-fem}_{-\,c\text{-}com}$ had pricked herself$\{^{fem}_{c\text{-}com}\}$...

b. *Target-mismatch; distractor-match*
The surgeon$^{-fem}_{+\,c\text{-}com}$ who treated Jennifer$^{+fem}_{-\,c\text{-}com}$ had pricked herself$\{^{fem}_{c\text{-}com}\}$...

Modeling retrieval processes in parsing

Lewis & Vasishth 2005, Engelmann, Jäger, Vasishth 2016



Modeling retrieval processes in parsing

Lewis & Vasishth 2005, Engelmann, Jäger, Vasishth 2016

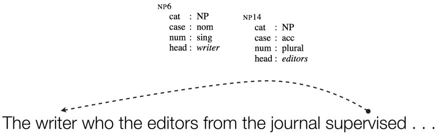
Agreement attraction could also be an instance of similarity-based interference:

(10) a. The key$_{+sing}$ to the cabinet$_{+sing}$ is in the box.
b. The key$_{+sing}$ to the cabinets$_{+plur}$ is in the box.
c. * The key$_{+sing}$ to the cabinet$_{+sing}$ are in the box.
d. * The key$_{+sing}$ to the cabinets$_{+plur}$ are in the box.

Modeling retrieval processes in parsing

Lewis & Vasishth 2005, Engelmann, Jäger, Vasishth 2016

Lewis & Vasishth (2005) Parser



- Activation decay —> distance effects
- Associative retrieval —> Similarity-based interference
- Deterministic rule application —> Expectation effects, reanalysis

Modeling retrieval processes in parsing

Engelmann, Jäger, Vasishth 2016

[1] Richard L. Lewis and Shravan Vasishth.
An activation-based model of sentence processing as skilled memory retrieval.
Cognitive Science, 29:1–45, May 2005.

[2] P. N. Johnson-Laird.
Mental Models: Towards a cognitive science of language, inference, and consciousness.
Harvard University Press, Cambridge, MA, 1983.

[3] Philip Resnik.
Left–corner parsing and psychological plausibility.
In Proceedings of COLING, pages 191–197, address not known, 1992.

[4] S. Abney and M. Johnson.
Memory requirements and local ambiguities of parsing strategies.
Journal of Psycholinguistic Research, 20(3):233–250, 1991.

[5] George Miller and Noam Chomsky.
Finitary models of language users.
In R. Duncan Luce, Robert R. Bush, and Eugene Galanter, editors, Handbook of Mathematical Psychology, Volume II, pages 419–492. John Wiley, 1963.

[6] Daniel Jurafsky.
A probabilistic model of lexical and syntactic access and disambiguation.
Cognition, 20:137–194, 1996.

[7] Roger Levy.
Expectation-based syntactic comprehension.
Cognition, 106:1126–1177, 2008.

[8] Marisa F. Boston, John T. Hale, Shravan Vasishth, and Reinhold Kliegl.
Parallel processing and sentence comprehension difficulty.
Language and Cognitive Processes, 26(3):301–349, 2011.

[9] Lyn Frazier.
On Comprehending Sentences: Syntactic Parsing Strategies.
PhD thesis, University of Massachusetts, Amherst, 1979.

[10] Lyn Frazier and Keith Rayner.
Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences.
Cognitive Psychology, 14:178–210, 1982.

[11] L. Frazier.
Syntactic complexity.
In David R. Dowty, Lauri Kartunnen, and Arnold M. Zwicky, editors, Natural language parsing: Psychological, Computational, and Theoretical perspectives, pages 129–189. Cambridge University Press, 1985.

[12] John T. Hale.
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, Pittsburgh, PA, 2001.

[13] John Hale.
The information conveyed by words in sentences.
Journal of Psycholinguistic Research, 32(2):101–123, March 2003.

[14] Matthew J Traxler, Martin J Pickering, and Charles Clifton.
Adjunct attachment is not a form of lexical ambiguity resolution.
Journal of Memory and Language, 39(4):558–592, 1998.

[15] Roger P.G. van Gompel, Martin J. Pickering, and Matthew J. Traxler.
Unrestricted race: A new model of syntactic ambiguity resolution.
In A Kennedy, R Radach, D Heller, and J Pynte, editors, Reading as a perceptual process. Elsevier, Oxford, 2000.

[16] Pavel Logačev and Shravan Vasishth.
A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension.
Cognitive Science, 2015.

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

[18] C. Clifton, B. Juhasz, J. Ashby, M. J. Traxler, M. T. Mohamed, R. S. Williams, R. K. Morris, and K. Rayner.
The use of thematic role information in parsing: Syntactic processing autonomy revisited.
Journal of Memory and Language, 49:317–334, 2003.

[19] Daniel Grodner and Edward Gibson.
Consequences of the serial nature of linguistic input.
Cognitive Science, 29:261–290, 2005.

[20] Brian Bartek, Richard L. Lewis, Shravan Vasishth, and Mason Smith.
In search of on-line locality effects in sentence comprehension.
Journal of Experimental Psychology: Learning, Memory and Cognition, 37(5):1178–1198, 2011.

[21] Whitney Tabor, Bruno Galantucci, and Daniel Richardson.
Effects of merely local syntactic coherence on sentence processing.
Journal of Memory and Language, 50:355–370, 2004.

[22] Dan Parker and Colin Phillips.
Negative polarity illusions and the format of hierarchical encodings in memory.
Under review with Cognition, 2016.

[23] Shravan Vasishth, Sven Bruessow, Richard L. Lewis, and Heiner Drenhaus.
Processing polarity: How the ungrammatical intrudes on the grammatical.
Cognitive Science, 32(4), 2008.

[24] M. Xiang, B. Dillon, and C. Phillips.
Illusory licensing effects across dependency types: ERP evidence.
Brain and Language, 108(1):40–55, 2009.

[25] Brian Dillon, Alan Mishler, Shayne Sloggett, and Colin Phillips.
A computational cognitive model of syntactic priming.
Journal of Memory and Language, 69(4):85–103, 2011.