

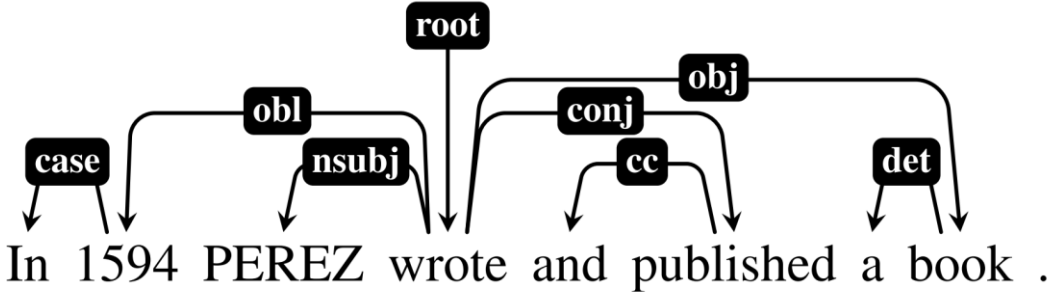
Coordinate Constructions in English Enhanced UD: Analysis and Computational Modeling

Stefan Grünewald, Prisca Piccirilli, Annemarie Friedrich

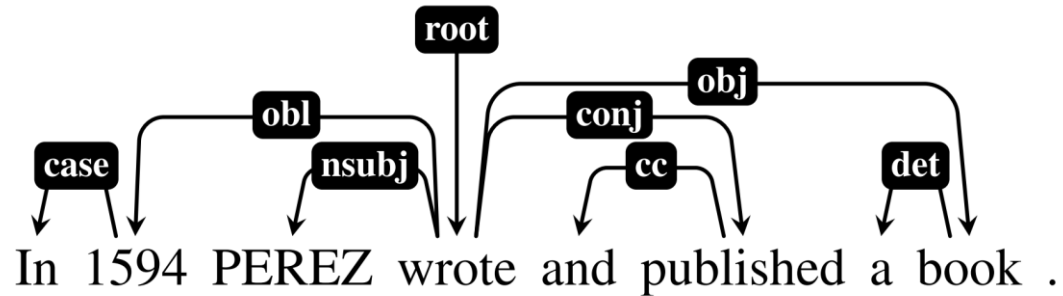
EACL 2021



Coordinate Constructions in Enhanced UD



Coordinate Constructions in Enhanced UD



BASIC UD:

Who wrote a book?

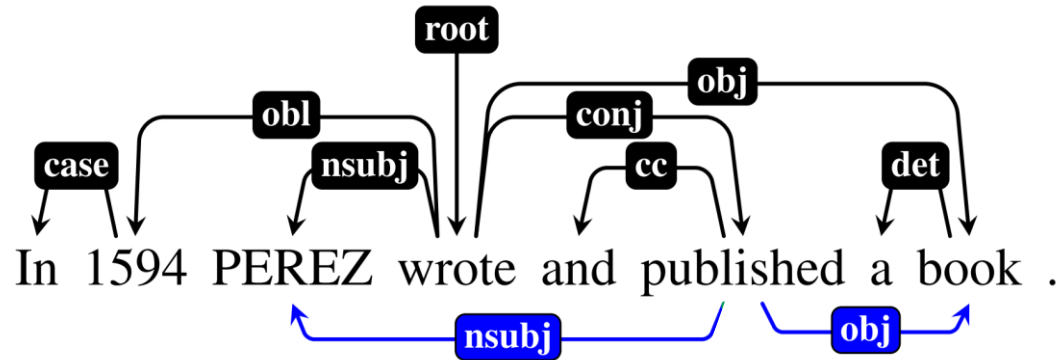
When was the book written?

What was written?

→ manually created

→ English Web Treebank (EWT)

Coordinate Constructions in Enhanced UD



BASIC UD:

Who wrote a book?
When was the book written?
What was written?

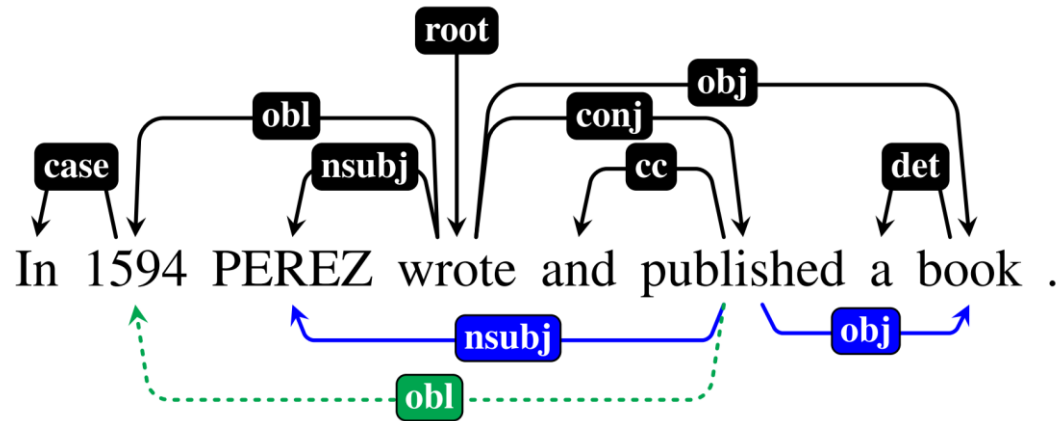
→ manually created
→ English Web Treebank (EWT)

ENHANCED UD:

Who published a book?
What was published?

→ rule-based converter
[Schuster & Manning, 2016]

Coordinate Constructions in Enhanced UD



FURTHER PROPAGATION:

When was the book published?

→ extension

BASIC UD:

Who wrote a book?

When was the book written?

What was written?

→ manually created

→ English Web Treebank (EWT)

ENHANCED UD:

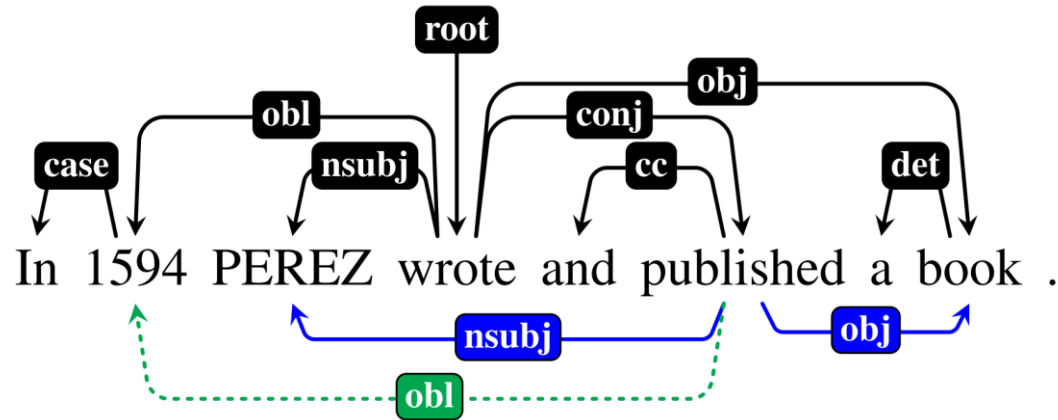
Who published a book?

What was published?

→ rule-based converter

[Schuster & Manning, 2016]

Coordinate Constructions in Enhanced UD



FURTHER PROPAGATION:

When was the book published?

→ extension

BASIC UD:

Who wrote a book?

When was the book written?

What was written?

→ manually created

→ English Web Treebank (EWT)

ENHANCED UD:

Who published a book?

What was published?

→ rule-based converter

[Schuster & Manning, 2016]

1) manual corpus validation + extension
→ **high quality gold standard dataset**

2) Computational models for conjunction propagation

Corpus Study and Manual Annotations (EWT)

- Extracted the sentences containing **conjoined verb phrases** (*VERB conj VERB*)
- Modification / verification by expert annotator of **enhanced links** involved in (all) **coordinate structures**
- New: propagate **non-core dependents**: *obl, advcl, advmod*

Corpus Study and Manual Annotations (EWT)

- Extracted the sentences containing **conjoined verb phrases** (*VERB conj VERB*)
- Modification / verification by expert annotator of **enhanced links** involved in (all) **coordinate structures**
- New: propagate **non-core dependents**: *obl, advcl, advmod*

	conj.sent	edited
train	1,926	999
dev	222	222
test	196	196
total	2,344	1,417

Modifications:

- often added: *nsubj* (160), *obl* (72), *nmod* (32), *advmod* (46)
- often removed: *nsubj:pass* (18), *nsubj* (30)

statistics by label → paper

Coordinate constructions dataset statistics

Corpus Study and Manual Annotations (EWT)

- Extracted the sentences containing **conjoined verb phrases** (*VERB conj VERB*)
- Modification / verification by expert annotator of **enhanced links** involved in (all) **coordinate structures**
- New: propagate **non-core dependents**: *obl*, *advcl*, *advmod*

	conj.sent	edited
train	1,926	999
dev	222	222
test	196	196
total	2,344	1,417

Modifications:

- often added: *nsubj* (160), *obl* (72), *nmod* (32), *advmod* (46)
- often removed: *nsubj:pass* (18), *nsubj* (30)

statistics by label → paper

Coordinate constructions dataset statistics

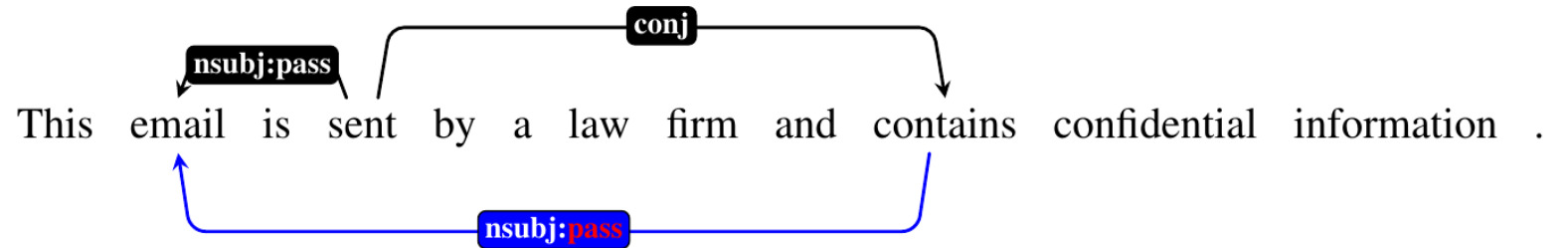
	A	B	C
A	-	90.1	94.9
B	95.2	-	97.2
C	80.5	77.9	-

Inter-annotator agreement study:
100 sentences: precision/recall

Corpus Annotation of Coordinate Structures

Analysis: Examples

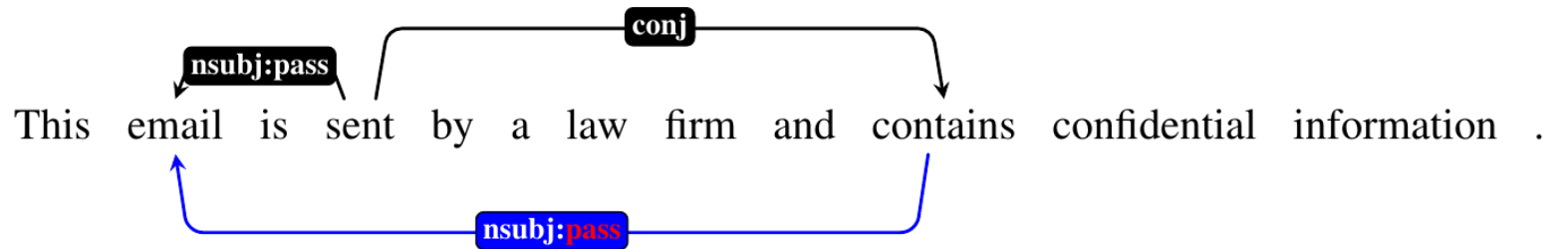
Passive Propagation:
18 times



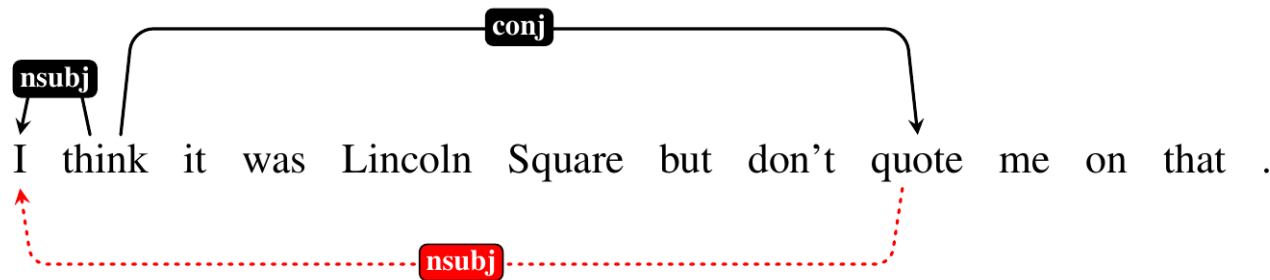
Corpus Annotation of Coordinate Structures

Analysis: Examples

Passive Propagation:
18 times



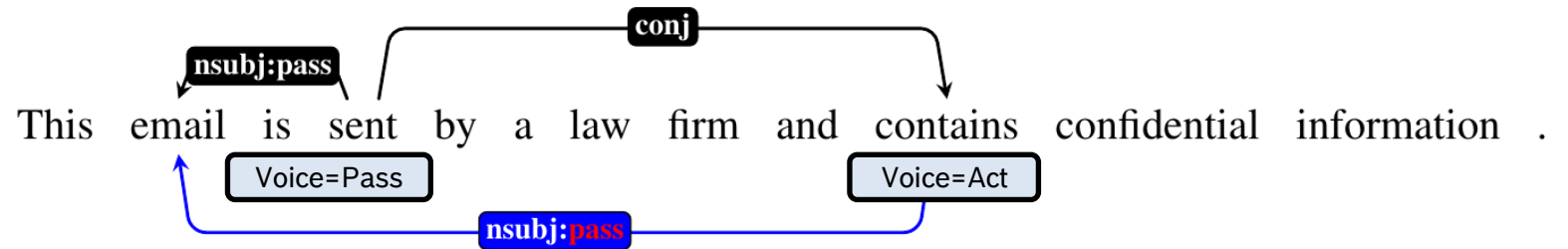
Imperatives:
12 times



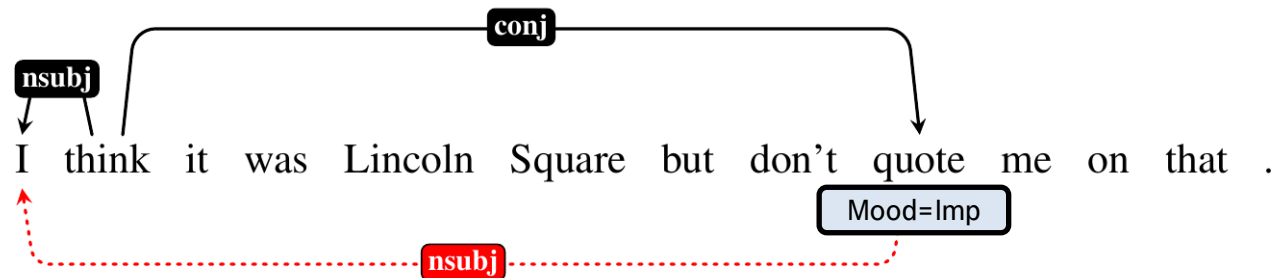
Corpus Annotation of Coordinate Structures

Analysis: Examples

Passive Propagation:
18 times



Imperatives:
12 times

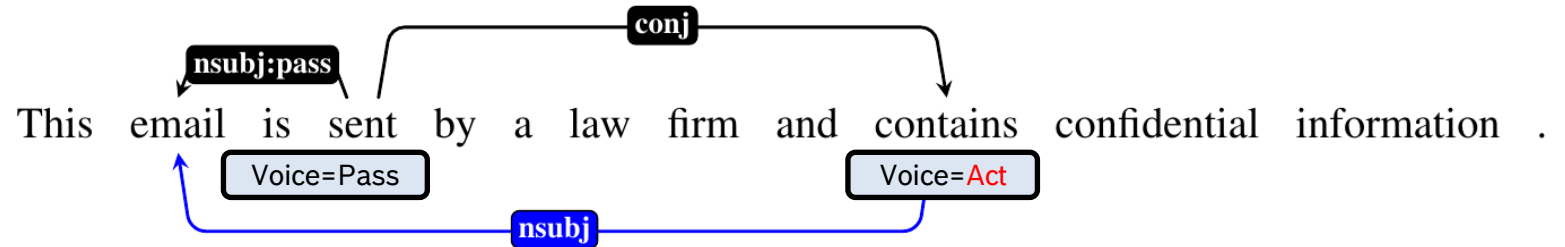


Proposed fix: take into account **morphosyntactic features** of words

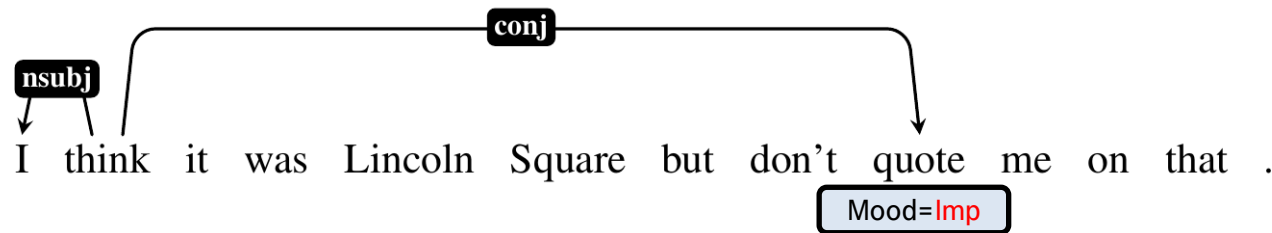
Corpus Annotation of Coordinate Structures

Analysis: Examples

Passive Propagation:
18 times



Imperatives:
12 times

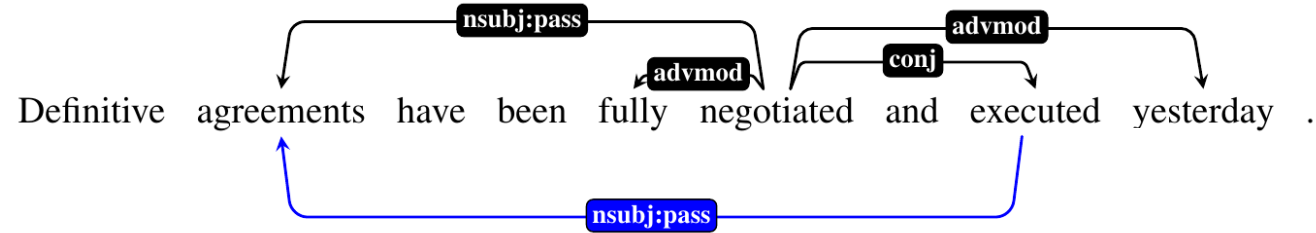
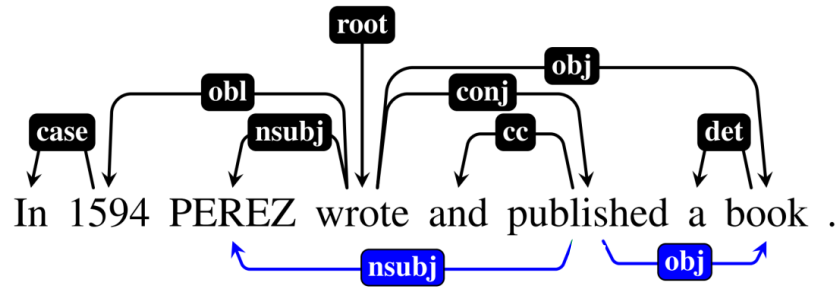


Proposed fix: take into account **morphosyntactic features** of words

Corpus Annotation of Coordinate Structures

Analysis: Examples

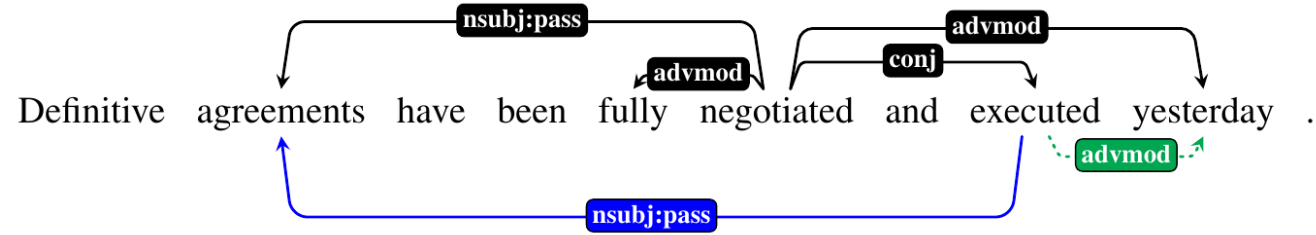
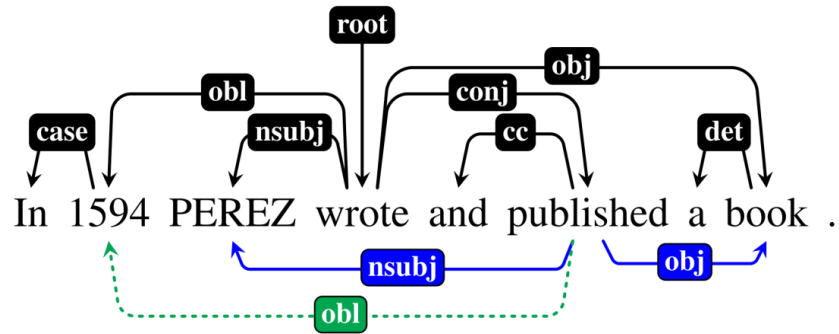
Non-core dependents:



Corpus Annotation of Coordinate Structures

Analysis: Examples

Non-core dependents:

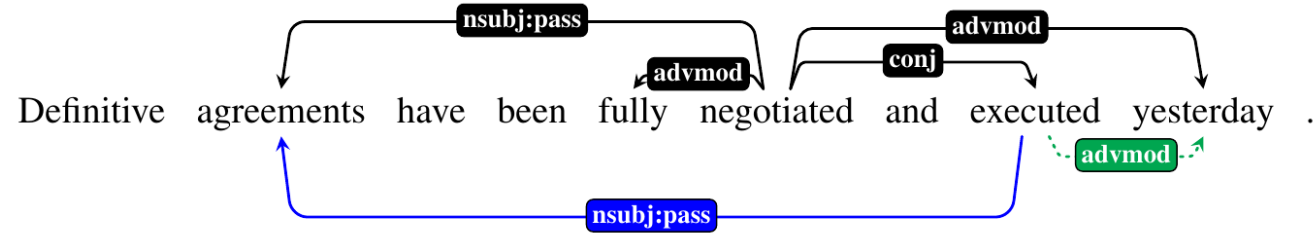
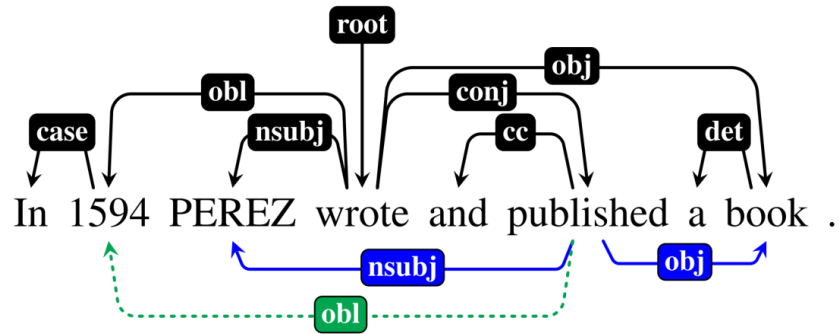


- We propose propagation of non-core dependents: extend propagation to such relations by the RBC

Corpus Annotation of Coordinate Structures

Analysis: Examples

Non-core dependents:

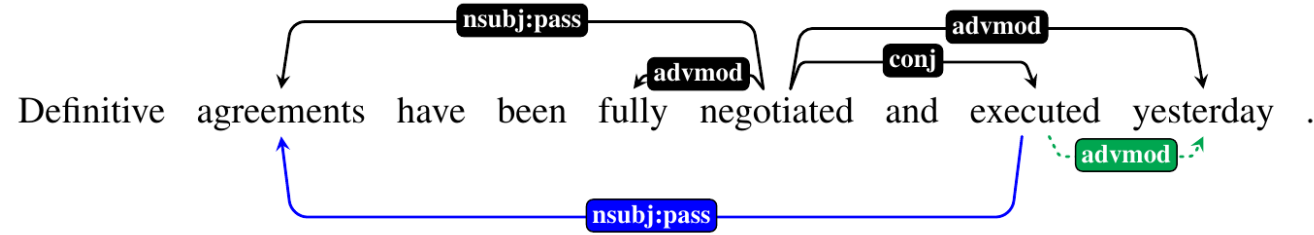
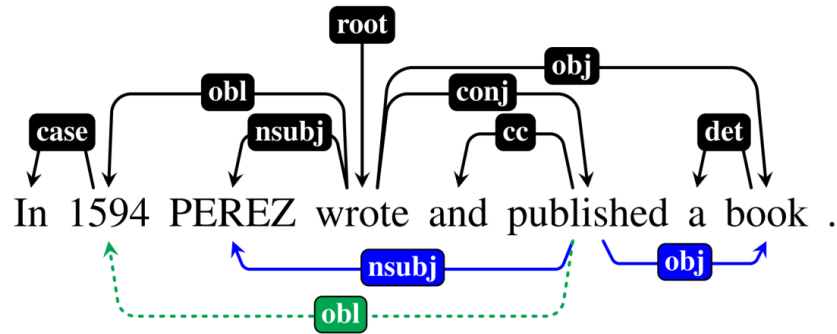


- We propose propagation of non-core dependents: extend propagation to such relations by the RBC
- This phenomenon requires *semantic* information for disambiguation when automatically propagating

Corpus Annotation of Coordinate Structures

Analysis: Examples

Non-core dependents:

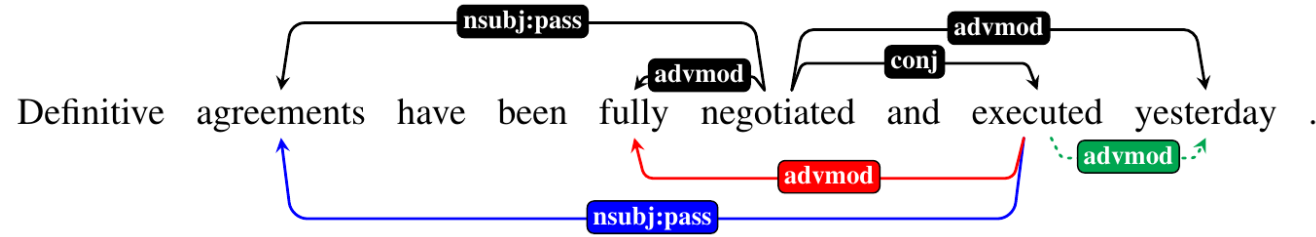
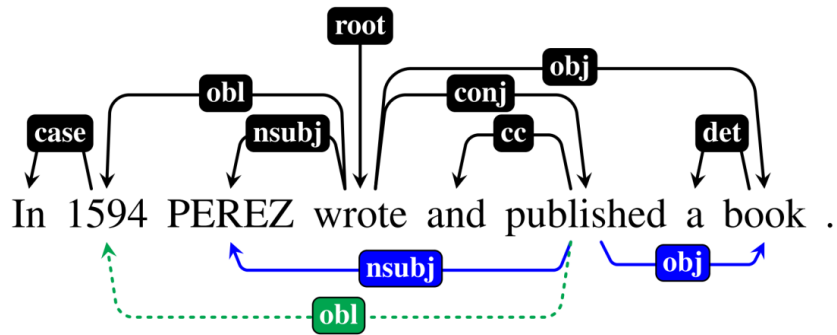


- We propose propagation of non-core dependents: extend propagation to such relations by the RBC
- This phenomenon requires *semantic* information for disambiguation when automatically propagating
- Basic layer partially disambiguates these cases: propagation if dep comes after 2nd conjunct

Corpus Annotation of Coordinate Structures

Analysis: Examples

Non-core dependents:



- We propose propagation of non-core dependents: extend propagation to such relations by the RBC
- This phenomenon requires *semantic* information for disambiguation when automatically propagating
- Basic layer partially disambiguates these cases: propagation if dep comes after 2nd conjunct
- We keep this syntactic constraint in the RBC: requires **human annotator** to make the decision

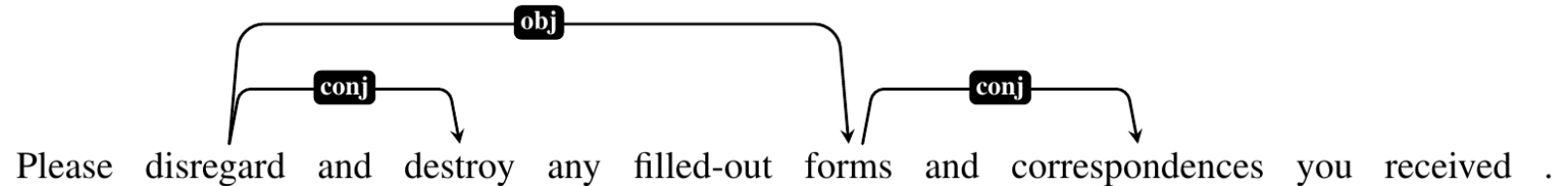
Corpus Annotation of Coordinate Structures

Analysis: Examples

Multiple coordinations:

- „single pass“ over the dependencies in the sentence
- **relations that result from a propagation cannot themselves be propagated!**

Input sentence:



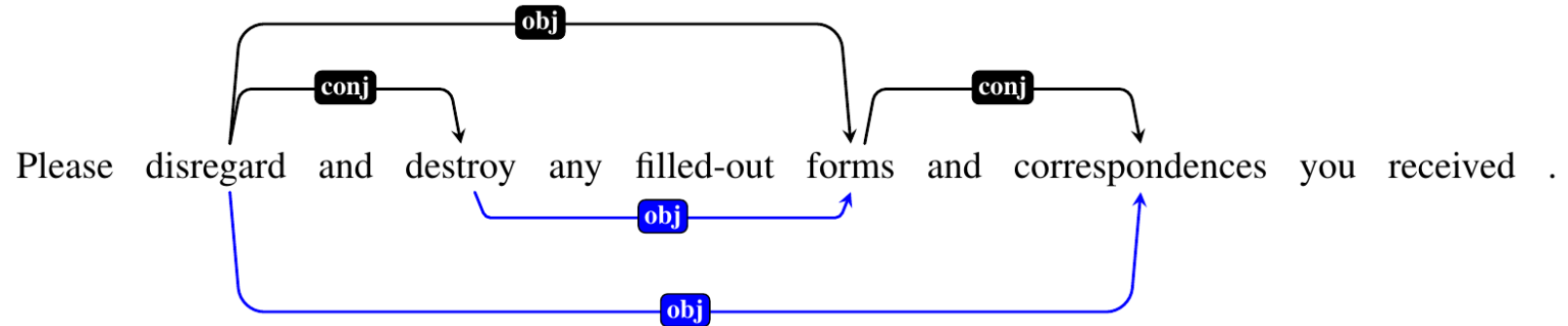
Corpus Annotation of Coordinate Structures

Analysis: Examples

Multiple coordinations:

- „single pass“ over the dependencies in the sentence
- **relations that result from a propagation cannot themselves be propagated!**

After 1st pass:



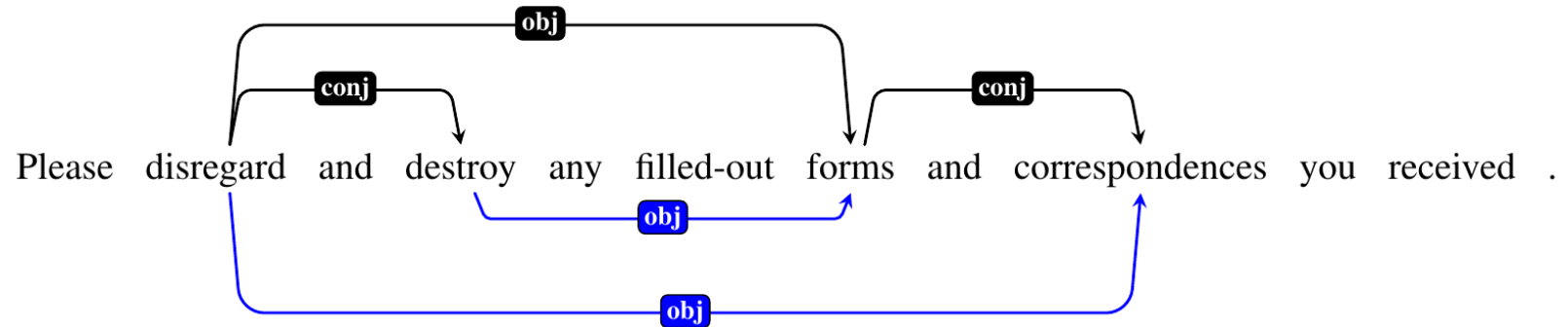
Corpus Annotation of Coordinate Structures

Analysis: Examples

Multiple coordinations:

- „single pass“ over the dependencies in the sentence
- **relations that result from a propagation cannot themselves be propagated!**

After 1st pass:



Proposed fix: Repeat the propagation process until the graph does not change anymore

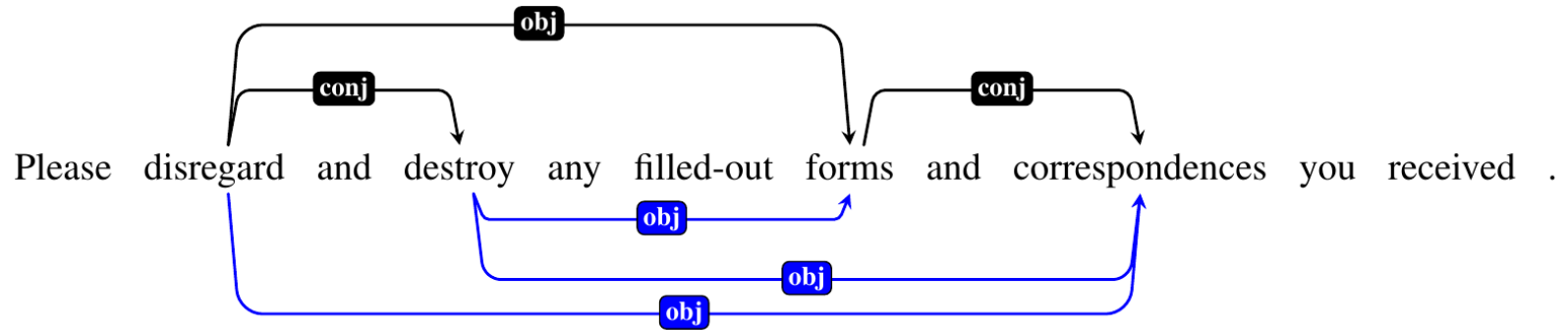
Corpus Annotation of Coordinate Structures

Analysis: Examples

Multiple coordinations:

- „single pass“ over the dependencies in the sentence
- **relations that result from a propagation cannot themselves be propagated!**

After 2nd pass:



Proposed fix: Repeat the propagation process until the graph does not change anymore

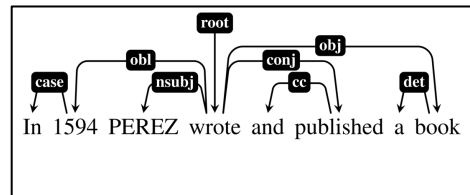
Computational Modeling of Coordinate Structures

Experiments: Retrieval of Propagated Relations

Computational Modeling of Coordinate Structures

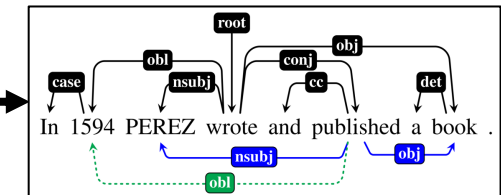
Experiments: Retrieval of Propagated Relations

Gold basic annotations



*Rule-based
or ML-based
converters*

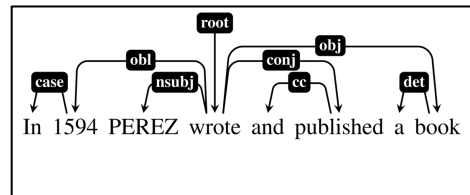
Enhanced annotations



Computational Modeling of Coordinate Structures

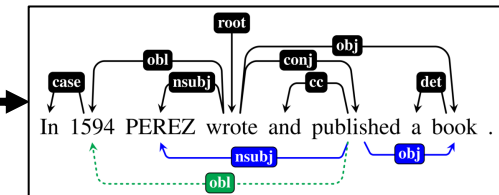
Experiments: Retrieval of Propagated Relations

Gold basic annotations

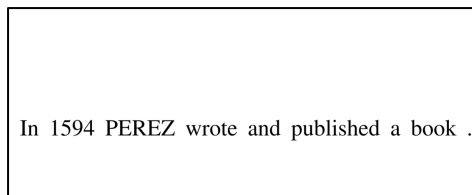


*Rule-based
or ML-based
converters*

Enhanced annotations

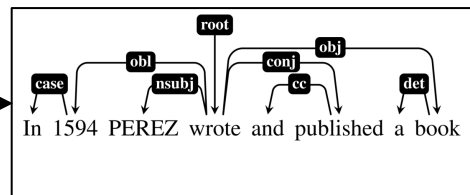


Raw tokens

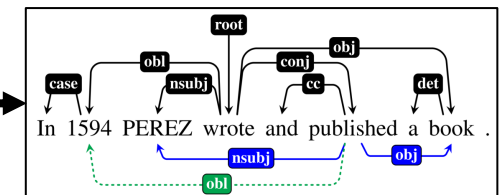


UDify

Predicted basic annotations



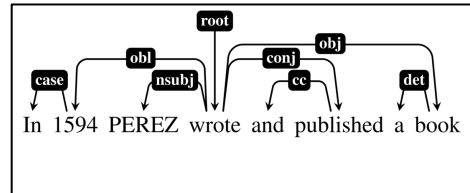
Enhanced annotations



Computational Modeling of Coordinate Structures

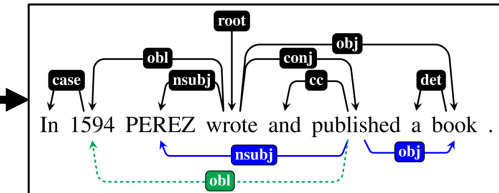
Experiments: Retrieval of Propagated Relations

Gold basic annotations

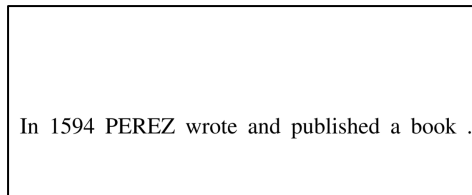


*Rule-based
or ML-based
converters*

Enhanced annotations

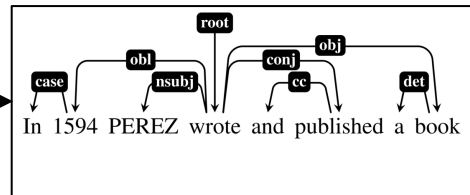


Raw tokens

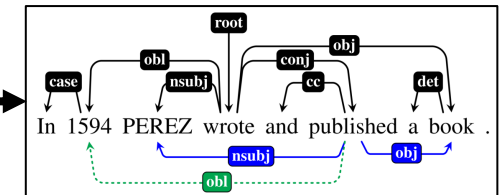


UDify

Predicted basic annotations



Enhanced annotations

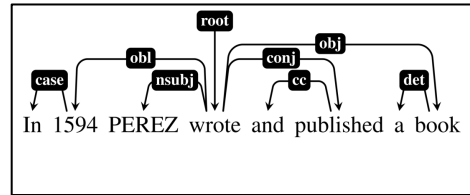


Graph-based parser

Computational Modeling of Coordinate Structures

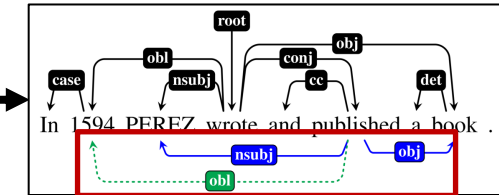
Experiments: Retrieval of Propagated Relations

Gold basic annotations

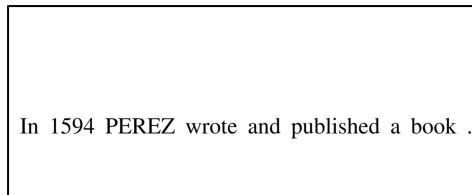


Rule-based
or ML-based
converters

Enhanced annotations

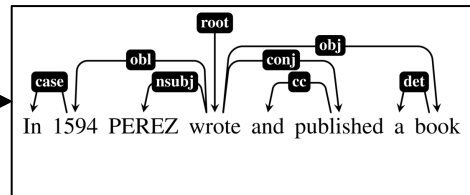


Raw tokens

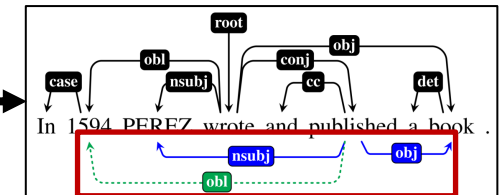


UDify

Predicted basic annotations



Enhanced annotations



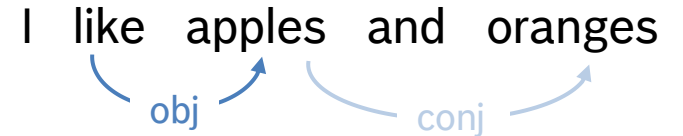
Graph-based parser

Evaluation metrics: Precision, recall, F-score on relations resulting from propagation comparing against our **corrected** dev and test sets

Computational Modeling of Coordinate Structures

ML-based Classifiers

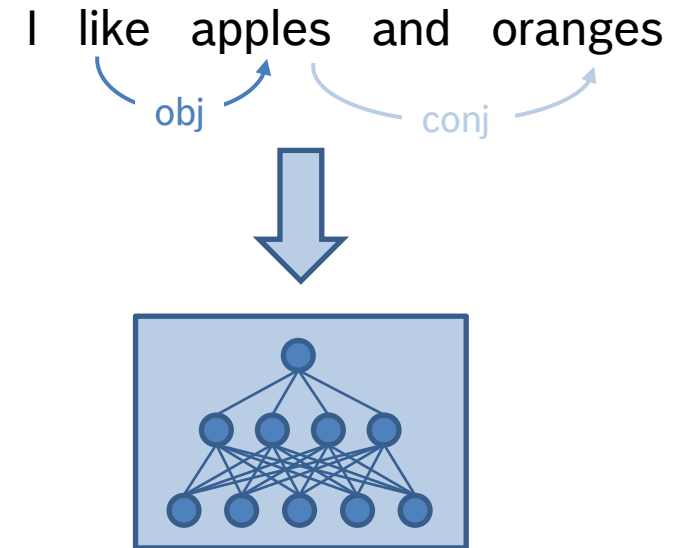
- Use machine learning to learn whether dependencies are propagated or not
- SVM (following [Nyblom et al., 2013](#)) and NN-based classifiers (own implementation, multilayer perceptron)
- Features:
 - **Instance features:** Dependency label, incoming vs. outgoing dependency
 - **Tree features:** E.g. linear dependency direction, number of items in the coordination
 - **Token features:** Morphological features (SVM) or RoBERTa embeddings (NN)



Computational Modeling of Coordinate Structures

ML-based Classifiers

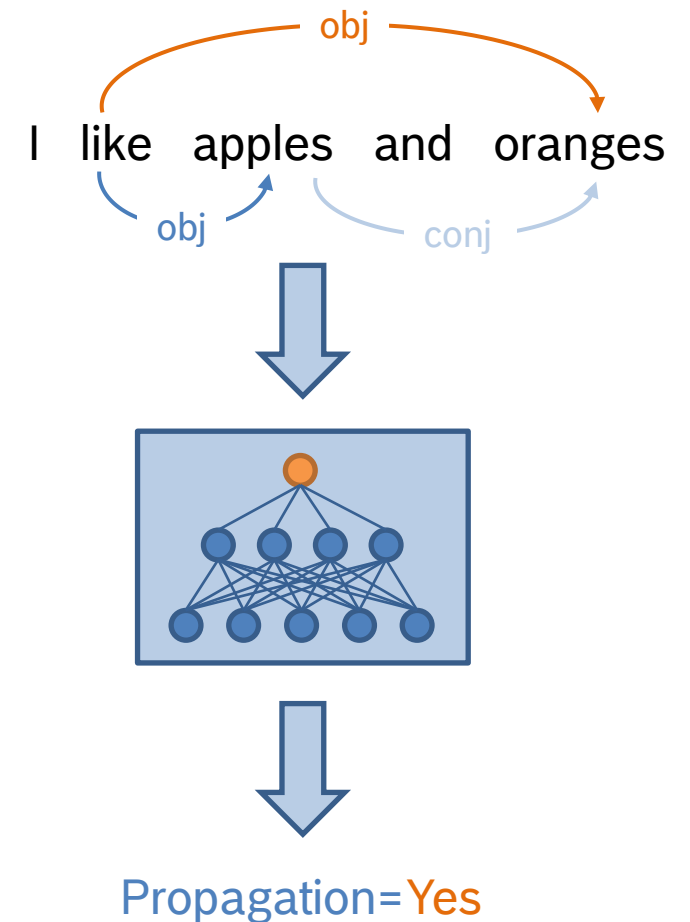
- Use machine learning to learn whether dependencies are propagated or not
- SVM (following [Nyblom et al., 2013](#)) and NN-based classifiers (own implementation, multilayer perceptron)
- Features:
 - **Instance features:** Dependency label, incoming vs. outgoing dependency
 - **Tree features:** E.g. linear dependency direction, number of items in the coordination
 - **Token features:** Morphological features (SVM) or RoBERTa embeddings (NN)



Computational Modeling of Coordinate Structures

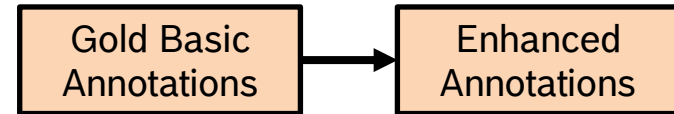
ML-based Classifiers

- Use machine learning to learn whether dependencies are propagated or not
- SVM (following [Nyblom et al., 2013](#)) and NN-based classifiers (own implementation, multilayer perceptron)
- Features:
 - **Instance features:** Dependency label, incoming vs. outgoing dependency
 - **Tree features:** E.g. linear dependency direction, number of items in the coordination
 - **Token features:** Morphological features (SVM) or RoBERTa embeddings (NN)



Computational Modeling of Coordinate Structures

Results (Gold Basic Dependencies)

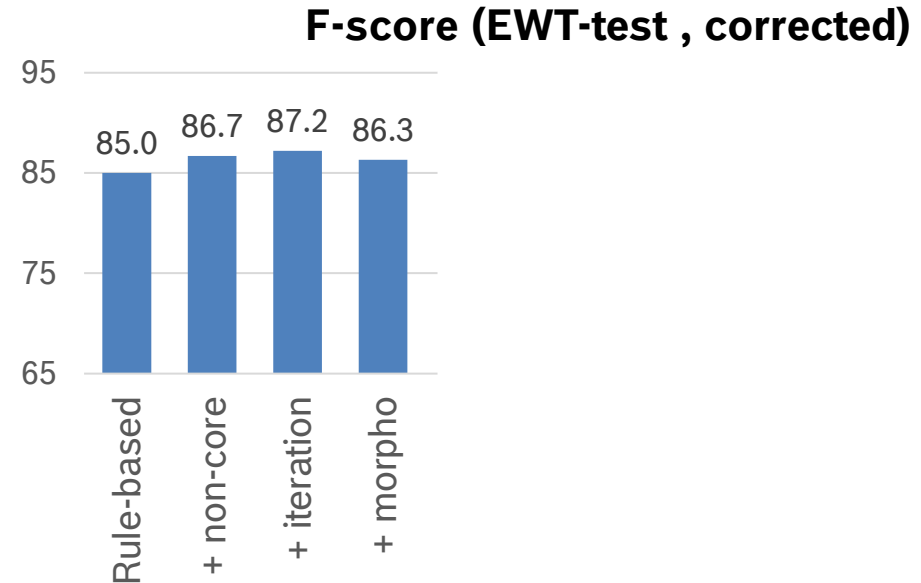
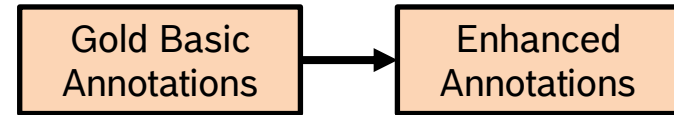


Computational Modeling of Coordinate Structures

Results (Gold Basic Dependencies)

- Rule-based converter modifications yield slight improvements at best:
 - Sensitive to basic-layer annotation errors

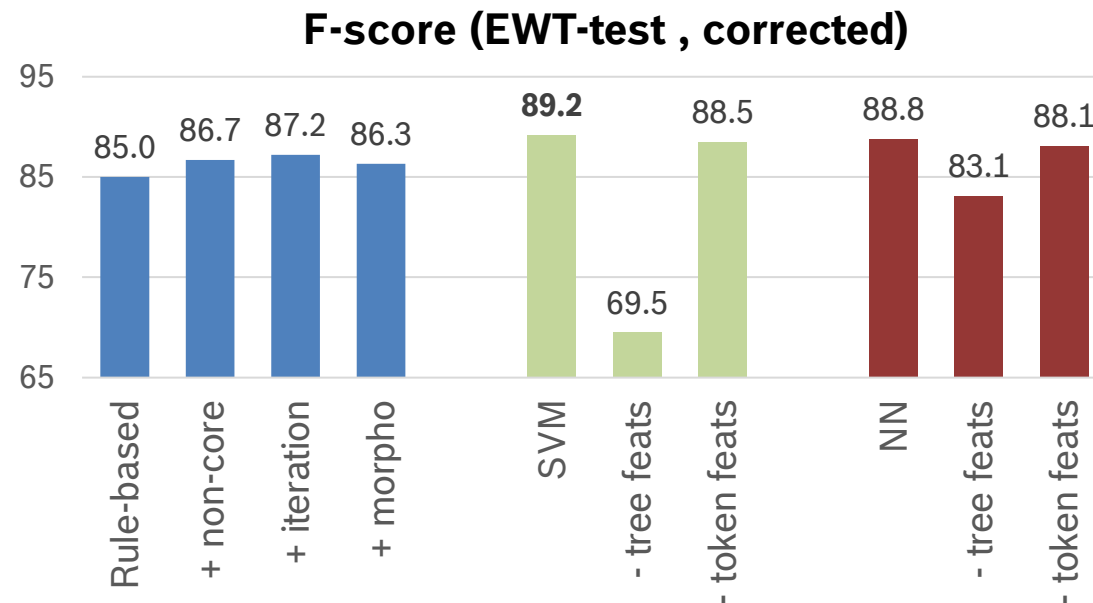
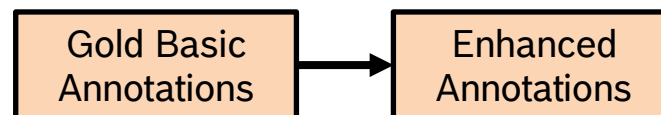
[More details → paper](#)



Computational Modeling of Coordinate Structures

Results (Gold Basic Dependencies)

- Rule-based converter modifications yield slight improvements at best:
 - Sensitive to basic-layer annotation errors
- More details → [paper](#)
- SVM and NN perform similarly
 - Outperforming rule-based converter on EWT-test
 - Both rely heavily on tree-based features, but the NN less so



Computational Modeling of Coordinate Structures

Graph-Parser Prediction

Computational Modeling of Coordinate Structures

Graph-Parser Prediction

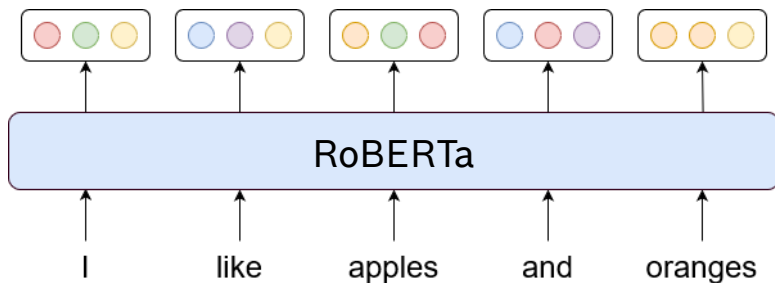
- Predict enhanced dependencies directly
- Essentially graph-based dependency parsing for basic+enhanced layer

Computational Modeling of Coordinate Structures

Graph-Parser Prediction

- Predict enhanced dependencies directly
- Essentially graph-based dependency parsing for basic+enhanced layer
- **Neural architecture:** Biaffine classifier on top of fine-tuned RoBERTa embeddings

RobertNLP (Grünwald & Friedrich, 2020) for details



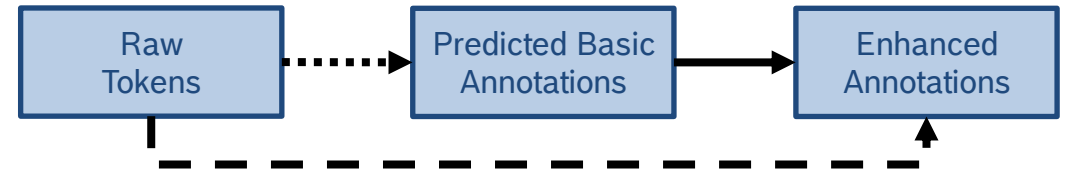
	I	like	apples	and	oranges
I					
like	\emptyset	\emptyset	\emptyset	\emptyset	\emptyset
apples	\emptyset	\emptyset	\emptyset	\emptyset	\emptyset
and	\emptyset	\emptyset	\emptyset	\emptyset	\emptyset
oranges	\emptyset	\emptyset	\emptyset	\emptyset	\emptyset

Enhanced dependencies (row labels):

- like: **nsubj**
- apples: **obj**
- and: **conj**
- oranges: **cc**

Computational Modeling of Coordinate Structures

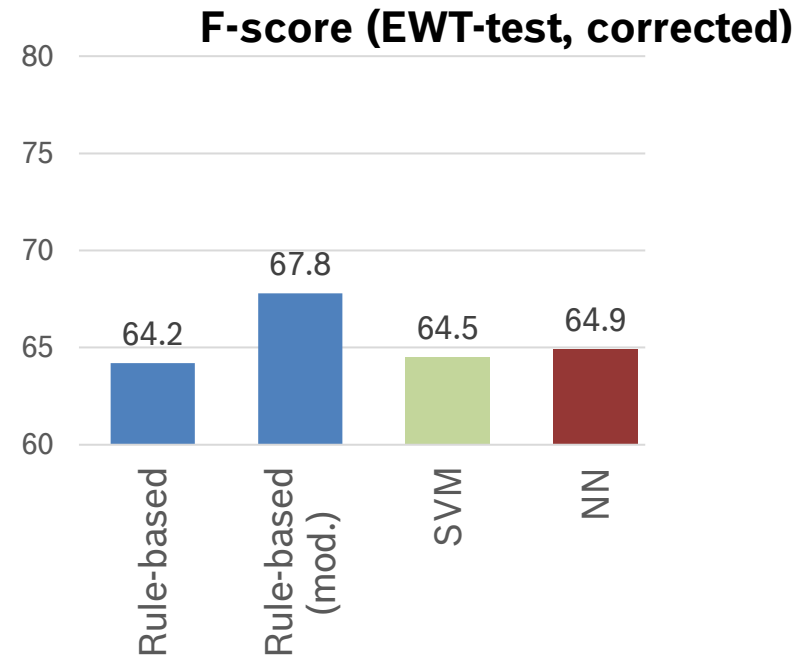
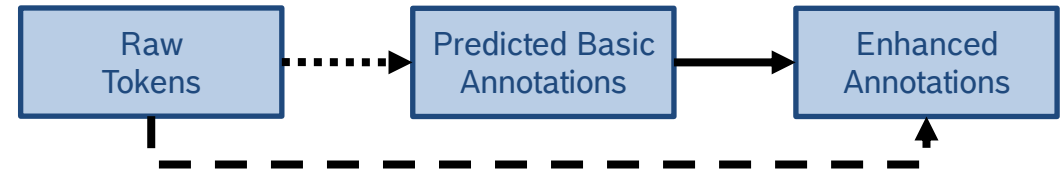
Results (Raw Tokens)



Computational Modeling of Coordinate Structures

Results (Raw Tokens)

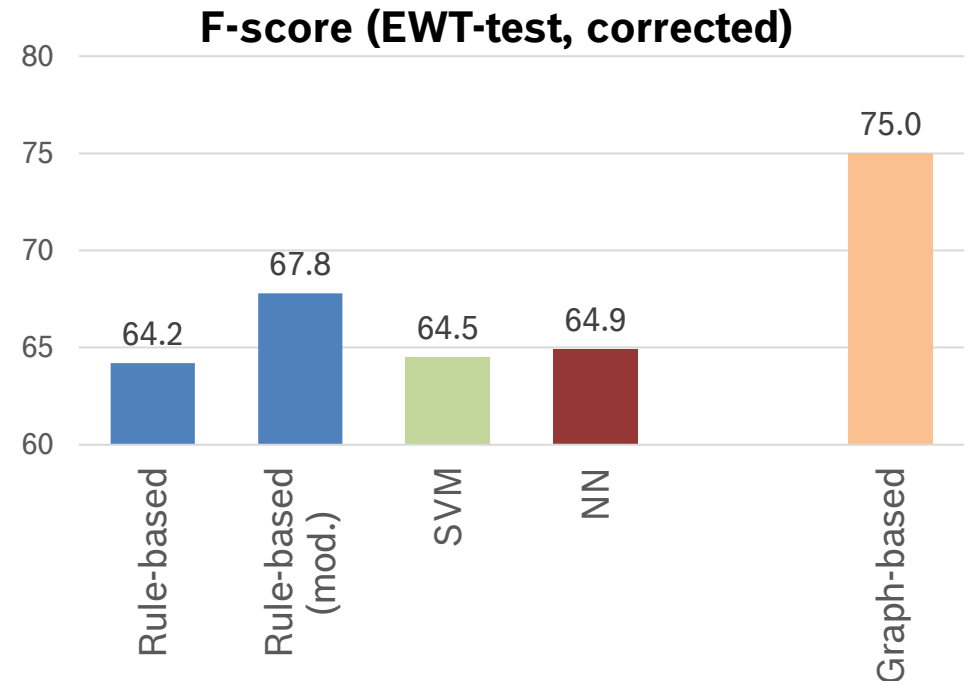
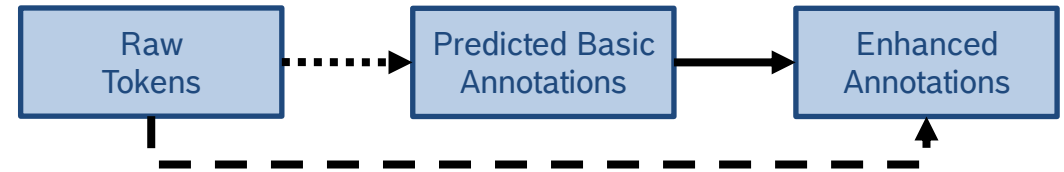
- Pipeline-based systems show rather poor performance
 - Our modified rule-based converter performs best among pipeline systems



Computational Modeling of Coordinate Structures

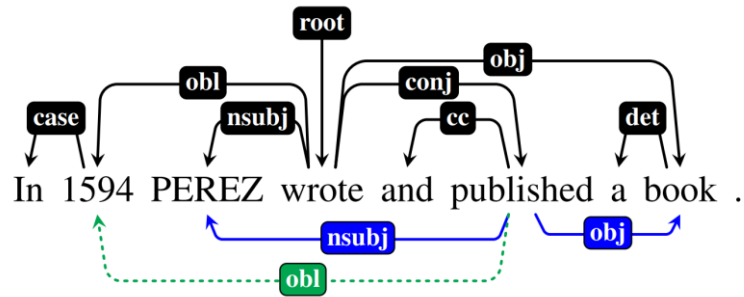
Results (Raw Tokens)

- Pipeline-based systems show rather poor performance
 - Our modified rule-based converter performs best among pipeline systems
- Our graph-based parser performs best by a wide margin
 - Implicit modeling of conjunction propagation avoids error propagation from the basic layer

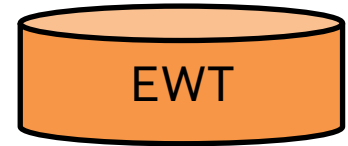


Contributions and Conclusion

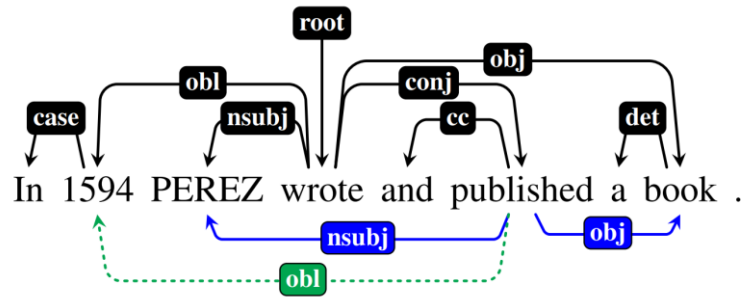
Contributions and Conclusion



First manually validated dataset for **conjunction propagation** in English Enhanced UD

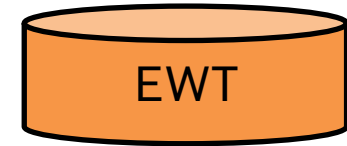


Contributions and Conclusion

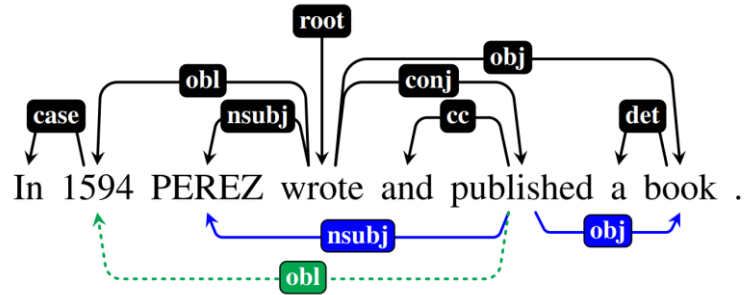


First manually validated dataset for **conjunction propagation** in English Enhanced UD

Semantic annotation task,
high inter-annotator agreement

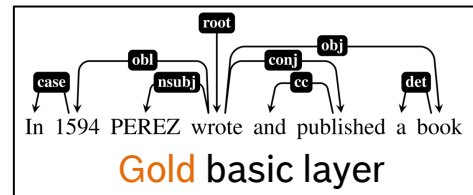
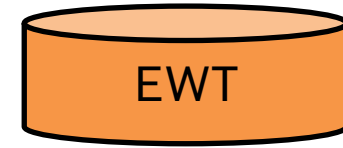


Contributions and Conclusion

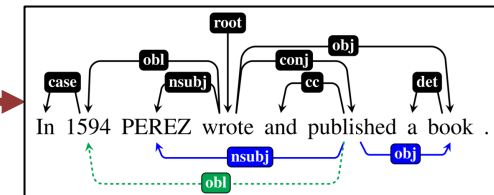


First manually validated dataset for **conjunction propagation** in English Enhanced UD

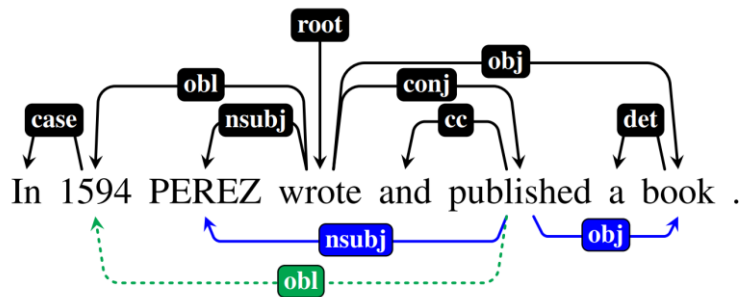
Semantic annotation task, high inter-annotator agreement



Rule-based
or ML-based
converters

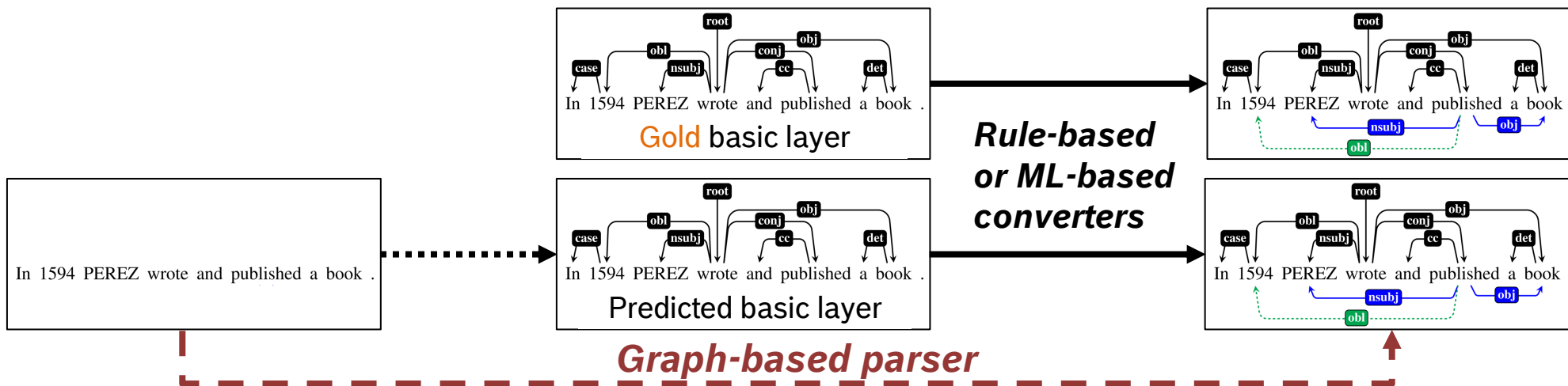
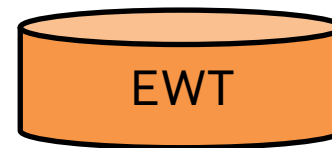


Contributions and Conclusion

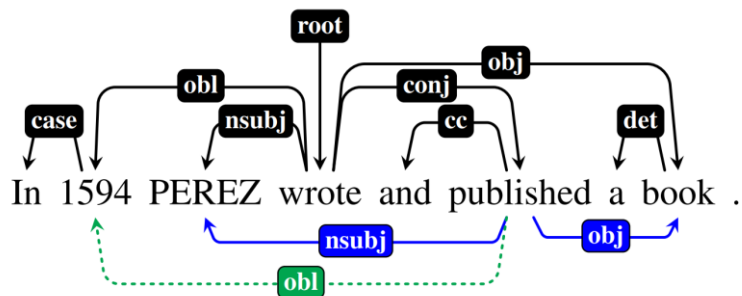


First manually validated dataset for **conjunction propagation** in English Enhanced UD

Semantic annotation task, high inter-annotator agreement

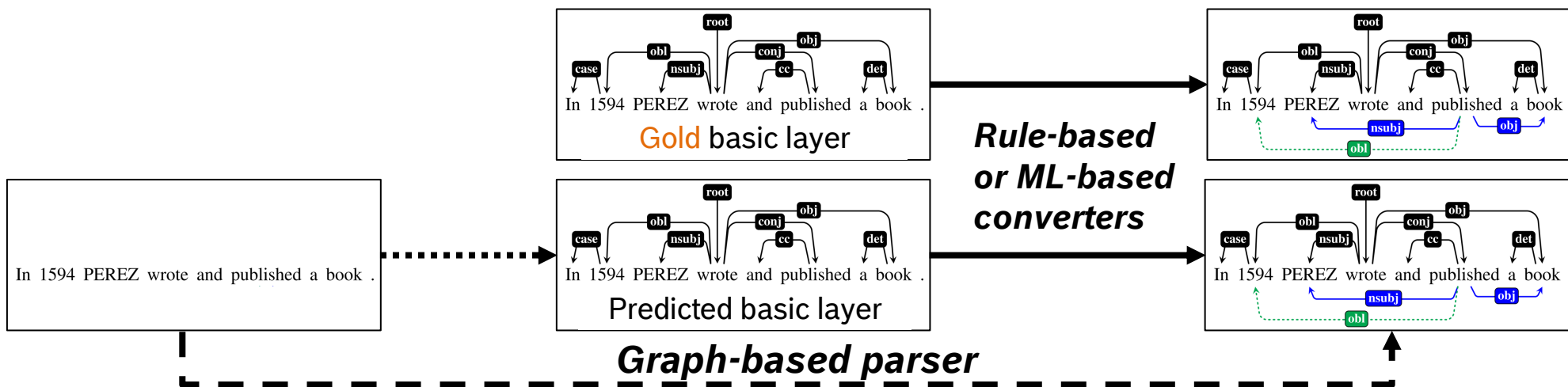
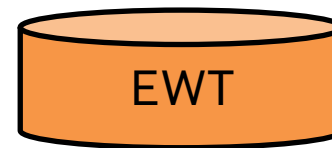


Contributions and Conclusion



First manually validated dataset for **conjunction propagation** in English Enhanced UD

Semantic annotation task, high inter-annotator agreement



https://github.com/boschresearch/coordinate_constructions_english_enhanced_ud_eacl2021