Phrase-structures and Dependencies for End-to-End Coreference Resolution

Anders Björkelund **Jonas Kuhn**
Institut für Maschinelle Sprachverarbeitung, University of Stuttgart
{anders,jonas}@ims.uni-stuttgart.de

End-to-End Coreference

- Example

  [Information International Inc.] said [it] was sued by a buyer of [[its] computerized newspaper-publishing system], alleging that [the company] failed to correct deficiencies in [the system]. A spokesman for [Information International] said the lawsuit by [two units of Morris Communications Corp.] seeks restitution of the [system 's] about $3 million price and cancellation of a software license provided by [the Morris units] to [Information International] for alleged failure to pay royalties. [Information International] said [it] believes that the complaints, filed in federal court in Georgia, are without merit.

- End-to-End coreference:
  Identify mentions **and** cluster them into coreference chains

Grammatical Factors in Coreference Resolution

- Prototypical examples based on Binding Theory

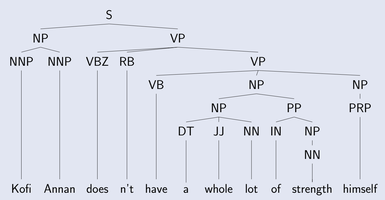
1. John$_i$ thinks that Bill$_j$ hurt himself$_{*i/j}$. Principle A – Reflexives must be coreferent with an element inside their local clause
2. John$_i$ thinks that Bill$_j$ hurt him$_{i/*j}$. Principle B – Non-reflexives must have an antecedent outside their local clause
3. He$_i$ hurt John$_{*i/j}$. Principle C – Full NPs (e.g. proper names) must not be preceded by a coreferent NP in the same sentence

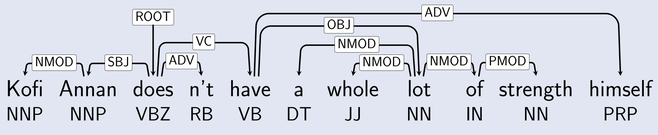
- However, there are exceptions:

4. Ronni$_i$ suspected that was probably true [...] [S]omething else [...] had provoked her$_i$ own furious outburst [...] Some more personal resentment that had come from within herself$_i$.
5. John did not have any money on him (/*himself).

- Additionally, statistical parsers do make mistakes –
  Trying to enforce binding theory constraints might lead to unexpected results

Syntactic Paradigms





- The syntax path in the phrase-structure tree indicates that *Kofi Annan* and *himself* are in the same clause
- The dependency tree states clearly that *Kofi Annan* is a subject, while clause boundaries are less overt

Brief System Description

- State-of-the-art system (Björkelund and Farkas, 2012) from this years CoNLL Shared Task (Pradhan et al. 2012)
- Second best system in the Shared Task, available at `http://www.ims.uni-stuttgart.de/~anders`
- Pairwise system with vast feature space (including lexical features, syntax paths, syntactic categories etc)
- Automatic feature selection to find the optimal feature set given either type of syntax

Conclusions at a glance

- Phrase-structure syntax is superior to dependency syntax for English coreference resolution
- A combination of features drawing on both syntactic paradigms improve performance, esp. for pronouns
- Independently predicting the two syntactic representations is better than converting the phrase-structures

Experimental setup

- Predicted syntax
  - CoNLL phrase-structure (**PS**) trees (Charniak parser)
  - Stanford dependencies (**DT**$_{Stanf}$) by converting the predicted PS trees
  - Predicted dependencies (**DT**$_{Choi}$) using the conversion of Choi and Palmer (2010), predicted with the parser of Bohnet (2010)
- Five systems that only differ in their feature sets wrt syntax:
  1. Baseline: the Björkelund & Farkas system stripped of syntax-based features
  2. Reference: the Björkelund & Farkas system
  3. Choi deps: the Baseline system extended with features from Choi dependencies
  4. Choi deps + PS: the Reference system extended with features from Choi dependencies
  5. Stanford deps + PS: the Reference system extended with features from Stanford dependencies
- To evaluate the impact of the features we keep the mention extraction fixed (based on the PS trees)

Overall Results

System	Feature set	MD	MUC	BCUB	CEAFE	CoNLL
1	BL	73.64	65.64	70.45	45.43	60.51
2	BL+PS	74.96	67.12	71.18	46.84	61.71
3	BL+DT$_{Choi}$	74.54	66.74	70.98	46.50	61.42
4	BL+PS+DT$_{Choi}$	**75.23**	**67.69**	**71.48**	47.02	**62.07**
5	BL+PS+DT$_{Stanf}$	**75.23**	67.46	71.22	**47.18**	61.96

- Evaluation metrics
  - Mention Detection F_1 (**MD**)
  - **MUC** F_1 (link-based)
  - **BCUB** F_1 (mention-based)
  - **CEAFE** F_1 (entity-based)
  - **CoNLL** average, i.e. the average of MUC, BCUB, and CEAFE

- All systems are significantly better than the Baseline in all metrics (using paired t-test)
- Systems 2, 4, and 5 are significantly better than system 3 in all metrics
- Systems 4 and 5 are better than system 2, however only MUC improvement for system 4 is significant

Results on pronouns

System	Feature set	Standard	Possessive	Reflexive	All
1	BL	68.47	68.65	69.07	68.51
2	BL+PS	69.35	71.00	68.04	69.64
3	BL+DT$_{Choi}$	68.95	69.86	65.98	69.09
4	BL+PS+DT$_{Choi}$	**70.00**	71.63	**74.23**	**70.35**
5	BL+PS+DT$_{Stanf}$	69.51	**71.69**	69.07	69.91
Total occurences		7,497	1,745	97	9,339

- Pronoun accuracy: Every pronoun (in the gold) is regarded as correct if
  1. The nearest predicted antecedent to the left is in the same cluster (in gold)
  2. The pronoun is not part of a cluster in either prediction or gold

- The improvements over the baseline are about 1-1.5% absolute and all are significant
- Note that the improvement for system 4 over system 2 is about 0.7% absolute and is also significant
- It seems like the small improvements in the coreference metrics stem from better handling of pronouns

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) via Project D8 of the SFB 732