



CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

# Coarse-To-Fine Parsing for Expressive Grammar Formalisms

**Christoph Teichmann** <sup>1</sup>

Alexander Koller <sup>1</sup>

Jonas Groschwitz <sup>1,2</sup>

<sup>1</sup>Saarland University

<sup>2</sup>Macquarie University



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

# Overview



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

- **Coarse-To-Fine parsing reduces grammar constant**
- Apply it to TAG, HRG, Bottom-up tree transducers, Synchronous CFG/TAG/HRG, Multi context-free grammars . . .
- IRTGs generalize all above
- Coarse-To-Fine hurts → connection to permissiveness



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

- Coarse-To-Fine parsing reduces grammar constant
- Apply it to TAG, HRG, Bottom-up tree transducers, Synchronous CFG/TAG/HRG, Multi context-free grammars . . .
- IRTGs generalize all above
- Coarse-To-Fine hurts → connection to permissiveness



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

- Coarse-To-Fine parsing reduces grammar constant
- Apply it to TAG, HRG, Bottom-up tree transducers, Synchronous CFG/TAG/HRG, Multi context-free grammars . . .
- IRTGs generalize all above
- Coarse-To-Fine hurts → connection to permissiveness



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

**IRTG**

CTF with  
IRTG

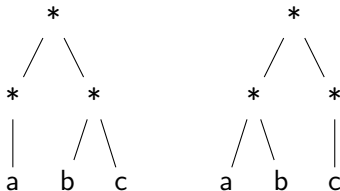
Evaluation

Analysis

Conclusion

# Interpreted Regular Tree Grammar

a b c





CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

$$A \rightarrow r_1(B, C) : 0.7 \quad | \quad A \rightarrow r_4(C, B) : 0.3$$

$$B \rightarrow r_2 : 0.6 \quad | \quad B \rightarrow r_5 : 0.4$$

$$C \rightarrow r_3 : 0.8 \quad | \quad C \rightarrow r_6 : 0.2$$





CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

**IRTG**

CTF with  
IRTG

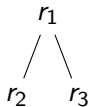
Evaluation

Analysis

Conclusion

Regular Tree  
Grammar

generates  
→



CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

$$A \rightarrow r_1(B, C) : 0.7$$

$$h(r_1) = *(x_1, x_2)$$

$$A \rightarrow r_4(C, B) : 0.3$$

$$h(r_4) = *(x_1, x_2)$$

$$B \rightarrow r_2 : 0.6$$

$$h(r_2) = a$$

$$B \rightarrow r_5 : 0.4$$

$$h(r_5) = c$$

$$C \rightarrow r_3 : 0.8$$

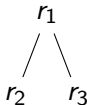
$$h(r_3) = *(b, c)$$

$$C \rightarrow r_6 : 0.2$$

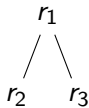
$$h(r_6) = *(a, b)$$



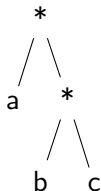
Regular Tree  
Grammar



generates  
→

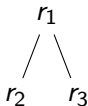


homomorphism  
→

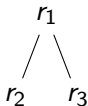




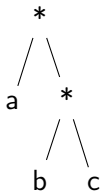
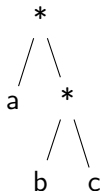
Regular Tree  
Grammar



generates →



homomorphism →



algebra →

"a b c"



# Parse Chart for "a b c"

CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

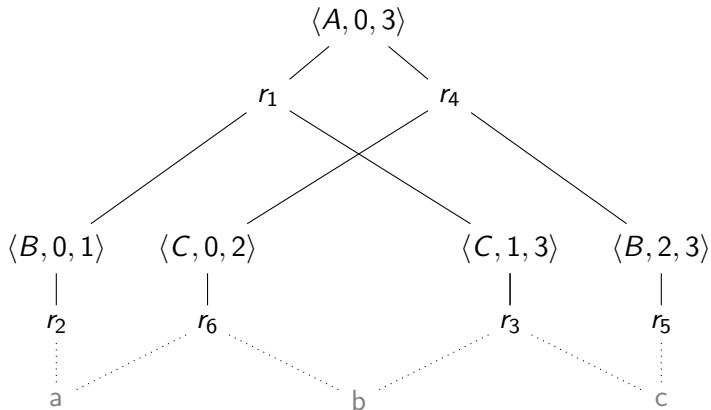
IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion



CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

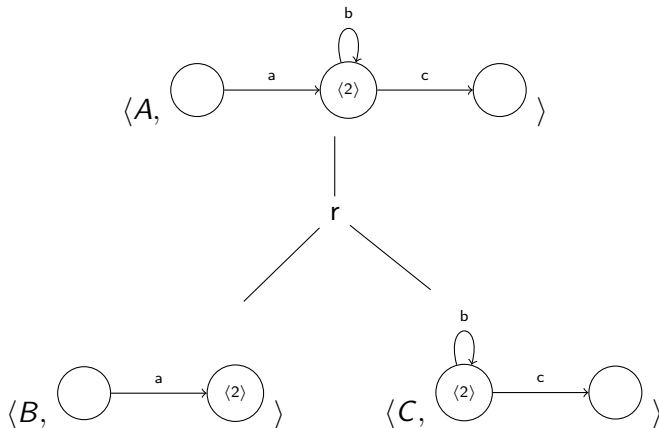
IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion



CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

Grammar formalism	Algebra 1	Algebra 2
PCFGs	strings	–
Tree-adjoining grammar (TAG)	TAG strings	–
Multi context-free grammars (MCFGs)	Lawvere strings	–
Regular tree grammars (RTGs)	trees	–
Hyperedge replacement grammars (HRGs)	graphs	–
Synchronous CFGs	strings	strings
Synchronous TAG	TAG strings	TAG strings
Bottom-up tree transducers	trees	trees
Tree-to-string transducers	trees	strings
Synchronous HRGs	strings	graphs
Semantically interpreted grammars	strings	sets

- Multiple mappings  $\rightarrow$  synchronous grammars
- Other algebras  $\rightarrow$  changes objects generated / generative power



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

# Coarse-To-Fine with Interpreted Regular Tree Grammar



- Introduced by [Charniak et al., 2006] for PCFG
- Parse with coarser grammar first
- Fine-To-Coarse mapping clusters nonterminals into groups

$$X \begin{cases} B \\ C \end{cases}$$

$$A \rightarrow X, X : 0.5 \begin{cases} A \rightarrow B, C : 0.3 \\ A \rightarrow C, B : 0.2 \end{cases}$$

- Compute score (inside  $\times$  outside) for each edge
- Set threshold

If  $\langle A, 0, 2 \rangle \rightarrow \langle X, 0, 1 \rangle, \langle X, 1, 2 \rangle$  below threshold  $\rightarrow$   
Skip  $\langle A, 0, 2 \rangle \rightarrow \langle B, 0, 1 \rangle, \langle C, 1, 2 \rangle$  and  
 $\langle A, 0, 2 \rangle \rightarrow \langle C, 0, 1 \rangle, \langle B, 1, 2 \rangle$  in fine chart

CTF for  
IRTGs

 Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

 CTF with  
IRTG

Evaluation

Analysis

Conclusion

$$A \rightarrow r_1(B, C) : 0.7$$

$$h(r_1) = *(x_1, x_2)$$

$$A \rightarrow r_4(C, B) : 0.3$$

$$h(r_4) = *(x_1, x_2)$$

$$B \rightarrow r_2 : 0.6$$

$$h(r_2) = a$$

$$B \rightarrow r_5 : 0.4$$

$$h(r_5) = c$$

$$C \rightarrow r_3 : 0.8$$

$$h(r_3) = *(b, c)$$

$$C \rightarrow r_6 : 0.2$$

$$h(r_6) = *(a, b)$$

CTF for  
IRTGs

 Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

 CTF with  
IRTG

Evaluation

Analysis

Conclusion

$$A \rightarrow r_1(X, X) : 0.7$$

$$h(r_1) = *(x_1, x_2)$$

$$A \rightarrow r_4(X, X) : 0.3$$

$$h(r_4) = *(x_1, x_2)$$

$$X \rightarrow r_2 : 0.6$$

$$h(r_2) = a$$

$$X \rightarrow r_5 : 0.4$$

$$h(r_5) = c$$

$$X \rightarrow r_3 : 0.8$$

$$h(r_3) = *(b, c)$$

$$X \rightarrow r_6 : 0.2$$

$$h(r_6) = *(a, b)$$

$$\begin{array}{l}
 A \quad \rightarrow \{r_1, r_4\}(X, X) : 1.0 \\
 h(\{r_1, r_2\}) = *(x_1, x_2)
 \end{array}$$

$$\begin{array}{l}
 X \quad \rightarrow r_2 : 0.6 \\
 h(r_2) = a
 \end{array}$$

$$\begin{array}{l}
 X \quad \rightarrow r_3 : 0.8 \\
 h(r_3) = *(b, c)
 \end{array}$$

$$\begin{array}{l}
 X \quad \rightarrow r_5 : 0.4 \\
 h(r_5) = c
 \end{array}$$

$$\begin{array}{l}
 X \quad \rightarrow r_6 : 0.2 \\
 h(r_6) = *(a, b)
 \end{array}$$

CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

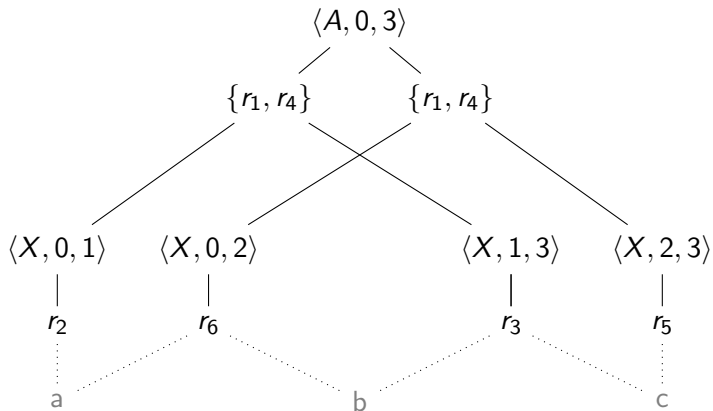
IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion



CTF for  
IRTGs

 Teichmann,  
Groschwitz,  
Koller

Overview

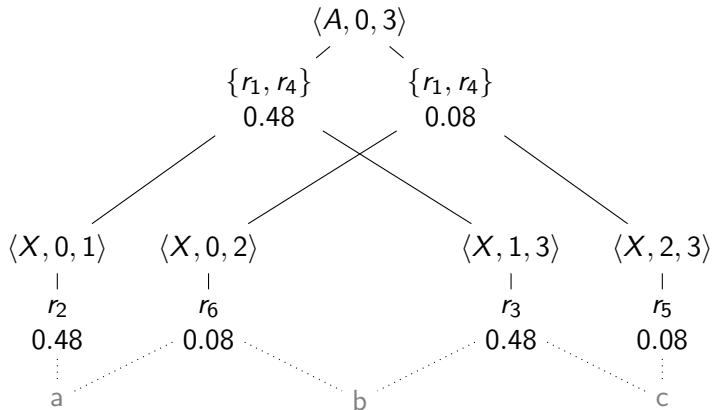
IRTG

 CTF with  
IRTG

Evaluation

Analysis

Conclusion





CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

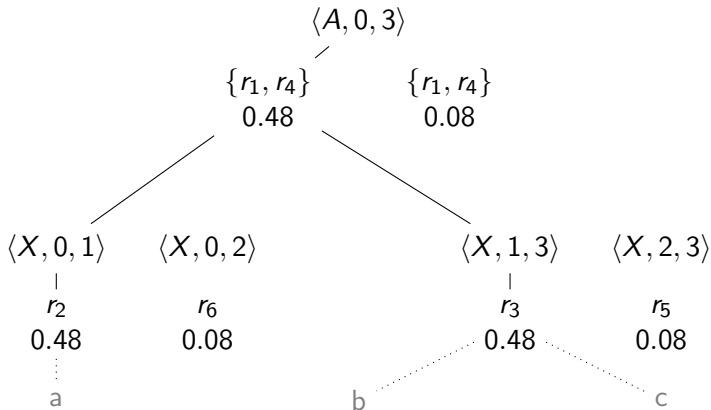
IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion







CTF for  
IRTGs

 Teichmann,  
Groschwitz,  
Koller

Overview

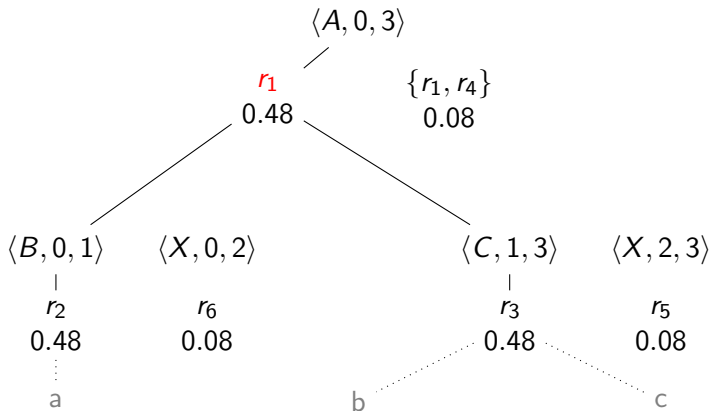
IRTG

 CTF with  
IRTG

Evaluation

Analysis

Conclusion



CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

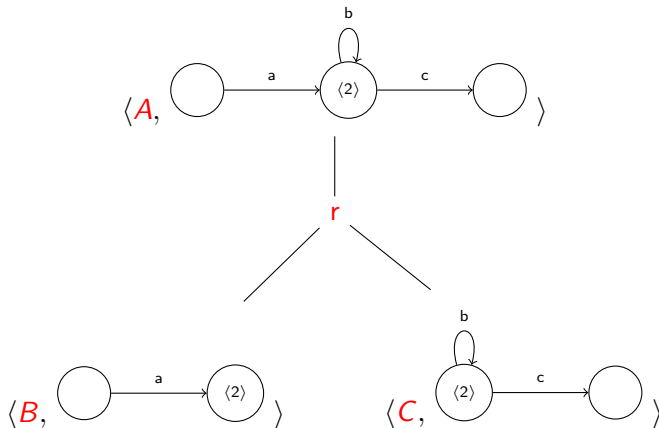
IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion



CTF for  
IRTGs

 Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

 CTF with  
IRTG

Evaluation

Analysis

Conclusion

Grammar formalism	Algebra 1	Algebra 2
PCFGs	strings	–
Tree-adjoining grammar (TAG)	TAG strings	–
Multi context-free grammars (MCFGs)	Lawvere strings	–
Regular tree grammars (RTGs)	trees	–
Hyperedge replacement grammars (HRGs)	graphs	–
Synchronous CFGs	strings	strings
Synchronous TAG	TAG strings	TAG strings
Bottom-up tree transducers	trees	trees
Tree-to-string transducers	trees	strings
Synchronous HRGs	strings	graphs
Semantically interpreted grammars	strings	sets

- Completely generic
- CTF for all formalisms above
- Only need to specify Fine-to-Coarse mapping



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

**Evaluation**

Analysis

Conclusion

# Evaluation

CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

- Induced from Penn Treebank [Marcus et al., 1993] – directly from trees
- Only POS Tags, no words
- Fine-to-Coarse map from [Charniak et al., 2006]

Approach	f-score	time (ms)
Standard	73.7	1230
CTF	73.9	58

Generic Algorithm works!

- Grammar induced from “Little Prince” portion of AMR bank [Banarescu et al., 2013] by [Groschwitz et al., 2015]
- Nonterminals keep track of position of root
- Fine-to-Coarse Mapping: remove root information

Approach	best %	time (ms)
Standard	100.0	622
CTF	96.0	3279

CTF hurts! (It also did not help for other mappings and TAG)



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

**Analysis**

Conclusion

# Analysis of Results



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

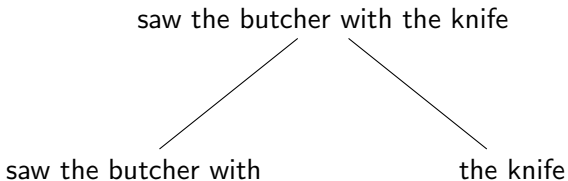
IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion



Likely to be allowed by grammar [Klein and Manning, 2001]

CTF for  
IRTGsTeichmann,  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

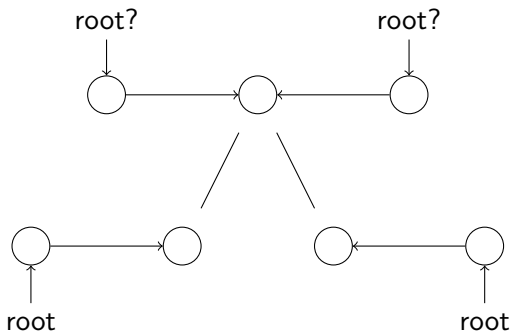
Evaluation

Analysis

Conclusion

<b>Approach</b>	<b>time (ms)</b>	<b>Substructure Combinations</b>
<b>Standard</b>	1230	2190
<b>CTF</b>	58	2260

Coarsest grammar does not substantially increase number of splits explored!



Not allowed by grammar



CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

<b>Approach</b>	<b>time (ms)</b>	<b>Substructure Combinations</b>
<b>Standard</b>	622	9042
<b>CTF</b>	3279	35476

Coarsest grammar substantially increases number of splits  
explored!

*Is this a general problem?*

# Ways to Divide Graph

CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

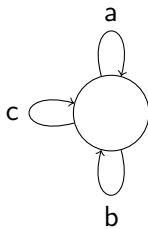
IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion





CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

- Fine grammar permissive  $\rightarrow$  CTF works well
- Fine grammar restrictive  $\rightarrow$  CTF might hurt
- Expressive grammar formalisms need restrictive grammars for efficient parsing



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

Conclusion



CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

- CTF parser for many grammar formalisms
- Generic implementation helps where original algorithm works
- Less useful for restrictive grammars





CTF for  
IRTGs

**Teichmann,**  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion

- Investigate grammar dependence of other speed-up techniques
- Use techniques that rule out possible substructures early

CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

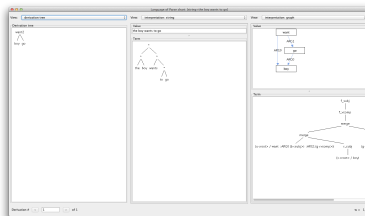
CTF with  
IRTG

Evaluation

Analysis

Conclusion

- Grammars and Fine-to-Coarse mappings available — see paper
- Code available as part of Alto toolkit – try out your own grammar



CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion



Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., and Schneider, N. (2013).

Abstract meaning representation for sembanking.

*In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186.



Charniak, E., Johnson, M., Elsner, M., Austerweil, J., Ellis, D., Haxton, I., Hill, C., Shrivaths, R., Moore, J., Pozar, M., and Vu, T. (2006).

Multilevel coarse-to-fine pcfg parsing.

*In Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*.



CTF for  
IRTGs

Teichmann,  
Groschwitz,  
Koller

Overview

IRTG

CTF with  
IRTG

Evaluation

Analysis

Conclusion



Groschwitz, J., Koller, A., and Teichmann, C. (2015).  
Graph parsing with S-graph Grammars.

*In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 1481–1490.*



Klein, D. and Manning, C. D. (2001).

Parsing with treebank grammars: Empirical bounds, theoretical models, and the structure of the penn treebank.  
*In Proceedings of 39th Annual Meeting of the Association for Computational Linguistics.*



Koller, A. and Kuhlmann, M. (2011).

A generalized view on parsing and translation.

*In Proceedings of the 12th International Conference on Parsing Technologies*, pages 2–13.



Koller, A. and Kuhlmann, M. (2012).

Decomposing tag algorithms using simple algebraizations.

*In Proceedings of the 11th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+11)*, pages 135–143.



Marcus, M. P., Marcinkiewicz, M. A., and Santorini, B. (1993).

Building a large annotated corpus of english: The penn treebank.

*Computational Linguistics*, 19(2):313–330.