A Corpus-based Syntactic Analysis of Two-termed Unlike Coordination

Julie Kallini and **Christiane Fellbaum**

Findings of EMNLP 2021

Part 1: Introduction

- Two-termed Coordination
- LCL
- Motivation and Goal

Part 1: Introduction

- → Two-termed Coordination
- LCL
- Motivation and Goal

Coordination groups two or more phrases together:

Part 1: Introduction

- → Two-termed Coordination
- LCL
- Motivation and Goal

Coordination groups two or more phrases together:

[NP The chicken] and [NP the rice] go well together.

Part 1: Introduction

- → Two-termed Coordination
- LCL
- Motivation and Goal

Coordination groups two or more phrases together:

[NP The chicken] and [NP the rice] go well together.

Conjuncts

Part 1: Introduction

→ Two-termed Coordination

- LCL
- Motivation and Goal

Coordination groups two or more phrases together:

[NP The chicken] and [NP the rice] go well together.

Coordinating Conjunction

Part 1: Introduction

- Two-termed Coordination
- → LCL
- Motivation and Goal

Law of Coordination of Likes (LCL)

mandates like conjunct categories

(Prazmowska, 2015; Williams, 1981)

Part 1: Introduction

- Two-termed Coordination
- → LCL
- Motivation and Goal

Law of Coordination of Likes (LCL)

mandates like conjunct categories

What about unlike coordinations...

Pat is [NP a Republican] and [AP proud of it].

(Prazmowska, 2015; Williams, 1981)

Part 1: Introduction

- Two-termed Coordination
- LCL
- → Motivation and Goal

Motivation: the extent and limits of unlike coordination have not been defined

Part 1: Introduction

- Two-termed Coordination
- LCL
- → Motivation and Goal

Motivation: the extent and limits of unlike coordination have not been defined

Goal: analyze unlike coordination through a quantitative corpus analysis

Part 2: Background

- Linguistics
  Syntax
- Structure of
  Coordination

Part 2: Background

- → Linguistics Syntax
- Structure of Coordination

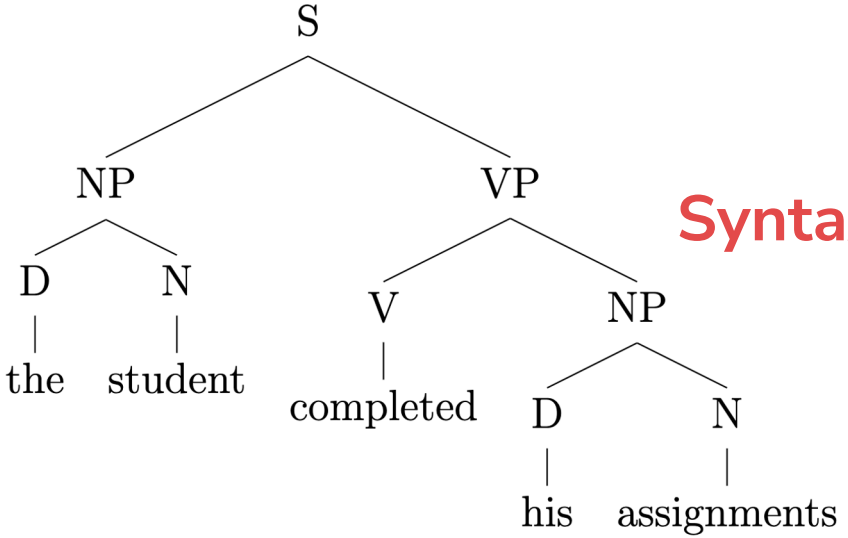
Syntax: the study of sentence structure

Part 2: Background

→ Linguistics Syntax

- Structure of Coordination

Syntax: the study of sentence structure



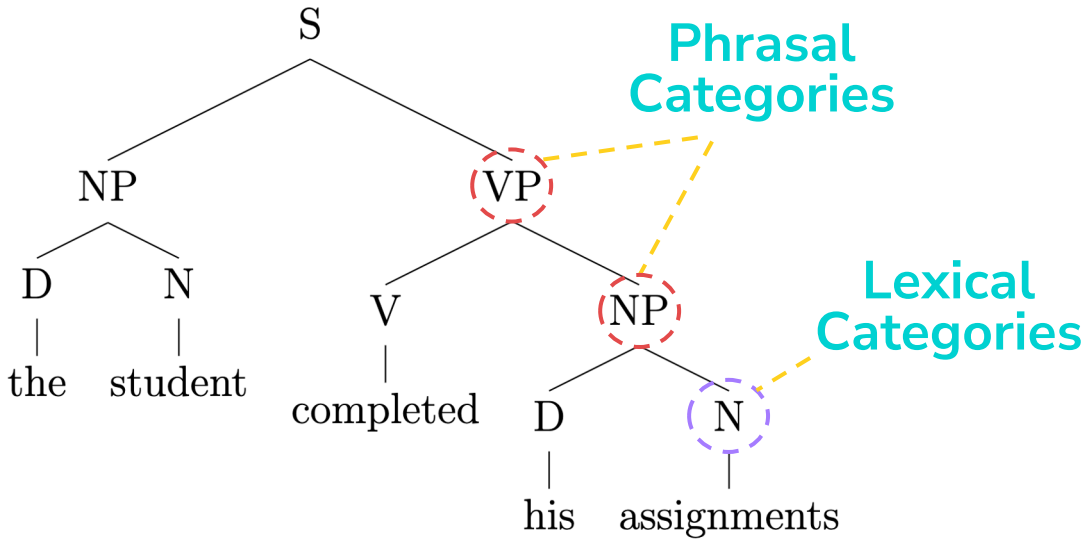
Syntax Trees

Part 2: Background

→ Linguistics Syntax

- Structure of Coordination

Syntax: the study of sentence structure

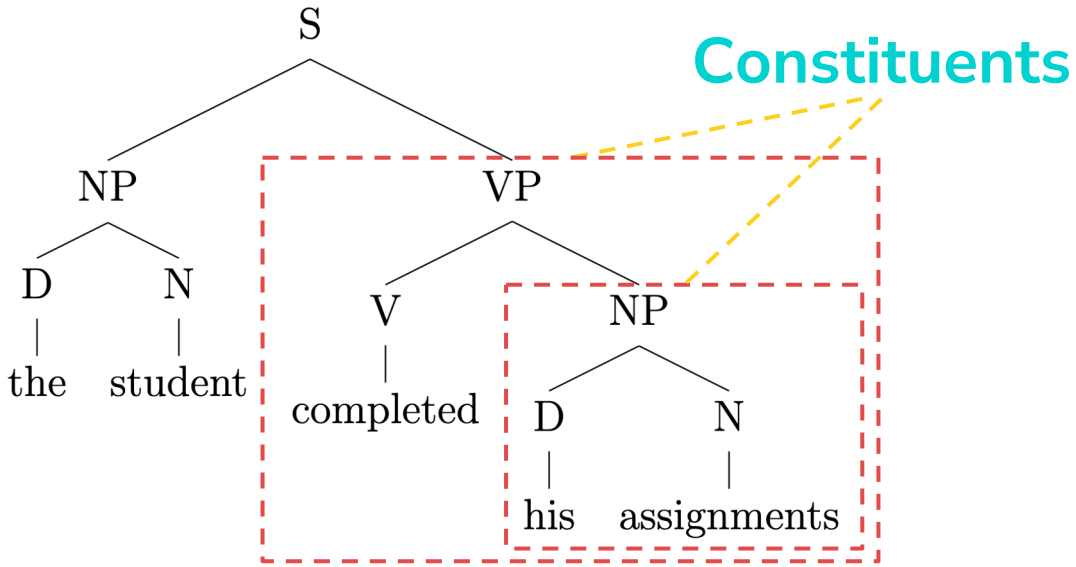


Part 2: Background

→ Linguistics Syntax

- Structure of Coordination

Syntax: the study of sentence structure

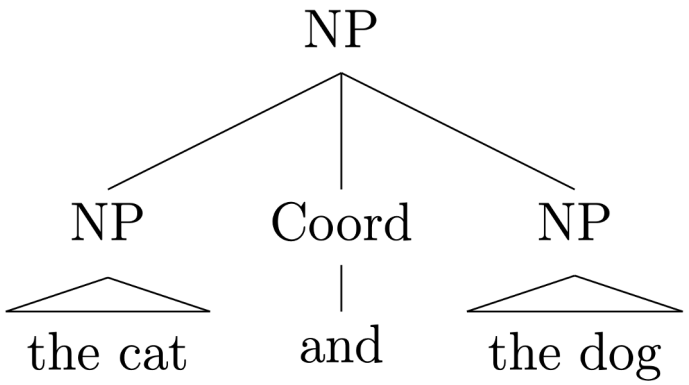


Part 2: Background

- Linguistics
  Syntax
- → Structure of Coordination

Structure of coordination

[NP the cat] and [NP the dog]



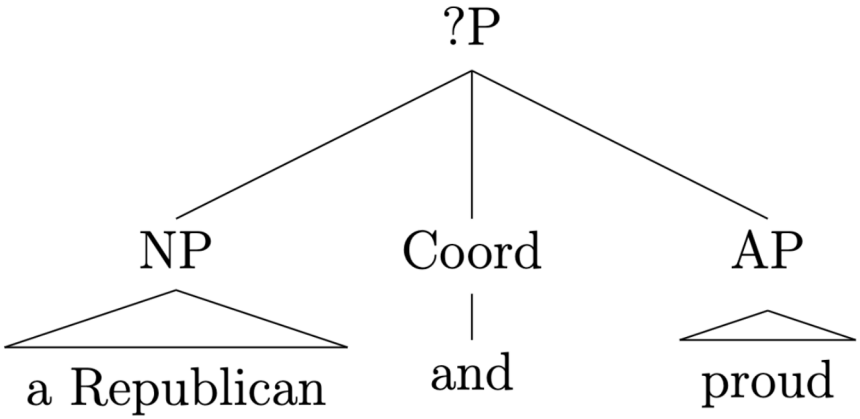
(Progovac, 1998a; Chomsky, 1981; Prazmowska, 2015)

Part 2: Background

- Linguistics Syntax
- → Structure of Coordination

Structure of coordination

[NP a Republican] and [AP proud]



(Progovac, 1998a; Johannessen; 1998; Zoerner, 1995)

Part 3: Approach

- Key Idea
- Data
- Syntactic Analysis

Part 3: Approach

→ Key Idea
- Data
- Syntactic Analysis

Key Idea: process and analyze corpus data to study the attested usage of unlike coordination in English

Part 3: Approach

→ Key Idea
- Data
- Syntactic Analysis

Key Idea: process and analyze corpus data to study the attested usage of unlike coordination in English

- Place emphasis on empirical rather than intuitive data

Part 3: Approach

→ Key Idea
- Data
- Syntactic Analysis

Key Idea: process and analyze corpus data to study the attested usage of unlike coordination in English

- Place emphasis on empirical rather than intuitive data
- Identify the constraints on coordination

Part 3: Approach

- Key Idea
- → Data
- Syntactic Analysis

Two Data Sources

- **Corpus of Contemporary American English (COCA)**
  - >450 million words of text
  - Balanced data from five genres

(Davies, 2015; Ficler and Goldberg, 2016; Marcus et al., 1993)

Part 3: Approach

- Key Idea
- → Data
- Syntactic Analysis

Two Data Sources

- **Corpus of Contemporary American English (COCA)**
  - >450 million words of text
  - Balanced data from five genres
- **Penn Treebank (PTB)** with manual Coordination Annotation Extension

(Davies, 2015; Ficler and Goldberg, 2016; Marcus et al., 1993)

Part 3: Approach

- Key Idea
- Data
- → Syntactic Analysis

Coordination Phrase Extraction

- Use Berkeley Neural Parser to produce trees of COCA sentences
- Search for coordination structures within parse trees
  - In both COCA and PTB

(Kitaev and Klein, 2018)

Part 3: Approach

- Key Idea
- Data
- → Syntactic Analysis

Coordination Extraction from COCA

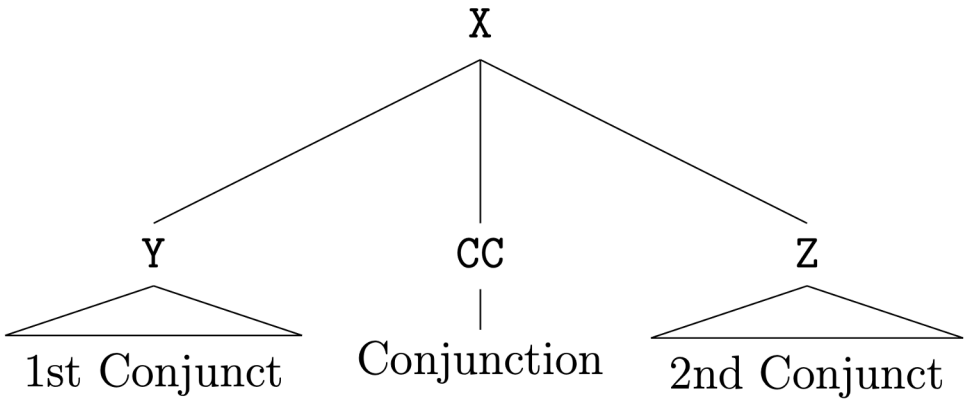
Three patterns of two-termed coordination…

Part 3: Approach

- Key Idea
- Data
- → Syntactic Analysis

Coordination Extraction from COCA

Pattern 1: Simple Ternary-branching Tree

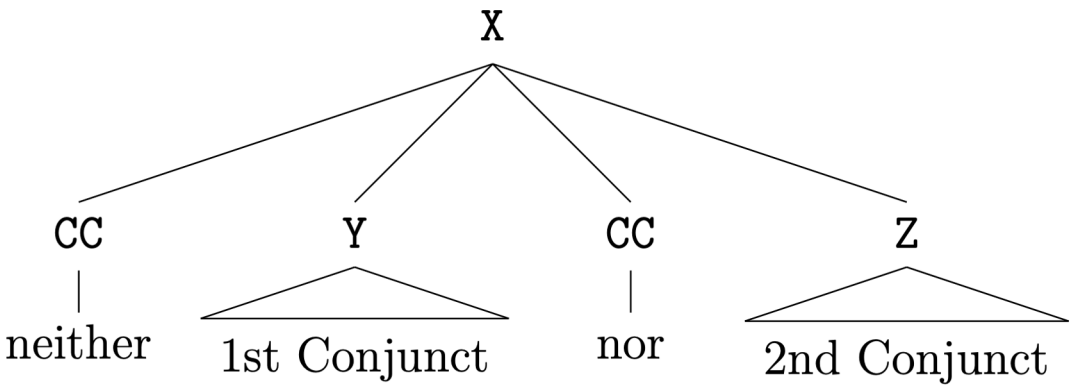


Part 3: Approach

- Key Idea
- Data
- → Syntactic Analysis

Coordination Extraction from COCA

Pattern 2: *Neither-Nor* Tree

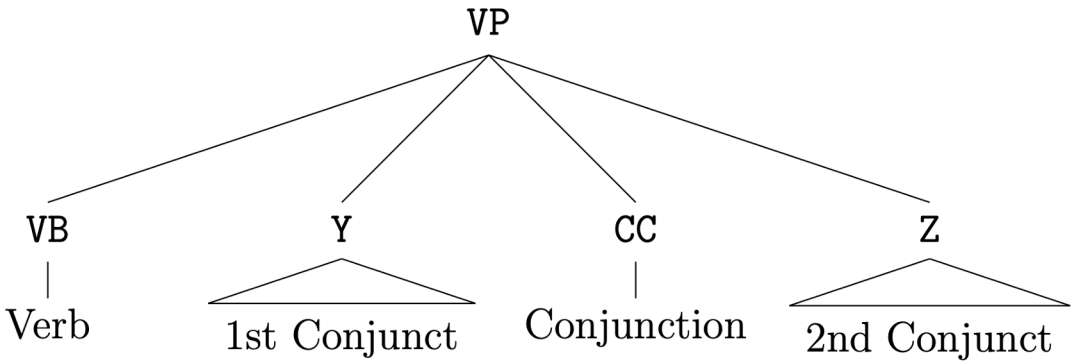


Part 3: Approach

- Key Idea
- Data
- → Syntactic Analysis

Coordination Extraction from COCA

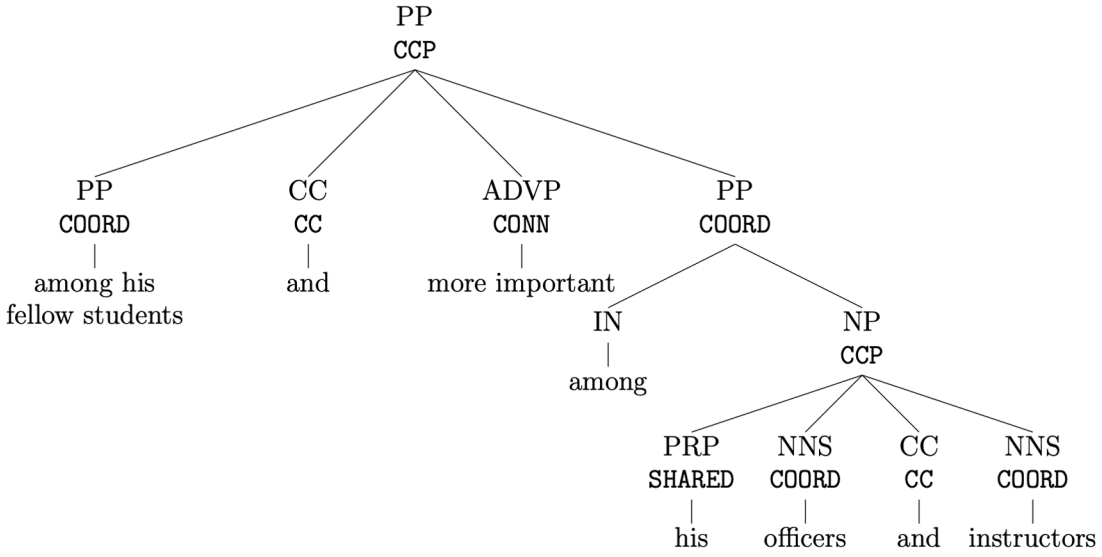
Pattern 3: Verb-Complement Tree



Part 3: Approach

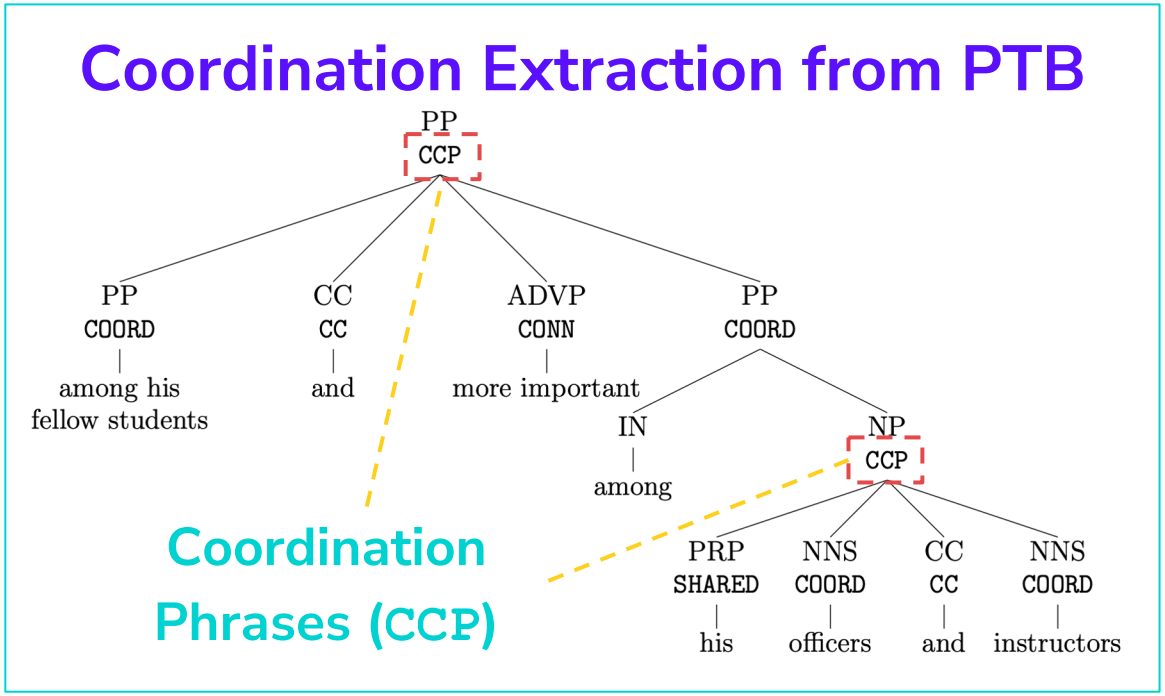
- Key Idea
- Data
- → Syntactic Analysis

Coordination Extraction from PTB



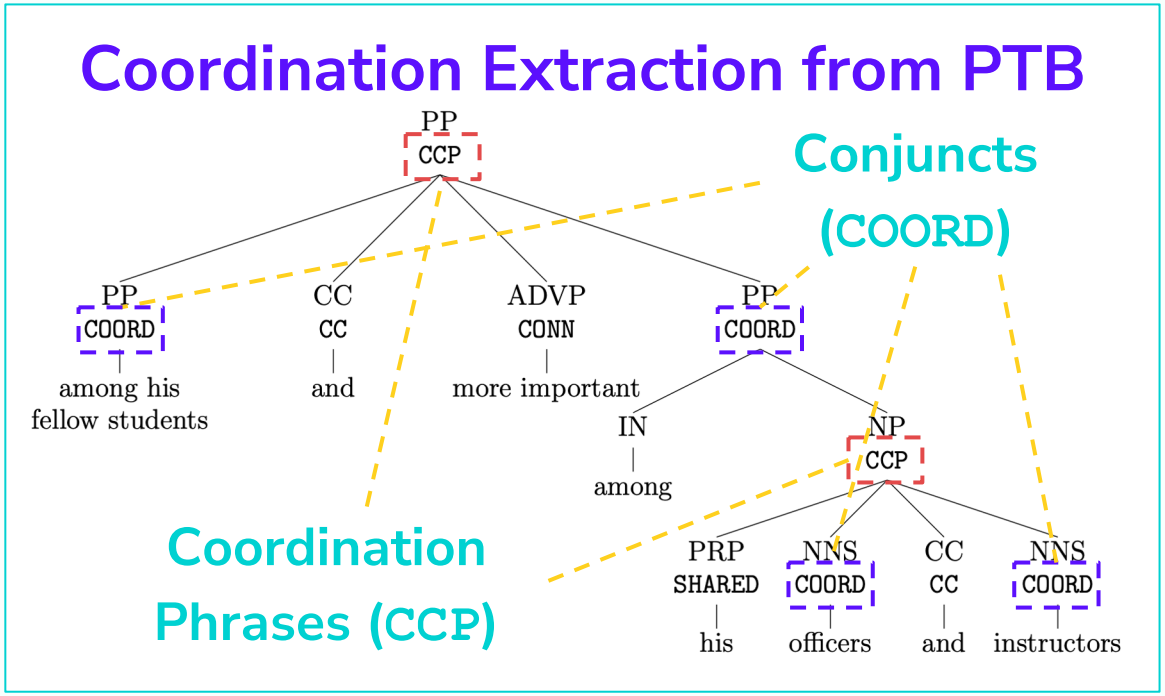
Part 3: Approach

- Key Idea
- Data
- → Syntactic Analysis



Part 3: Approach

- Key Idea
- Data
- → Syntactic Analysis



Part 4: Results

- Frequent Unlike Coordinations
- Differences from Position
- Evaluation

Results

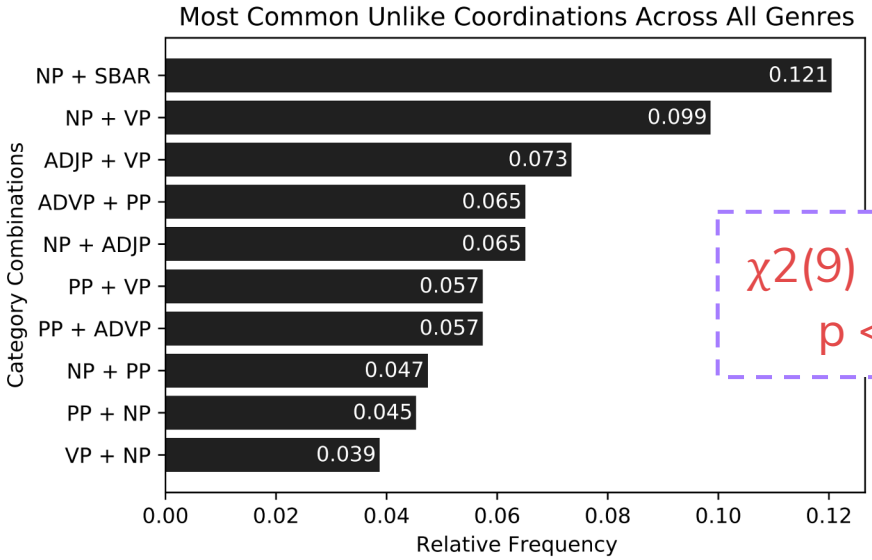
- Presentation and discussion of most important results
- Consider $p < .05$ to be significant
- Include post-tests

(In the interest of time, this presentation includes a subset of all results)

Part 4: Results

- → Frequent Unlike Coordinations
- Differences from Position
- Evaluation

Most Frequent Unlike Coordinations



$\chi^2(9) = 3142.0$
$p < .001$

*SBAR = Subordinate Clause

Part 4: Results

→ Frequent Unlike Coordinations

- Differences from Position
- Evaluation

Most Frequent Unlike Coordinations

- High **NP+SBAR** frequency might be explained by sentential subjects

  We disliked [NP Mary's attitude] and [SBAR that she left early].

  [SBAR That Mary left] disappointed us.

- This is more evidence that clauses distribute like NPs

(Lohndal, 2014)

*SBAR = Subordinate Clause

Part 4: Results

- Frequent Unlike Coordinations
- → Differences from Position
- Evaluation

Differences Between Conjunct Positions

Cat.	1st Conjunct (%)	2nd Conjunct (%)	χ2	N	V
NP	70.75	29.24	3200.7	18582	.415
VP	32.42	67.58	1764.4	14277	.352
PP	53.47	46.53	68.789	14248	.069
ADJP	55.73	44.27	125.57	9566	.114
ADVP	48.71	51.29	5.076	7645	.026
SBAR	23.97	76.03	2385.8	8800	.521

Part 4: Results

- Frequent Unlike Coordinations
- → Differences from Position
- Evaluation

Differences Between Conjunct Positions

Cat.	1st Conjunct (%)	2nd Conjunct (%)	χ2	N	V
NP	70.75	29.24	3200.7	18582	.415
VP	32.42	67.58	1764.4	14277	.352
PP	53.47	46.53	68.789	14248	.069
ADJP	55.73	44.27	125.57	9566	.114
ADVP	48.71	51.29	5.076	7645	.026
SBAR	23.97	76.03	2385.8	8800	.521

*p < .05, SBAR = Subordinate Clause

Part 4: Results

- Frequent Unlike Coordinations
- → Differences from Position
- Evaluation

Differences Between Conjunct Positions

Cat.	1st Conjunct (%)	2nd Conjunct (%)	χ2	N	V
NP	70.75	29.24	3200.7	18582	.415
VP	32.42	67.58	1764.4	14277	.352
PP	53.47	46.53	68.789	14248	.069
ADJP	55.73	44.27	125.57	9566	.114
ADVP	48.71	51.29	5.076	7645	.026
SBAR	23.97	76.03	2385.8	8800	.521

*$p < .05$, SBAR = Subordinate Clause

Part 4: Results

- Frequent Unlike Coordinations
- Differences from Position
- → Evaluation

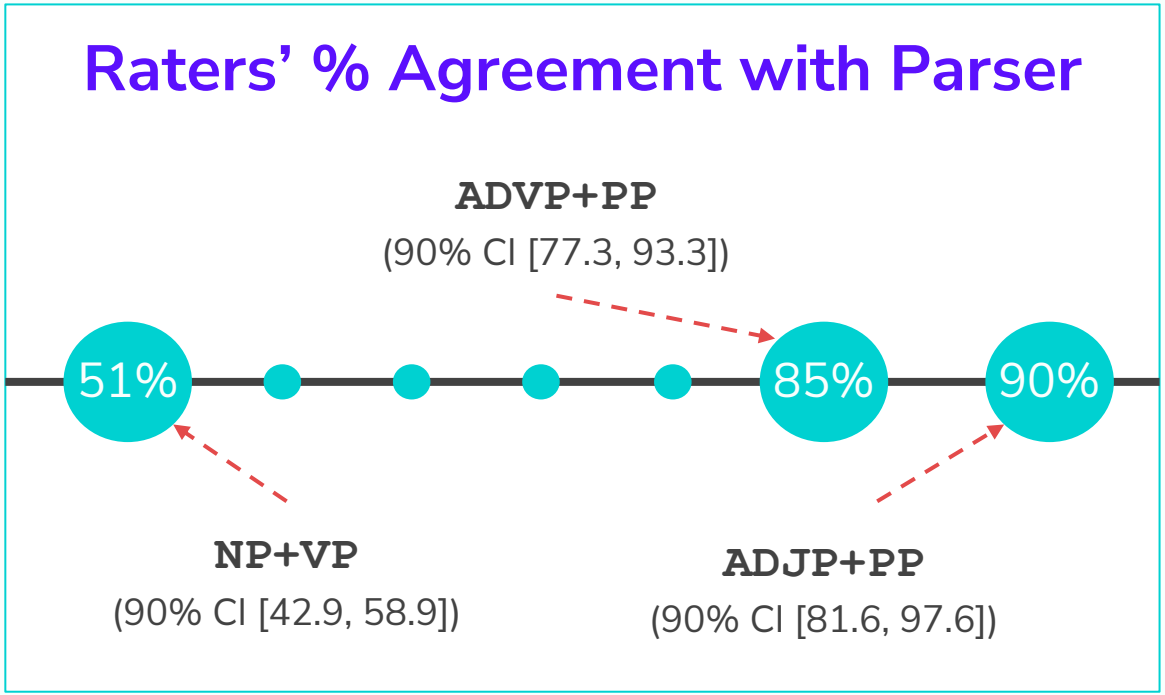
Evaluation Plan

- Randomly sampled unlike coordinations
- 3 independent raters evaluated agreement with parser's labelings

```
If you 're [VP married] and [PP in uniform] , what have
you learned after a decade of war?
Coordination phrase: [VP married] and [PP in uniform]
Is this coordination phrase correctly labeled? (y/n)
```

Part 4: Results

- Frequent Unlike Coordinations
- Differences from Position
- → Evaluation



*For agreement between raters, Fleiss's Kappa K = 0.29 (fair) (McHugh, 2015)

Part 5: Conclusion

- Concluding Remarks
- Future Work

Part 5: Conclusion

→ Concluding Remarks

- Future Work

Concluding Remarks

- **Constraints on unlike coordination** are still elusive to linguists

Part 5: Conclusion

→ Concluding Remarks

- Future Work

Concluding Remarks

- **Constraints on unlike coordination** are still elusive to linguists
- We use a **computational syntactic analysis** to study coordination

Part 5: Conclusion

→ Concluding Remarks

- Future Work

Concluding Remarks

- **Constraints on unlike coordination** are still elusive to linguists
- We use a **computational syntactic analysis** to study coordination
- Our analysis provides new data and perspectives to **shape future theories**

Part 5: Conclusion

- Concluding Remarks
- → Future Work

Future Work

- **Semantic** constraints of coordination

Part 5: Conclusion

- Concluding Remarks
- → Future Work

Future Work

- **Semantic** constraints of coordination
- Ungrammatical like coordinations…

  * John ate [PP with his mother] and [PP with good appetite].

Acknowledgments

- Thank you to the second reader **Srinivas Bangalore** and the students of his *Machine Translation* class for help with the evaluation plan
- Thank you to the **EMNLP reviewers** for their feedback and suggestions

Thank you!

References

Haldun Akoglu. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18.

John Bowers. 1993. The syntax of predication. *Linguistic Inquiry*, 24(4):591–656.

Noam Chomsky. 1981. *Lectures On Government and Binding*. Foris Publications.

Timothy Colleman, Bernard De Clerck, and Magda SBAR Devos. 2010. Prepositional dative constructions in english and dutch: A contrastive semantic analysis. *Neuphilologische Mitteilungen*, 111(2):131–152.

Mark Davies. 2015. Corpus of Contemporary American English (COCA).

Jessica Ficler and Y. Goldberg. 2016. Coordination annotation extension in the penn treebank. *ArXiv*, abs/1606.02529.

References

Janne Bondi Johannessen. 1998. *Coordination.* Oxford University Press.

Richard Kayne. 1994. *The Antisymmetry of Syntax*. MIT Press.

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)*, Melbourne, Australia. Association for Computational Linguistics.

Terje Lohndal. 2014. Sentential subjects in english and norwegian. *Syntax and Semantics*, 15:81–113.

Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. *Comput. Linguist.*, 19(2):313–330.

References

Mary L. McHugh. 2015. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.

Anna Prazmowska. 2015. Is unlike coordination against the law (of the coordination of likes)?

Ljiljana Progovac. 1998a. Structure for coordination: Part 1. In *GLOT International* 3.7.

Ivan Sag, Gerald Gazdar, Thomas Wasow, and Steven Weisler. 1985. Coordination and how to distinguish categories. *Natural Language and Linguistic Theory*, 3:117–171.

Edwin S. Williams. 1981. Transformationless grammar. In *Linguistic Inquiry 12*, pages 645–653.

Cyril Edward Zoerner. 1995. *Coordination: The Syntax of &P*. UCI dissertations in linguistics. University of California, Irvine.