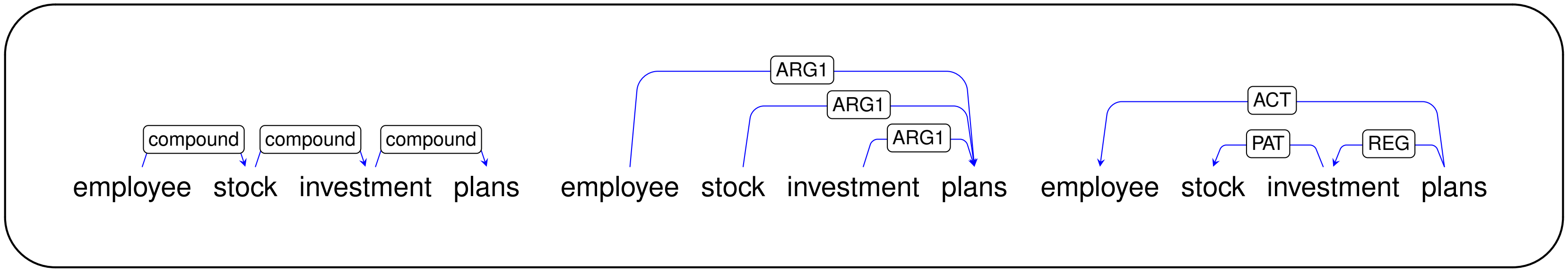



# SDP 2014 (SemEval Task 8)

## Broad-Coverage Semantic Dependency Parsing

**Stephan Oepen**
Universitetet i Oslo & Universität Potsdam

Marco Kuhlmann, Yusuke Miyao, Daniel Zeman,
Dan Flickinger, Jan Hajič, Angelina Ivanova, Yi Zhang

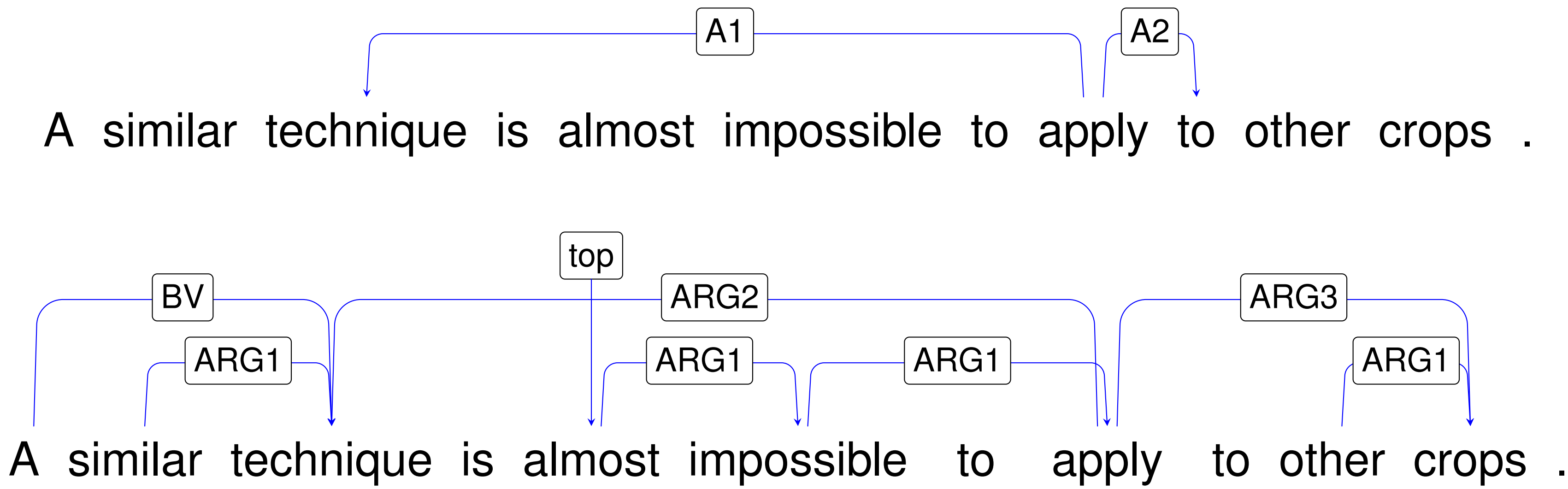
sdp-organizers@emmtee.net

# By Way of Introduction: Semantic Dependencies

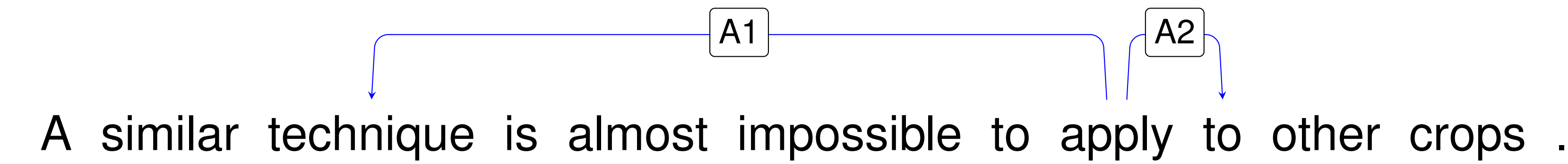
A1 A2

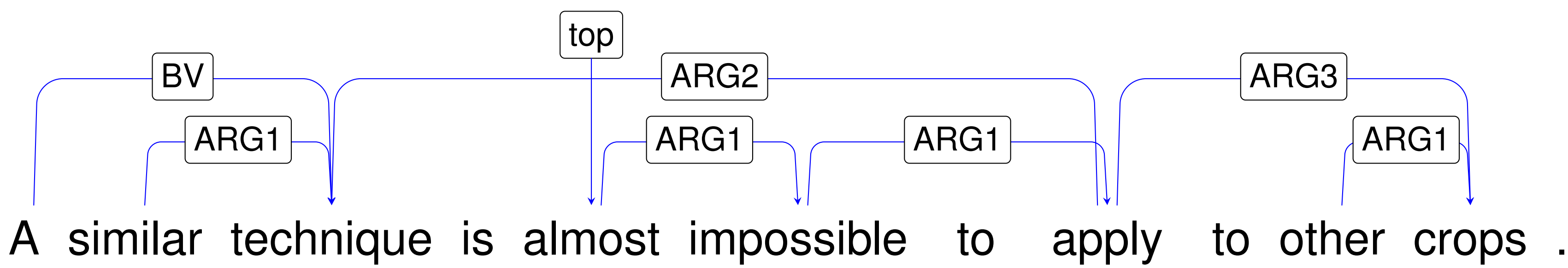
A similar technique is almost impossible to apply to other crops .

# By Way of Introduction: Semantic Dependencies



# By Way of Introduction: Semantic Dependencies




### High-Level Linguistic and Formal Properties

- Core *semantic* predicate–argument structure, or ‘Who did What to Whom?’
- argument sharing: graph re-entrancies; vacuous words: unattached nodes;
- designated *top* node (not root): semantic head, highest-scoping predicate.

# Task Definition, Goals, and High Hopes

> *We define* BROAD-COVERAGE SEMANTIC DEPENDENCY PARSING *(SDP) as the task of recovering sentence-internal predicate–argument relationships for* ALL CONTENT WORDS*, i.e. the semantic structure constituting the relational core of sentence meaning.*
>
> [2014 Task Description]

# Task Definition, Goals, and High Hopes

*We define* BROAD-COVERAGE SEMANTIC DEPENDENCY PARSING *(SDP) as the task of recovering sentence-internal predicate–argument relationships for* ALL CONTENT WORDS*, i.e. the semantic structure constituting the relational core of sentence meaning.*
[2014 Task Description]

**Push Dependency Parsing Towards Directed Graphs**

- Higher degree of abstraction: deep syntax or predicate–argument structure;
- e.g. shared arguments (control, relative clauses); vacuous word classes.

**Define 'Semantic Role Labeling' for All Content Words**

- Argument labeling for phenomena like negation, comparatives, possessives.

# Three Parallel Annotations of the WSJ Corpus

**DM: DELPH-IN MRS-Derived Bi-Lexical Dependencies**

- DeepBank: Fresh HPSG-style annotation, including logical-form semantics;
- ‘lossy’ reduction of MRS meaning representations to bi-lexical dependencies.

# Three Parallel Annotations of the WSJ Corpus

**DM: DELPH-IN MRS-Derived Bi-Lexical Dependencies**

- DeepBank: Fresh HPSG-style annotation, including logical-form semantics;
- 'lossy' reduction of MRS meaning representations to bi-lexical dependencies.

**PAS: Enju Predicate–Argument Structures**

- Enju Treebank: Projection of (complete) PTB syntax to HPSG derivations;
- semantic analyses take form of lexicalized predicate–argument structures.

# Three Parallel Annotations of the WSJ Corpus

**DM: DELPH-IN MRS-Derived Bi-Lexical Dependencies**

- DeepBank: Fresh HPSG-style annotation, including logical-form semantics;
- ‘lossy’ reduction of MRS meaning representations to bi-lexical dependencies.

**PAS: Enju Predicate–Argument Structures**

- Enju Treebank: Projection of (complete) PTB syntax to HPSG derivations;
- semantic analyses take form of lexicalized predicate–argument structures.

**PCEDT: Parts of the Prague Tectogrammatical Layer**

- Include all nodes from Prague t-trees that correspond to surface tokens;
- re-attach functors of generated nodes; project dependencies to conjuncts.

# Three Parallel Annotations of the WSJ Corpus

**DM: DELPH-IN MRS-Derived Bi-Lexical Dependencies**

- DeepBank: Fresh HPSG-style annotation, including logical-form semantics;
- ‘lossy’ reduction of MRS meaning representations to bi-lexical dependencies.

**PAS: Enju Predicate–Argument Structures**

- Enju Treebank: Projection of (complete) PTB syntax to HPSG derivations;
- semantic analyses take form of lexicalized predicate–argument structures.

**PCEDT: Parts of the Prague Tectogrammatical Layer**

- Include all nodes from Prague t-trees that correspond to surface tokens;
- re-attach functors of generated nodes; project dependencies to conjuncts.

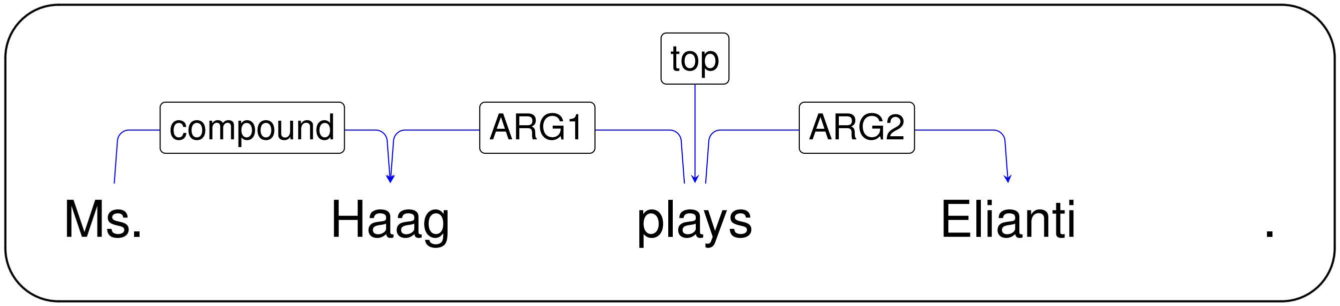
Sections 00–20 for Training (745,543 Tokens); Section 21 for Testing (29,808).

# Token-Oriented, Tabular Data Format

| *id* | *form* | *lemma* | *pos* | *top* | *pred* | *arg1* | *arg2* |
|---|---|---|---|---|---|---|---|
| #20200002 | | | | | | | |
| 1 | Ms. | Ms. | NNP | − | + | _ | _ |
| 2 | Haag | Haag | NNP | − | − | compound | ARG1 |
| 3 | plays | play | VBZ | + | + | _ | _ |
| 4 | Elianti | Elianti | NNP | − | − | _ | ARG2 |
| 5 | . | . | . | − | − | _ | _ |

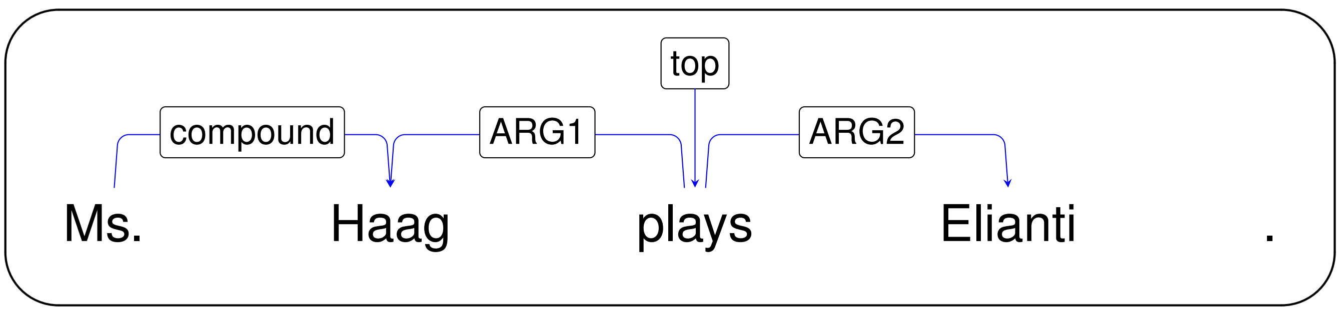
# Token-Oriented, Tabular Data Format

| *id* | *form* | *lemma* | *pos* | *top* | *pred* | *arg1* | *arg2* |
|---|---|---|---|---|---|---|---|
| #20200002 | | | | | | | |
| 1 | Ms. | Ms. | NNP | − | + | _ | _ |
| 2 | Haag | Haag | NNP | − | − | compound | ARG1 |
| 3 | plays | play | VBZ | + | + | _ | _ |
| 4 | Elianti | Elianti | NNP | − | − | _ | ARG2 |
| 5 | . | . | . | − | − | _ | _ |



# Token-Oriented, Tabular Data Format

| *id* | *form* | *lemma* | *pos* | *top* | *pred* | *arg1* | *arg2* |
|---|---|---|---|---|---|---|---|
| #20200002 | | | | | | | |
| 1 | Ms. | Ms. | NNP | − | + | _ | _ |
| 2 | Haag | Haag | NNP | − | − | compound | ARG1 |
| 3 | plays | play | VBZ | + | + | _ | _ |
| 4 | Elianti | Elianti | NNP | − | − | _ | ARG2 |
| 5 | . | . | . | − | − | _ | _ |




- Sub-set of fields from CoNLL 2009; simplified `pred` column, added `top`;

→ generic data format: labeled directed graphs with designated top node(s).

# Quantitative Comparison of Target Representations

| | | DM | PAS | PCEDT |
|---|---|---|---|---|
| (1) | **# labels** | 51 | 42 | 68 |
| (2) | **% singletons** | 22.62 | 4.49 | 35.79 |
| (3) | **# edge density** | 0.96 | 1.02 | 0.99 |
| (4) | **$\%_g$ trees** | 2.35 | 1.30 | 56.58 |
| (5) | **$\%_g$ projective** | 3.05 | 1.71 | 53.29 |
| (6) | **$\%_g$ fragmented** | 6.71 | 0.23 | 0.56 |
| (7) | **$\%_n$ reentrancies** | 27.35 | 29.40 | 9.27 |
| (8) | **$\%_g$ topless** | 0.28 | 0.02 | 0.00 |
| (9) | **# top nodes** | 0.9972 | 0.9998 | 1.1237 |
| (10) | **$\%_n$ non-top roots** | 44.71 | 55.92 | 4.36 |

# Quantitative Comparison of Target Representations

| | | DM | PAS | PCEDT |
|---|---|---|---|---|
| (1) | **# labels** | 51 | 42 | 68 |
| (2) | **% singletons** | 22.62 | 4.49 | 35.79 |
| (3) | **# edge density** | 0.96 | 1.02 | 0.99 |
| (4) | **$\%_g$ trees** | 2.35 | 1.30 | 56.58 |
| (5) | **$\%_g$ projective** | 3.05 | 1.71 | 53.29 |
| (6) | **$\%_g$ fragmented** | 6.71 | 0.23 | 0.56 |
| (7) | **$\%_n$ reentrancies** | 27.35 | 29.40 | 9.27 |
| (8) | **$\%_g$ topless** | 0.28 | 0.02 | 0.00 |
| (9) | **# top nodes** | 0.9972 | 0.9998 | 1.1237 |
| (10) | **$\%_n$ non-top roots** | 44.71 | 55.92 | 4.36 |

- PCEDT is most fine-grained in labels (1); also most 'tree-like' (4, 7, 10);
- PAS is most 'covering' (2) and most connected (3; ignoring singletons);
- DM has some structural red flags: fragmented and topless graphs (6, 8).

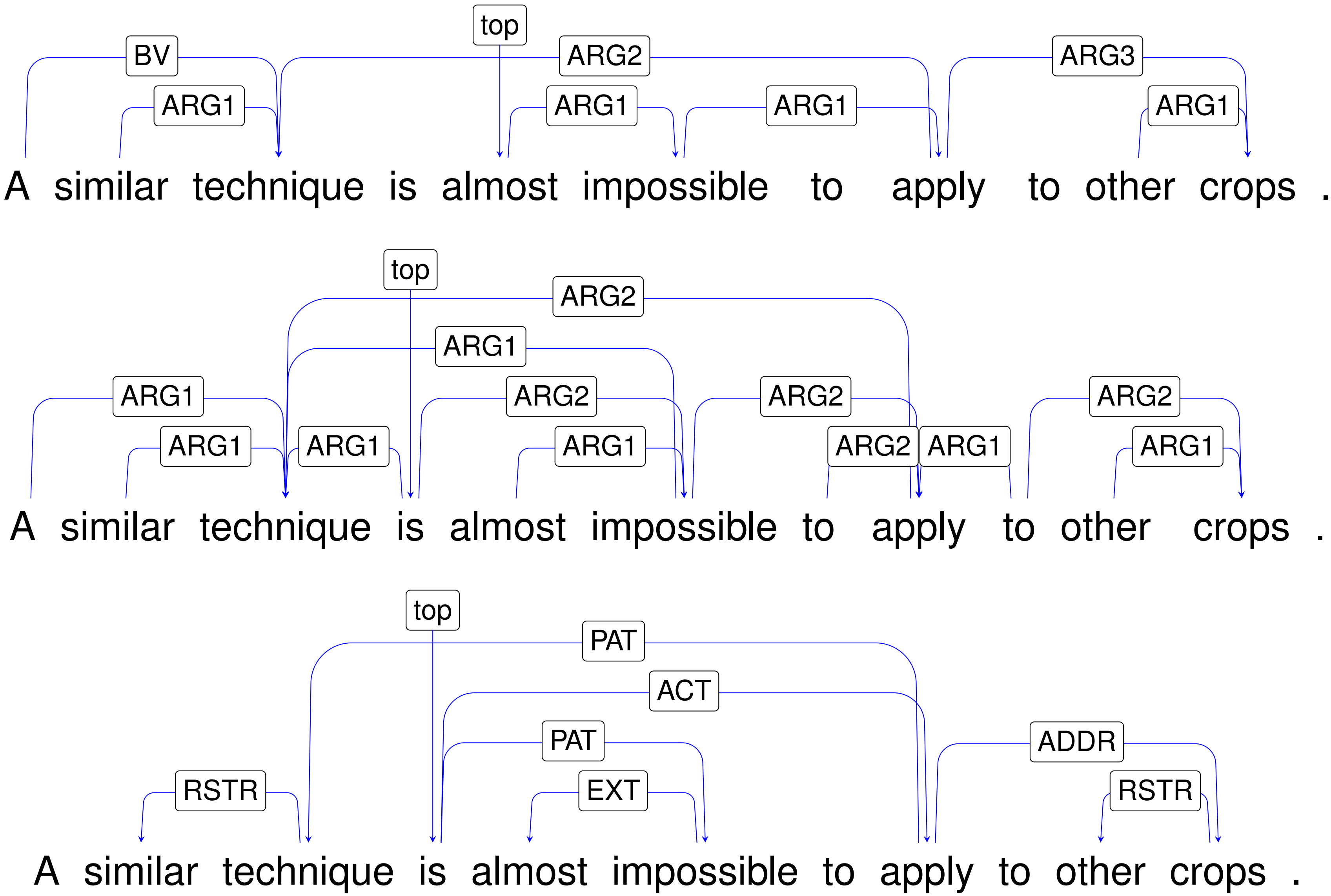
# Pairwise Similarity (Unlabeled Dependency $F_1$)

| | Directed | | | Undirected | | |
|---|---|---|---|---|---|---|
| | DM | PAS | PCEDT | DM | PAS | PCEDT |
| DM | – | .6425 | .2612 | – | .6719 | .5675 |
| PAS | .6688 | – | .2963 | .6993 | – | .5490 |
| PCEDT | .2636 | .2963 | – | .5743 | .5630 | – |

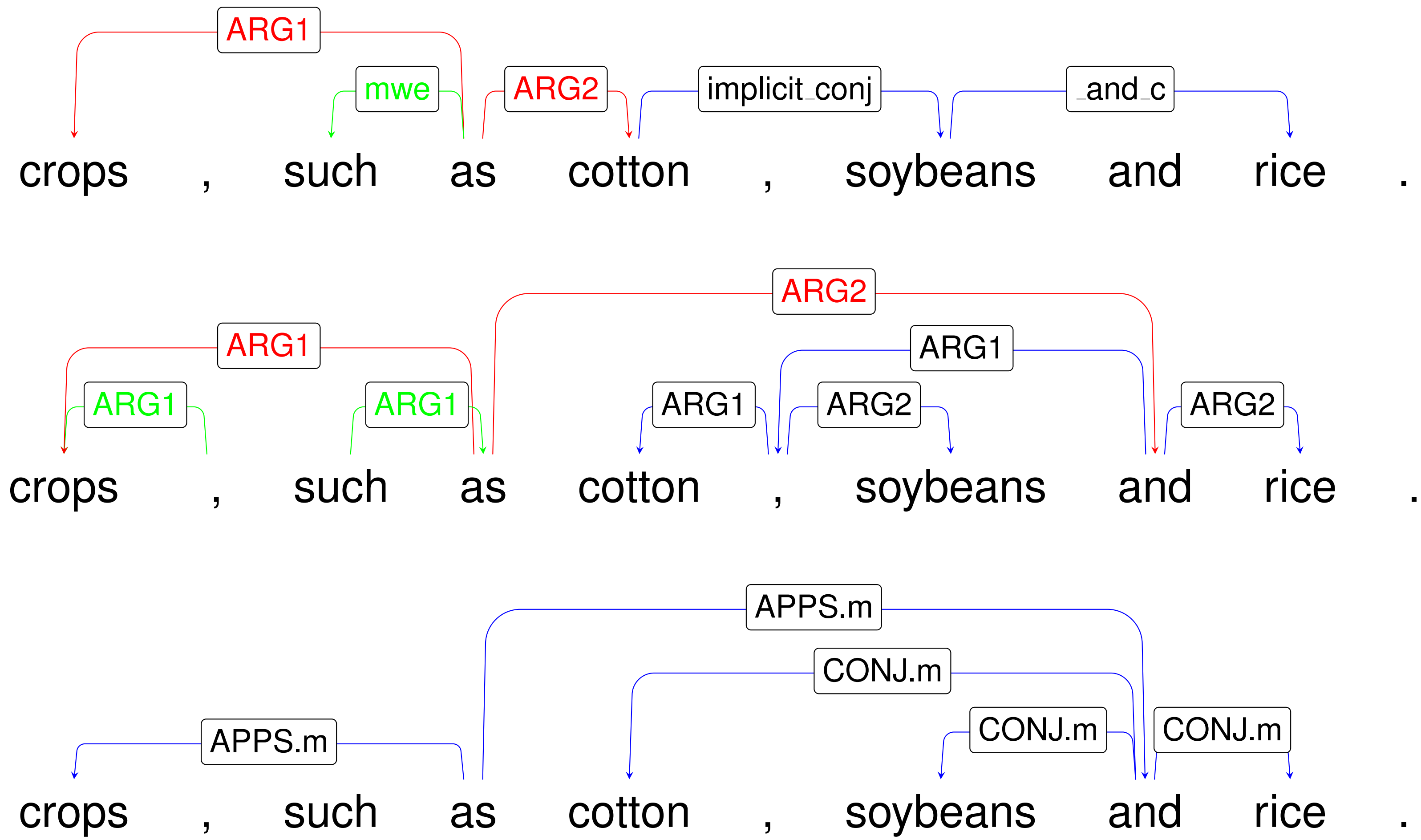
(Upper Right Diagonals: Including punctuation; Lower Left: Ignoring It)

- DM and PAS structurally much closer to each other than either to PCEDT;
- effect stronger when ignoring dependencies involving punctuation marks;
- directionality of dependencies one of the major sources of divergence.

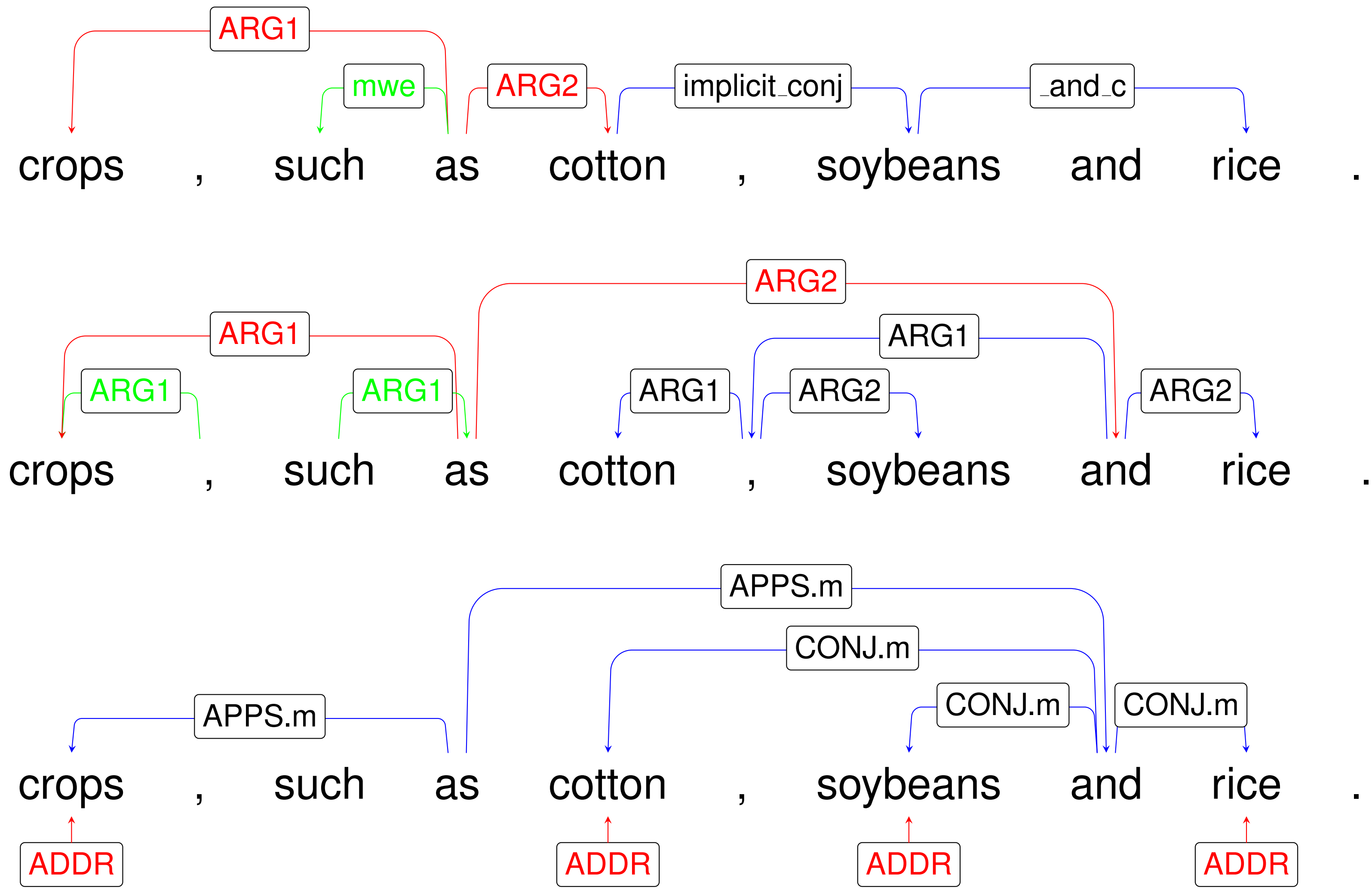
# Linguistic Comparison of Target Representations



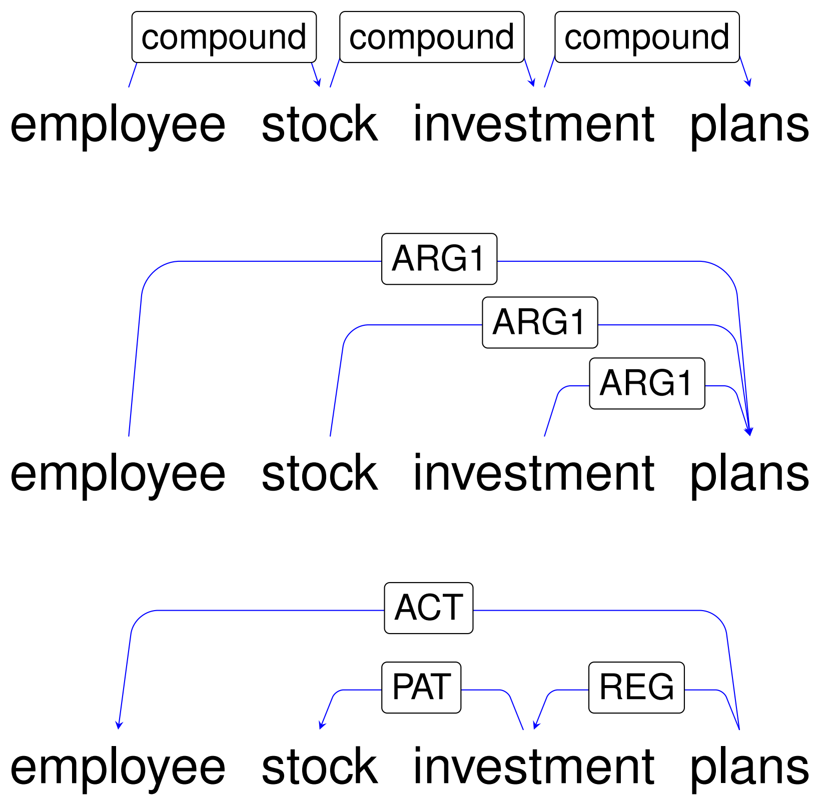
# Different Approaches to Coordinate Structures



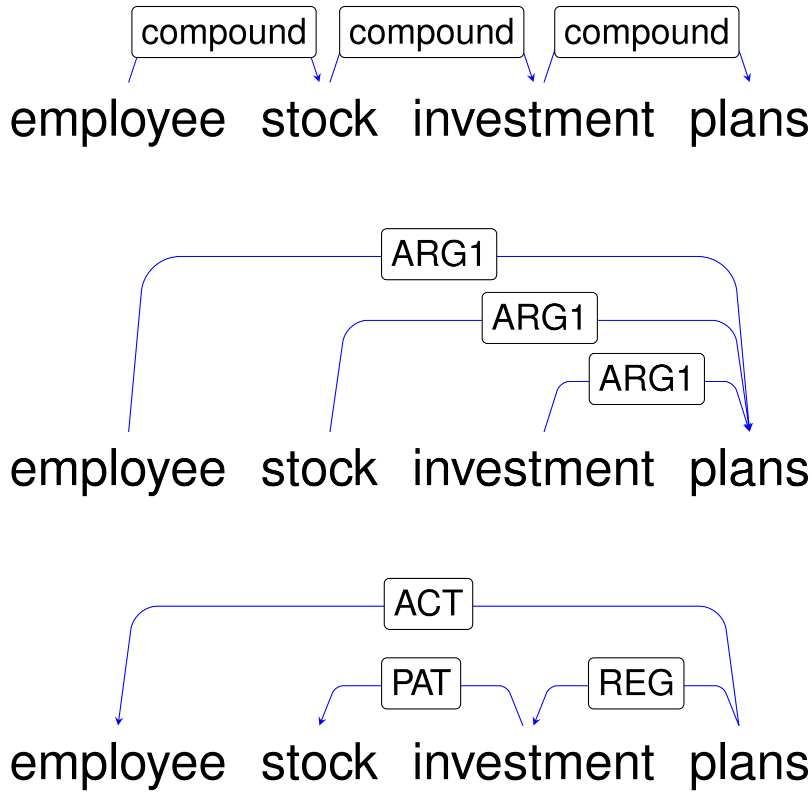
# Different Approaches to Coordinate Structures



# Diverging Ambitions: Sentence vs. Speaker Meaning



# Diverging Ambitions: Sentence vs. Speaker Meaning




? Meaning determined by linguistic signal alone vs. by utterance context;

- internal bracketing arguably grammaticalized, but not role interpretation.

# Task Setup

### Evaluation Metrics

- Labeled and unlabeled precision, recall, and $F_1$ of individual dependencies;
- additionally, labeled and unlabeled exact sentence accuracy (much stricter);
- identification of *top* node(s) considered additional, 'virtual' dependencies.

### Closed vs. Open Tracks

- Beyond lemma and part of speech, no representation of syntax in Task data;
- investigate role of syntax in separate, 'open' track: (almost) no holds barred;
- 'companion' analyses: simplified PTB phrase structure and Stanford Basic.

# Participating Teams and Approaches

| Team | Track | Approach | Resources |
|---|---|---|---|
| Alpage | C & O | transition-based parsing for DAGs, logistic regression, structured perceptron | companion, Brown clusters |
| CMU | O | edge classification by logistic regression, edge-factored structured SVM | companion |
| Copenhagen-Malmö | C | graph-to-tree transformation, Mate | — |
| In-House | O | pre-existing parsers developed by the organizers | grammars |
| Linköping | C | extension of Eisner's algorithm for DAGs, edge-factored structured perceptron | — |
| Peking | C | transition-based parsing for DAGs, graph-to-tree transformation, parser ensemble | — |
| Potsdam | C & O | graph-to-tree transformation, Mate | companion |
| Priberam | C & O | model with second-order features, decoding with dual decomposition, MIRA | companion |
| Turku | O | cascade of SVM classifiers (dependency recognition, label classification, top recognition) | companion, syntactic n-grams, word2vec |

# Official Results: ‘Closed’ and ‘Open’ Tracks

| | | DM | | | | PAS | | | | PCEDT | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\overline{\text{LF}}$ | LP | LR | LF | LM | LP | LR | LF | LM | LP | LR | LF | LM |
| Peking | 85.91 | 90.27 | 88.54 | 89.40 | 26.71 | 93.44 | 90.69 | 92.04 | 38.13 | 78.75 | 73.96 | 76.28 | 11.05 |
| Priberam | 85.24 | 88.82 | 87.35 | 88.08 | 22.40 | 91.95 | 89.92 | 90.93 | 32.64 | 78.80 | 74.70 | 76.70 | 09.42 |
| Copenhagen-Malmö | 80.77 | 84.78 | 84.04 | 84.41 | 20.33 | 87.69 | 88.37 | 88.03 | 10.16 | 71.15 | 68.65 | 69.88 | 08.01 |
| Potsdam | 77.34 | 79.36 | 79.34 | 79.35 | 07.57 | 88.15 | 81.60 | 84.75 | 06.53 | 69.68 | 66.25 | 67.92 | 05.19 |
| Alpage | 76.76 | 79.42 | 77.24 | 78.32 | 09.72 | 85.65 | 82.71 | 84.16 | 17.95 | 70.53 | 65.28 | 67.81 | 06.82 |
| Linköping | 72.20 | 78.54 | 78.05 | 78.29 | 06.08 | 76.16 | 75.55 | 75.85 | 01.19 | 60.66 | 64.35 | 62.45 | 04.01 |

# Official Results: ‘Closed’ and ‘Open’ Tracks

| | | DM | | | | PAS | | | | PCEDT | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\overline{\text{LF}}$ | LP | LR | LF | LM | LP | LR | LF | LM | LP | LR | LF | LM |
| Peking | 85.91 | 90.27 | 88.54 | 89.40 | 26.71 | 93.44 | 90.69 | 92.04 | 38.13 | 78.75 | 73.96 | 76.28 | 11.05 |
| Priberam | 85.24 | 88.82 | 87.35 | 88.08 | 22.40 | 91.95 | 89.92 | 90.93 | 32.64 | 78.80 | 74.70 | 76.70 | 09.42 |
| Copenhagen-Malmö | 80.77 | 84.78 | 84.04 | 84.41 | 20.33 | 87.69 | 88.37 | 88.03 | 10.16 | 71.15 | 68.65 | 69.88 | 08.01 |
| Potsdam | 77.34 | 79.36 | 79.34 | 79.35 | 07.57 | 88.15 | 81.60 | 84.75 | 06.53 | 69.68 | 66.25 | 67.92 | 05.19 |
| Alpage | 76.76 | 79.42 | 77.24 | 78.32 | 09.72 | 85.65 | 82.71 | 84.16 | 17.95 | 70.53 | 65.28 | 67.81 | 06.82 |
| Linköping | 72.20 | 78.54 | 78.05 | 78.29 | 06.08 | 76.16 | 75.55 | 75.85 | 01.19 | 60.66 | 64.35 | 62.45 | 04.01 |

| | | DM | | | | PAS | | | | PCEDT | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\overline{\text{LF}}$ | LP | LR | LF | LM | LP | LR | LF | LM | LP | LR | LF | LM |
| Priberam | 86.27 | 90.23 | 88.11 | 89.16 | 26.85 | 92.56 | 90.97 | 91.76 | 37.83 | 80.14 | 75.79 | 77.90 | 10.68 |
| CMU | 82.42 | 84.46 | 83.48 | 83.97 | 08.75 | 90.78 | 88.51 | 89.63 | 26.04 | 76.81 | 70.72 | 73.64 | 07.12 |
| Turku | 80.49 | 80.94 | 82.14 | 81.53 | 08.23 | 87.33 | 87.76 | 87.54 | 17.21 | 72.42 | 72.37 | 72.40 | 06.82 |
| Potsdam | 78.60 | 81.32 | 80.91 | 81.11 | 09.05 | 89.41 | 82.61 | 85.88 | 07.49 | 70.35 | 67.33 | 68.80 | 05.42 |
| Alpage | 78.54 | 83.46 | 79.55 | 81.46 | 10.76 | 87.23 | 82.82 | 84.97 | 15.43 | 70.98 | 67.51 | 69.20 | 06.60 |
| In-House | 75.89 | 92.58 | 92.34 | 92.46 | 48.07 | 92.09 | 92.02 | 92.06 | 43.84 | 40.89 | 45.67 | 43.15 | 00.30 |

# Outlook: Candidate Revisions for SDP 2015

### Out-of-Domain Testing

?

### Complete-Predicate Scoring

?

### Harmonization of Target Representations

?

### Cross-Linguistic Variation

?

### Predicate Disambiguation

?