Retrieving structures from memory causes difficulty during incremental processing

Cory Shain[1], Marten van Schijndel[1], Richard Futrell[2], Edward Gibson[2], and William Schuler[1]

[1]Dept of Linguistics, The Ohio State University
[2]Dept of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Based on Shain et al. 2016

1 Apr. 2017, *CUNY 2017*

Introduction

- **Hypothesis:** Memory access in incremental parsing causes reading time delays.
- Predicted by a number of theories of sentence processing (e.g. Gibson 2000; Johnson-Laird 1983).
- Supported by numerous previous studies (e.g. Grodner and Gibson 2005; Boston et al. 2011; von der Malsburg, Kliegl, and Vasishth 2015).

Introduction

- **Hypothesis:** Memory access in incremental parsing causes reading time delays.
- Predicted by a number of theories of sentence processing (e.g. Gibson 2000; Johnson-Laird 1983).
- Supported by numerous previous studies (e.g. Grodner and Gibson 2005; Boston et al. 2011; von der Malsburg, Kliegl, and Vasishth 2015).

Introduction

- **Hypothesis:** Memory access in incremental parsing causes reading time delays.
- Predicted by a number of theories of sentence processing (e.g. Gibson 2000; Johnson-Laird 1983).
- Supported by numerous previous studies (e.g. Grodner and Gibson 2005; Boston et al. 2011; von der Malsburg, Kliegl, and Vasishth 2015).

Conclusion

- Memory retrieval effects exist in human sentence processing.

Acknowledgments

This work was supported by grants from the National Science Foundation: Graduate Research Fellowship Program Award DGE-1343012 to MvS; Doctoral Dissertation Research Improvement Award 1551543 to RF; Linguistics Program Award 1534318 to EG.

Introduction

So why are we still working on this?

Introduction

- **Previous findings:** Mixed.
  - **Constructed stimuli:** Strong effects
  - **Naturally-occurring stimuli:** Weak/null/negative effects

Introduction

- **Previous findings:** Mixed.
  - **Constructed stimuli:** Strong effects
  - **Naturally-occurring stimuli:** Weak/null/negative effects

Introduction

- **Previous findings:** Mixed.
  - **Constructed stimuli:** Strong effects
  - **Naturally-occurring stimuli:** Weak/null/negative effects

Introduction

- Potential weaknesses in each type of stimulus
  - **Constructed:**
    - Limited domain (e.g. relative clauses)
  - **Naturally-occurring:**

Introduction

- Potential weaknesses in each type of stimulus
  - **Constructed:**
    - Limited domain (e.g. relative clauses)
    - Confounds from lack of context or semantic strangeness
    - Lack of information theoretic controls (e.g. surprisal)
  - **Naturally-occurring:**
    - Limited number of subjects (10 in Dundee)
    - 'Easy' stimuli

Introduction

- Potential weaknesses in each type of stimulus
  - **Constructed:**
    - Limited domain (e.g. relative clauses)
    - Confounds from lack of context or semantic strangeness
    - Lack of information theoretic controls (e.g. surprisal)
  - **Naturally-occurring:**
    - Limited number of subjects (10 in Dundee)
    - 'Easy' stimuli

Introduction

- Potential weaknesses in each type of stimulus
  - **Constructed:**
    - Limited domain (e.g. relative clauses)
    - Confounds from lack of context or semantic strangeness
    - Lack of information theoretic controls (e.g. surprisal)
  - **Naturally-occurring:**
    - Limited number of subjects (10 in Dundee)
    - 'Easy' stimuli

Introduction

- Potential weaknesses in each type of stimulus
  - **Constructed:**
    - Limited domain (e.g. relative clauses)
    - Confounds from lack of context or semantic strangeness
    - Lack of information theoretic controls (e.g. surprisal)
  - **Naturally-occurring:**
    - Limited number of subjects (10 in Dundee)
    - 'Easy' stimuli

Introduction

- Potential weaknesses in each type of stimulus
  - **Constructed:**
    - Limited domain (e.g. relative clauses)
    - Confounds from lack of context or semantic strangeness
    - Lack of information theoretic controls (e.g. surprisal)
  - **Naturally-occurring:**
    - Limited number of subjects (10 in Dundee)
    - ‘Easy’ stimuli

Introduction

- Potential weaknesses in each type of stimulus
  - **Constructed:**
    - Limited domain (e.g. relative clauses)
    - Confounds from lack of context or semantic strangeness
    - Lack of information theoretic controls (e.g. surprisal)
  - **Naturally-occurring:**
    - Limited number of subjects (10 in Dundee)
    - 'Easy' stimuli

Introduction

- Potential weaknesses in each type of stimulus
  - **Constructed:**
    - Limited domain (e.g. relative clauses)
    - Confounds from lack of context or semantic strangeness
    - Lack of information theoretic controls (e.g. surprisal)
  - **Naturally-occurring:**
    - Limited number of subjects (10 in Dundee)
    - ‘Easy’ stimuli

Introduction

- Different theories of sentence processing
  - **The Dependency Locality Theory (DLT):** Cost as function of dependency length.
  - **Left-corner parsing:** Cost as function of parser operations that involve memory retrieval.
- Many plausible implementations of the cost function

Introduction

- Different theories of sentence processing
  - **The Dependency Locality Theory (DLT):** Cost as function of dependency length.
  - **Left-corner parsing:** Cost as function of parser operations that involve memory retrieval.
- Many plausible implementations of the cost function

Introduction

- Different theories of sentence processing
  - **The Dependency Locality Theory (DLT):** Cost as function of dependency length.
  - **Left-corner parsing:** Cost as function of parser operations that involve memory retrieval.
- Many plausible implementations of the cost function

Introduction

- Different theories of sentence processing
  - **The Dependency Locality Theory (DLT):** Cost as function of dependency length.
  - **Left-corner parsing:** Cost as function of parser operations that involve memory retrieval.
- Many plausible implementations of the cost function

Introduction

- This work:
  - 'Constructed-natural' stimuli (Natural Stories corpus) (Futrell et al. in prep).
  - Predictor selection on exploratory data.
  - Baseline information theoretic controls.

Introduction

- This work:
  - ‘Constructed-natural’ stimuli (Natural Stories corpus) (Futrell et al. in prep).
  - Predictor selection on exploratory data.
  - Baseline information theoretic controls.

Introduction

- This work:
  - 'Constructed-natural' stimuli (Natural Stories corpus) (Futrell et al. in prep).
  - Predictor selection on exploratory data.
  - Baseline information theoretic controls.

Introduction

- This work:
  - ‘Constructed-natural’ stimuli (Natural Stories corpus) (Futrell et al. in prep).
  - Predictor selection on exploratory data.
  - Baseline information theoretic controls.

Introduction

- We find strong broad-coverage evidence in self-paced reading (SPR) of:
  - Locality-based DLT effect.
  - Locality-independent left-corner effect.

Introduction

- We find strong broad-coverage evidence in self-paced reading (SPR) of:
  - Locality-based DLT effect.
  - Locality-independent left-corner effect.

Introduction

- We find strong broad-coverage evidence in self-paced reading (SPR) of:
  - Locality-based DLT effect.
  - Locality-independent left-corner effect.

Plan

Introduction

Experimental setup

The Dependency Locality Theory

Left-corner parsing

Conclusion

Plan

Experimental setup: The Natural Stories corpus

- Experiments used Natural Stories corpus (Futrell et al. in prep):
  - Constructed-natural stimuli:
    - Fluent narrative text
    - Memory-intensive constructions
  - SPR data collected
  - 10 texts, 10257 words, 181 subjects, 848,207 events
- Partitioned into exploratory (1/3) and confirmatory (2/3) corpora
- First evaluated all predictors on exploratory using LME regression.
- All results are in spillover 1 position.

Experimental setup: The Natural Stories corpus

- Experiments used Natural Stories corpus (Futrell et al. in prep):
  - Constructed-natural stimuli:
    - Fluent narrative text
    - Memory-intensive constructions
  - SPR data collected
  - 10 texts, 10257 words, 181 subjects, 848,207 events
- Partitioned into exploratory (1/3) and confirmatory (2/3) corpora
- First evaluated all predictors on exploratory using LME regression.
- All results are in spillover 1 position.

Experimental setup: The Natural Stories corpus

- Experiments used Natural Stories corpus (Futrell et al. in prep):
  - Constructed-natural stimuli:
    - Fluent narrative text
    - Memory-intensive constructions
  - SPR data collected
  - 10 texts, 10257 words, 181 subjects, 848,207 events
- Partitioned into exploratory (1/3) and confirmatory (2/3) corpora
- First evaluated all predictors on exploratory using LME regression.
- All results are in spillover 1 position.

Experimental setup: The Natural Stories corpus

- Experiments used Natural Stories corpus (Futrell et al. in prep):
  - Constructed-natural stimuli:
    - Fluent narrative text
    - Memory-intensive constructions
  - SPR data collected
  - 10 texts, 10257 words, 181 subjects, 848,207 events
- Partitioned into exploratory (1/3) and confirmatory (2/3) corpora
- First evaluated all predictors on exploratory using LME regression.
- All results are in spillover 1 position.

Experimental setup: The Natural Stories corpus

- Experiments used Natural Stories corpus (Futrell et al. in prep):
  - Constructed-natural stimuli:
    - Fluent narrative text
    - Memory-intensive constructions
  - SPR data collected
  - 10 texts, 10257 words, 181 subjects, 848,207 events
- Partitioned into exploratory (1/3) and confirmatory (2/3) corpora
- First evaluated all predictors on exploratory using LME regression.
- All results are in spillover 1 position.

Experimental setup: The Natural Stories corpus

- Experiments used Natural Stories corpus (Futrell et al. in prep):
  - Constructed-natural stimuli:
    - Fluent narrative text
    - Memory-intensive constructions
  - SPR data collected
  - 10 texts, 10257 words, 181 subjects, 848,207 events
- Partitioned into exploratory (1/3) and confirmatory (2/3) corpora
- First evaluated all predictors on exploratory using LME regression.
- All results are in spillover 1 position.

Experimental setup: The Natural Stories corpus

- Experiments used Natural Stories corpus (Futrell et al. in prep):
  - Constructed-natural stimuli:
    - Fluent narrative text
    - Memory-intensive constructions
  - SPR data collected
  - 10 texts, 10257 words, 181 subjects, 848,207 events
- Partitioned into exploratory (1/3) and confirmatory (2/3) corpora
- First evaluated all predictors on exploratory using LME regression.
- All results are in spillover 1 position.

Experimental setup: The Natural Stories corpus

- Experiments used Natural Stories corpus (Futrell et al. in prep):
  - Constructed-natural stimuli:
    - Fluent narrative text
    - Memory-intensive constructions
  - SPR data collected
  - 10 texts, 10257 words, 181 subjects, 848,207 events
- Partitioned into exploratory (1/3) and confirmatory (2/3) corpora
- First evaluated all predictors on exploratory using LME regression.
- All results are in spillover 1 position.

Experimental setup: The Natural Stories corpus

- Experiments used Natural Stories corpus (Futrell et al. in prep):
  - Constructed-natural stimuli:
    - Fluent narrative text
    - Memory-intensive constructions
  - SPR data collected
  - 10 texts, 10257 words, 181 subjects, 848,207 events
- Partitioned into exploratory (1/3) and confirmatory (2/3) corpora
- First evaluated all predictors on exploratory using LME regression.
- All results are in spillover 1 position.

Plan

The Dependency Locality Theory: Background

- DLT (Gibson 2000): Difficulty by dependency length (# discourse referents (DR))
- Intuition: Older things are harder to access in memory
- For simplicity, DR = nouns and finite verbs
- **Integration cost** is sum of:
  - Discourse cost: 1 for nouns/verbs, 0 otherwise
  - Dependency length: Length in DR of all dependencies with preceding words

The Dependency Locality Theory: Background

- DLT (Gibson 2000): Difficulty by dependency length (# discourse referents (DR))
- Intuition: Older things are harder to access in memory
- For simplicity, DR = nouns and finite verbs
- **Integration cost** is sum of:
  - Discourse cost: 1 for nouns/verbs, 0 otherwise
  - Dependency length: Length in DR of all dependencies with preceding words

The Dependency Locality Theory: Background

- DLT (Gibson 2000): Difficulty by dependency length (# discourse referents (DR))
- Intuition: Older things are harder to access in memory
- For simplicity, DR = nouns and finite verbs
- **Integration cost** is sum of:
  - **Discourse cost:** 1 for nouns/verbs, 0 otherwise.
  - **Dependency length:** Length in DR of all dependencies

The Dependency Locality Theory: Background

- DLT (Gibson 2000): Difficulty by dependency length (# discourse referents (DR))
- Intuition: Older things are harder to access in memory
- For simplicity, DR = nouns and finite verbs
- **Integration cost** is sum of:
  - **Discourse cost:** 1 for nouns/verbs, 0 otherwise.
  - **Dependency length:** Length in DR of all dependencies with preceding words

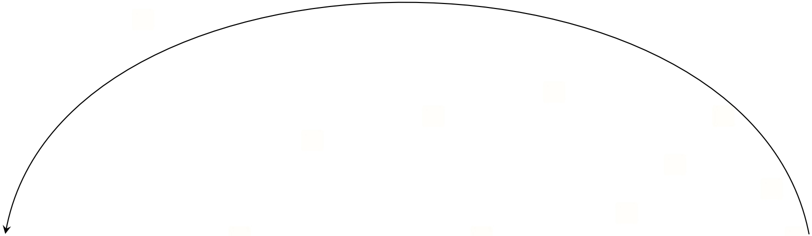
The Dependency Locality Theory: Background

- DLT (Gibson 2000): Difficulty by dependency length (# discourse referents (DR))
- Intuition: Older things are harder to access in memory
- For simplicity, DR = nouns and finite verbs
- **Integration cost** is sum of:
  - **Discourse cost:** 1 for nouns/verbs, 0 otherwise.
  - **Dependency length:** Length in DR of all dependencies with preceding words

The Dependency Locality Theory: Background

- DLT (Gibson 2000): Difficulty by dependency length (# discourse referents (DR))
- Intuition: Older things are harder to access in memory
- For simplicity, DR = nouns and finite verbs
- **Integration cost** is sum of:
  - **Discourse cost:** 1 for nouns/verbs, 0 otherwise.
  - **Dependency length:** Length in DR of all dependencies with preceding words

The Dependency Locality Theory: Example



Yesterday, the **person supervisors** and **co-workers caught** stealing **millions fled.**

Dependency length = 4 (4 intervening DR (bold))

The Dependency Locality Theory: Results

The Dependency Locality Theory: Results

	β (ms)	p-value
DLT	0.466	1.11e-05

Reliable broad-coverage DLT effect

The Dependency Locality Theory: Results

	β (ms)	*p*-value
DLT	0.466	1.11e-05
DLT (modified)	1.13	4.87e-10

Reliable broad-coverage DLT effect

Plan

Left-corner parsing

- Popular model of sentence processing
- Rosenkrantz and Lewis (1970), Johnson-Laird (1983), Abney and Johnson (1991), Gibson (1991), Resnik (1992), Stabler (1994), Lewis and Vasishth (2005), and van Schijndel, Exley, and Schuler (2013)

Left-corner parsing

- Popular model of sentence processing
- Rosenkrantz and Lewis (1970), Johnson-Laird (1983), Abney and Johnson (1991), Gibson (1991), Resnik (1992), Stabler (1994), Lewis and Vasishth (2005), and van Schijndel, Exley, and Schuler (2013)

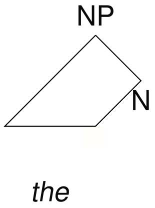
Left-corner parsing

- Maintains a store of derivation fragments $a/b, a'/b', \ldots$, each consisting of active category a lacking awaited category b.
- Incrementally assembles trees by forking/joining fragments.

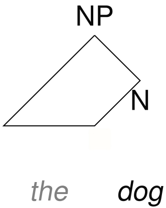
Left-corner parsing

- Maintains a store of derivation fragments $a/b, a'/b', \ldots$, each consisting of active category a lacking awaited category b.
- Incrementally assembles trees by forking/joining fragments.

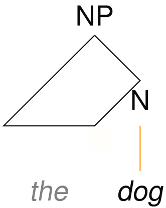
Left-corner parsing: Example



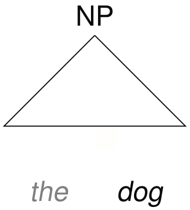
Left-corner parsing: Example



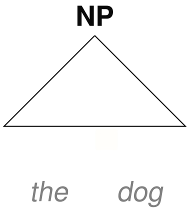
Left-corner parsing: Example



Left-corner parsing: Example



Left-corner parsing: Example



Left-corner parsing

- Many plausible predictor implementations (see Shain et al. 2016)
- Best predictor on exploratory: “no fork”
  - Top sign must be recalled from memory
  - Flags right edges of constituents
  - Boolean (locality-independent)

Left-corner parsing

- Many plausible predictor implementations (see Shain et al. 2016)
- Best predictor on exploratory: “no fork”
  - Top sign must be recalled from memory
  - Flags right edges of constituents
  - Boolean (locality-independent)

Left-corner parsing

- Many plausible predictor implementations (see Shain et al. 2016)
- Best predictor on exploratory: “no fork”
  - Top sign must be recalled from memory
  - Flags right edges of constituents
  - Boolean (locality-independent)

Left-corner parsing

- Many plausible predictor implementations (see Shain et al. 2016)
- Best predictor on exploratory: “no fork”
  - Top sign must be recalled from memory
  - Flags right edges of constituents
  - Boolean (locality-independent)

Left-corner parsing

- Many plausible predictor implementations (see Shain et al. 2016)
- Best predictor on exploratory: “no fork”
  - Top sign must be recalled from memory
  - Flags right edges of constituents
  - Boolean (locality-independent)

Left-corner parsing: Results

	β (ms)	*p*-value
DLT	0.466	1.11e-05
DLT (modified)	1.13	4.87e-10
Left corner (no-fork)	3.88	2.33e-14

Reliable broad-coverage left corner effect

Left-corner parsing: Results

- No-fork estimate (3.88ms) is entire effect.
- DLT effect has a large range (12) but is usually small (95th percentile = 2).
- No-fork effect estimate larger for most events, smaller for large DLT.

Left-corner parsing: Results

- No-fork estimate (3.88ms) is entire effect.
- DLT effect has a large range (12) but is usually small (95^{th} percentile = 2).
- No-fork effect estimate larger for most events, smaller for large DLT.

Left-corner parsing: Results

- No-fork estimate (3.88ms) is entire effect.
- DLT effect has a large range (12) but is usually small (95^{th} percentile = 2).
- No-fork effect estimate larger for most events, smaller for large DLT.

The DLT vs. left-corner

- We seem to have effects from both processing models.
- These models are often taken to be competing.
- Maybe they're measuring the same thing...

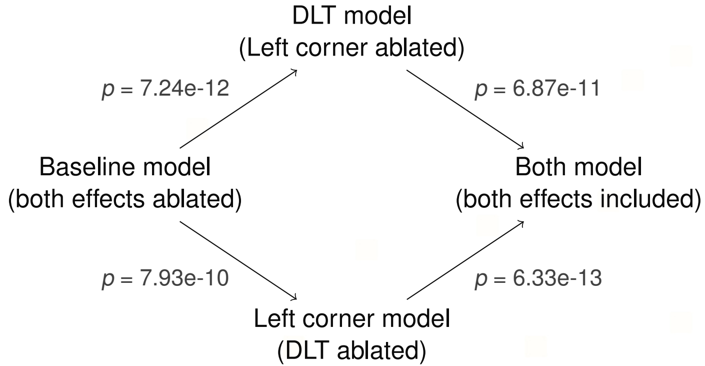
The DLT vs. left-corner

- We seem to have effects from both processing models.
- These models are often taken to be competing.
- Maybe they're measuring the same thing...

The DLT vs. left-corner

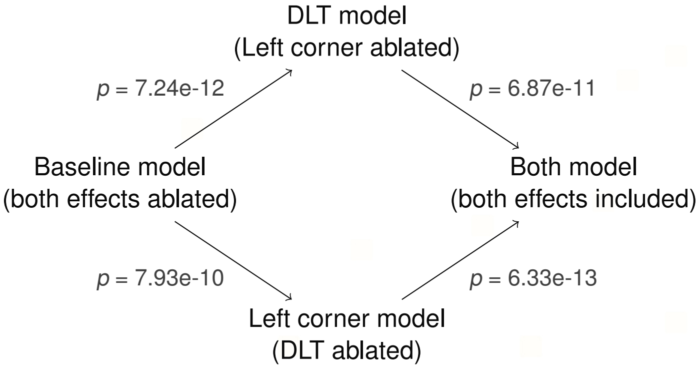
- We seem to have effects from both processing models.
- These models are often taken to be competing.
- Maybe they’re measuring the same thing...

The DLT vs. left-corner



- DLT and left-corner effects are independent via 'diamond' LRT.
- Both effects improve significantly over baseline and over each other.

The DLT vs. left-corner



- DLT and left-corner effects are independent via ‘diamond’ LRT.
- Both effects improve significantly over baseline and over each other.

How general are these effects?

Are these effects widespread, or are they driven by particular contexts?

How general are these effects?

	β-ms	p-value
DLT (modified) (N/V removed)	0.353	0.141
Left corner (no-fork) (N/V removed)	1.17	1.72e-05

- Left corner effect survives even when nouns and verbs are filtered out.
- DLT effect does not.

How general are these effects?

	β-ms	p-value
DLT (modified) (N/V removed)	0.353	0.141
Left corner (no-fork) (N/V removed)	1.17	1.72e-05

- Left corner effect survives even when nouns and verbs are filtered out.
- DLT effect does not.

Plan

Conclusion

- **Main contribution:** First strong evidence of broad-coverage memory effects in sentence processing.
- Supports psychological validity of dependency/constituency.

Conclusion

- **Main contribution:** First strong evidence of broad-coverage memory effects in sentence processing.
- Supports psychological validity of dependency/constituency.

Conclusion

- Significant independent contributions from both DLT and left corner
- Possibly semantics vs. syntax?
  - DLT has access to semantic information like head/dependent, referential status, etc. Left corner does not.
  - DLT effect driven by N/V, which introduce discourse referents.
- Separate contributions might indicate separate mechanisms for (semantic) dependency construction vs. retrieval of syntactic derivations.

Conclusion

- Significant independent contributions from both DLT and left corner
- Possibly semantics vs. syntax?
  - DLT has access to semantic information like head/dependent, referential status, etc. Left corner does not.
  - DLT effect driven by N/V, which introduce discourse referents.
- Separate contributions might indicate separate mechanisms for (semantic) dependency construction vs. retrieval of syntactic derivations.

Conclusion

- Significant independent contributions from both DLT and left corner
- Possibly semantics vs. syntax?
  - DLT has access to semantic information like head/dependent, referential status, etc. Left corner does not.
  - DLT effect driven by N/V, which introduce discourse referents.
- Separate contributions might indicate separate mechanisms for (semantic) dependency construction vs. retrieval of syntactic derivations.

Conclusion

- Significant independent contributions from both DLT and left corner
- Possibly semantics vs. syntax?
  - DLT has access to semantic information like head/dependent, referential status, etc. Left corner does not.
  - DLT effect driven by N/V, which introduce discourse referents.
- Separate contributions might indicate separate mechanisms for (semantic) dependency construction vs. retrieval of syntactic derivations.

Conclusion

- Significant independent contributions from both DLT and left corner
- Possibly semantics vs. syntax?
  - DLT has access to semantic information like head/dependent, referential status, etc. Left corner does not.
  - DLT effect driven by N/V, which introduce discourse referents.
- Separate contributions might indicate separate mechanisms for (semantic) dependency construction vs. retrieval of syntactic derivations.

Acknowledgments

This work was supported by grants from the National Science Foundation: Graduate Research Fellowship Program Award DGE-1343012 to MvS; Doctoral Dissertation Research Improvement Award 1551543 to RF; Linguistics Program Award 1534318 to EG.

References I

Abney, Steven P. and Mark Johnson (1991). "Memory Requirements and Local Ambiguities of Parsing Strategies". In: *J. Psycholinguistic Research* 20.3, pp. 233–250.

Boston, Marisa Ferrara et al. (2011). "Parallel processing and sentence comprehension difficulty". In: *Language and Cognitive Processes* 26.3, pp. 301–349.

Box, G. E. P. and D. R. Cox (1964). "An Analysis of Transformations". In: *Journal of the Royal Statistical Society, B* 26, pp. 211–234.

Futrell, Richard et al. (in prep). "Natural stories corpus".

Gibson, Edward (1991). "A computational theory of human linguistic processing: Memory limitations and processing breakdown". PhD thesis. Carnegie Mellon.

– (2000). "The dependency locality theory: A distance-based theory of linguistic complexity". In: *Image, language, brain: Papers from the first mind articulation project symposium.* Cambridge, MA: MIT Press, pp. 95–126.

Grodner, Daniel J. and Edward Gibson (2005). "Consequences of the serial nature of linguistic input". In: *Cognitive Science* 29, pp. 261–291.

References II

- Johnson-Laird, Philip N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA, USA: Harvard University Press. ISBN: 0-674-56882-6.
- Lewis, Richard L. and Shravan Vasishth (2005). "An activation-based model of sentence processing as skilled memory retrieval". In: *Cognitive Science* 29.3, pp. 375–419.
- Nguyen, Luan, Marten van Schijndel, and William Schuler (2012). "Accurate Unbounded Dependency Recovery using Generalized Categorial Grammars". In: *Proceedings of the 24th International Conference on Computational Linguistics (COLING '12)*. Mumbai, India, pp. 2125–2140.
- Resnik, Philip (1992). "Left-Corner Parsing and Psychological Plausibility". In: *Proceedings of COLING*. Nantes, France, pp. 191–197.
- Rosenkrantz, Stanley J. and Philip M. Lewis II (1970). "Deterministic left corner parser". In: *IEEE Conference Record of the 11th Annual Symposium on Switching and Automata*, pp. 139–152.

References III

- Shain, Cory et al. (2016). “Memory access during incremental sentence processing causes reading time latency”. In: *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, pp. 49–58.
- Stabler, Edward (1994). “The finite connectivity of linguistic structure”. In: *Perspectives on Sentence Processing*. Lawrence Erlbaum, pp. 303–336.
- van Schijndel, Marten, Andy Exley, and William Schuler (2013). “A model of language processing as hierarchic sequential prediction”. In: *Topics in Cognitive Science* 5.3. Ed. by John Hale and David Reitter, pp. 522–540.
- von der Malsburg, Titus, Reinhold Kliegl, and Shravan Vasishth (2015). “Determinants of Scanpath Regularity in Reading”. In: *Cognitive Science*.

Appendix: Experimental setup

- Filtered out sentence starts/ends (leaving 768,023 events)
- Models evaluated via likelihood ratio test (LRT)
- Reading times transformed by Box and Cox (1964) ($\lambda \approx 0.63$)
- **Baseline:**
  `boxcox(readingTime) ~ sentencePosition + wordLength + 5GramSurp + pcfgSurp + (1 + sentenceID + sentencePosition + wordLength + 5GramSurp + pcfgSurp + mainEffect | subject) + (1 | word) + (1 | sentenceID)`
- All predictors z-score normalized prior to evaluation
- β values above are divided by standard deviation and backtransformed into ms, only valid at mean
- Predictors computed over trees in Generalized Categorial Grammar (GCG) (Nguyen, van Schijndel, and Schuler 2012)
  - Automatically reannotated from Penn Treebank-style gold and hand-corrected
  - Contains implicit dependency and memory store representations, can be used to calculate all predictors from single source

The DLT variants

- In this study, we consider 3 additional broad-coverage modifications of the DLT:
  - **DLT-V**: *Verbs are more expensive*. Non-finite verbs receive a cost of 1 (instead of 0) and finite verbs receive a cost of 2 (instead of 1).
  - **DLT-C:** *Coordination is less expensive*. Dependencies out of coordinate structures skip preceding conjuncts. Dependencies with intervening coordinations just use heaviest conjunct.
  - **DLT-M:** *Exclude modifier dependencies*. Dependencies to preceding modifiers are ignored.
- Modifications can be applied in any combination, yielding 8 implementation variants of the DLT for this study.
- Best variant was DLT-C and DLT-M together.

The DLT variants

- In this study, we consider 3 additional broad-coverage modifications of the DLT:
  - **DLT-V**: *Verbs are more expensive*. Non-finite verbs receive a cost of 1 (instead of 0) and finite verbs receive a cost of 2 (instead of 1).
  - **DLT-C:** *Coordination is less expensive*. Dependencies out of coordinate structures skip preceding conjuncts. Dependencies with intervening coordinations just use heaviest conjunct.
  - **DLT-M:** *Exclude modifier dependencies*. Dependencies to preceding modifiers are ignored.
- Modifications can be applied in any combination, yielding 8 implementation variants of the DLT for this study.
- Best variant was DLT-C and DLT-M together.

The DLT variants

- In this study, we consider 3 additional broad-coverage modifications of the DLT:
  - **DLT-V**: *Verbs are more expensive*. Non-finite verbs receive a cost of 1 (instead of 0) and finite verbs receive a cost of 2 (instead of 1).
  - **DLT-C:** *Coordination is less expensive*. Dependencies out of coordinate structures skip preceding conjuncts. Dependencies with intervening coordinations just use heaviest conjunct.
  - **DLT-M:** *Exclude modifier dependencies*. Dependencies to preceding modifiers are ignored.
- Modifications can be applied in any combination, yielding 8 implementation variants of the DLT for this study.
- Best variant was DLT-C and DLT-M together.

The DLT variants

- In this study, we consider 3 additional broad-coverage modifications of the DLT:
  - **DLT-V**: *Verbs are more expensive*. Non-finite verbs receive a cost of 1 (instead of 0) and finite verbs receive a cost of 2 (instead of 1).
  - **DLT-C:** *Coordination is less expensive*. Dependencies out of coordinate structures skip preceding conjuncts. Dependencies with intervening coordinations just use heaviest conjunct.
  - **DLT-M:** *Exclude modifier dependencies*. Dependencies to preceding modifiers are ignored.
- Modifications can be applied in any combination, yielding 8 implementation variants of the DLT for this study.
- Best variant was DLT-C and DLT-M together.

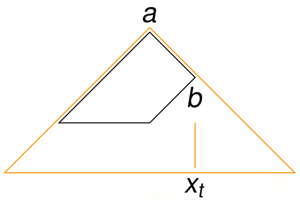
The DLT variants

- In this study, we consider 3 additional broad-coverage modifications of the DLT:
  - **DLT-V**: *Verbs are more expensive*. Non-finite verbs receive a cost of 1 (instead of 0) and finite verbs receive a cost of 2 (instead of 1).
  - **DLT-C:** *Coordination is less expensive*. Dependencies out of coordinate structures skip preceding conjuncts. Dependencies with intervening coordinations just use heaviest conjunct.
  - **DLT-M:** *Exclude modifier dependencies*. Dependencies to preceding modifiers are ignored.
- Modifications can be applied in any combination, yielding 8 implementation variants of the DLT for this study.
- Best variant was DLT-C and DLT-M together.

The DLT variants

- In this study, we consider 3 additional broad-coverage modifications of the DLT:
  - **DLT-V**: *Verbs are more expensive*. Non-finite verbs receive a cost of 1 (instead of 0) and finite verbs receive a cost of 2 (instead of 1).
  - **DLT-C:** *Coordination is less expensive*. Dependencies out of coordinate structures skip preceding conjuncts. Dependencies with intervening coordinations just use heaviest conjunct.
  - **DLT-M:** *Exclude modifier dependencies*. Dependencies to preceding modifiers are ignored.
- Modifications can be applied in any combination, yielding 8 implementation variants of the DLT for this study.
- Best variant was DLT-C and DLT-M together.

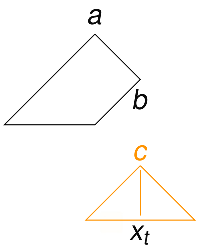
Left-corner parsing: Fork decision



No-fork (shift + match): Word satisfies b. a is complete.

$$\frac{a/b \quad x_t}{a}\; b \rightarrow x_t. \qquad (-F)$$

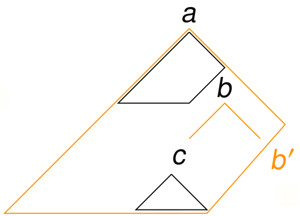
Left-corner parsing: Fork decision



Yes-fork (shift): Word does not satisfy b, fork off new complete category c.

$$\frac{a/b \quad x_t}{a/b \quad c} \; b \xrightarrow{+} c \ldots ; \quad c \rightarrow x_t. \tag{+F}$$

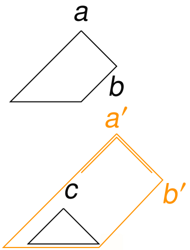
Left-corner parsing: Join decision



Yes-join (predict + match): Complete category c satisfies b while predicting b'. Store updates from $\langle \ldots, a/b, c \rangle$ to $\langle \ldots, a/b' \rangle$.

$$\frac{a/b \quad c}{a/b'} \; b \rightarrow c \; b'. \tag{+J}$$

Left-corner parsing: Join decision



No-join (predict): Complete category c does not satisfy b. Predict new a' and b' from c. Store updates from $\langle \ldots, a/b, c \rangle$ to $\langle \ldots, a/b, a'/b' \rangle$.

$$\frac{a/b \quad c}{a/b \quad a'/b'}\; b \xrightarrow{+} a' \ldots ; \quad a' \to c\; b'. \qquad (\text{–J})$$

Appendix: Left-corner predictors

- Memory effects are predicted when signs must be recalled by left-corner parser, but implementation details matter.
- We implemented 3 families of left-corner predictors:
  - **EMBD:** *End of embedded region.* True if –F+J or end of carrier, false otherwise.

Appendix: Left-corner predictors

- Memory effects are predicted when signs must be recalled by left-corner parser, but implementation details matter.
- We implemented 3 families of left-corner predictors:
  - **EMBD:** *End of embedded region.* True if –F+J or end of carrier, false otherwise.
  - **NoF:** *'No fork' (–F) operation.* True if F decision was negative.
  - **REINST:** *Reinstatement operation.* Union of EMBD and REINST.

Appendix: Left-corner predictors

- Memory effects are predicted when signs must be recalled by left-corner parser, but implementation details matter.
- We implemented 3 families of left-corner predictors:
  - **EMBD:** *End of embedded region.* True if –F+J or end of carrier, false otherwise.
  - **NoF:** *'No fork' (–F) operation.* True if F decision was negative.
  - **REINST:** *Reinstatement operation.* Union of EMBD and REINST.

Appendix: Left-corner predictors

- Memory effects are predicted when signs must be recalled by left-corner parser, but implementation details matter.
- We implemented 3 families of left-corner predictors:
  - **EMBD:** *End of embedded region.* True if –F+J or end of carrier, false otherwise.
  - **NoF:** *'No fork' (–F) operation.* True if F decision was negative.
  - **REINST:** *Reinstatement operation.* Union of EMBD and REINST.

Appendix: Left-corner predictors

- Memory effects are predicted when signs must be recalled by left-corner parser, but implementation details matter.
- We implemented 3 families of left-corner predictors:
  - **EMBD:** *End of embedded region.* True if –F+J or end of carrier, false otherwise.
  - **NoF:** *'No fork' (–F) operation.* True if F decision was negative.
  - **REINST:** *Reinstatement operation.* Union of EMBD and REINST.

Appendix: Left-corner predictors

- Also included distance-weighted variants of each of these (since last recall event) by:
  - Number of words
  - Number of DLT discourse referents
  - Number of verb-modified DLT discourse referents
- $3 \times 4 = 12$ total left-corner predictors.

Appendix: Left-corner predictors

- Also included distance-weighted variants of each of these (since last recall event) by:
  - Number of words
  - Number of DLT discourse referents
  - Number of verb-modified DLT discourse referents
- $3 \times 4 = 12$ total left-corner predictors.

Appendix: Left-corner predictors

- Also included distance-weighted variants of each of these (since last recall event) by:
  - Number of words
  - Number of DLT discourse referents
  - Number of verb-modified DLT discourse referents
- $3 \times 4 = 12$ total left-corner predictors.

Appendix: Left-corner predictors

- Also included distance-weighted variants of each of these (since last recall event) by:
  - Number of words
  - Number of DLT discourse referents
  - Number of verb-modified DLT discourse referents
- $3 \times 4 = 12$ total left-corner predictors.

Appendix: Left-corner predictors

- Also included distance-weighted variants of each of these (since last recall event) by:
  - Number of words
  - Number of DLT discourse referents
  - Number of verb-modified DLT discourse referents
- $3 \times 4 = 12$ total left-corner predictors.

Appendix: Full results

		Exploratory corpus				Confirmatory corpus			
		β	β-ms	t-value	p-value	β	β-ms	t-value	p-value
Best	**NoF-S1**	1.23e-4	1.29	6.66	1.45e-10	1.46e-4	1.54	8.15	2.33e-14
Best	**DLT-CM-S1**	1.11e-4	1.16	5.85	1.42e-8	9.63e-5	1.10	6.48	4.87e-10
Canon	**REINST-S1**	1.17e-4	1.23	6.33	1.60e-9	1.35e-4	1.43	8.01	5.77e-14
Canon	**DLT-S1**	8.04e-5	0.846	4.51	1.03e-05	6.04e-05	0.634	4.50	1.11e-05