A Multigraph Model for Coreference Resolution

Sebastian Martschat, Jie Cai, Samuel Broscheit, Éva Mújdricza-Maydt, Michael Strube
Heidelberg Institute for Theoretical Studies gGmbH

Heidelberg Institute for Theoretical Studies
HITS

Mention Extraction

We extract noun phrases from the provided parse and NE layers. For embedded mentions with the same head, we keep only the one with the largest span. We convert parse trees into dependency parse trees via Stanford's dependency parser. For every mention, we compute its mention type, number, gender, semantic class, grammatical function etc.

Mentions → nodes

Relations

Negative relations: indicate that the mentions are not coreferent.

Positive relations: clues for coreference between mentions.

Some example relations:

- N_Number: two mentions do not agree in number
- P_HeadMatch: heads of mentions match
- P_Nprn_Prn: antecedent is not a pronoun, anaphor is a pronoun
- P_Subject: anaphor is *it*/*they*, both mentions are subjects, sentence distance ≤ 1

Weights for relations. Each relation R has a weight w_R computed as follows. For each R we count

- n_R: number of all pairs of mentions in training data such that R holds
- c_R: number of **coreferent** pairs of such mentions

Then set $w_R = c_R / n_R$.

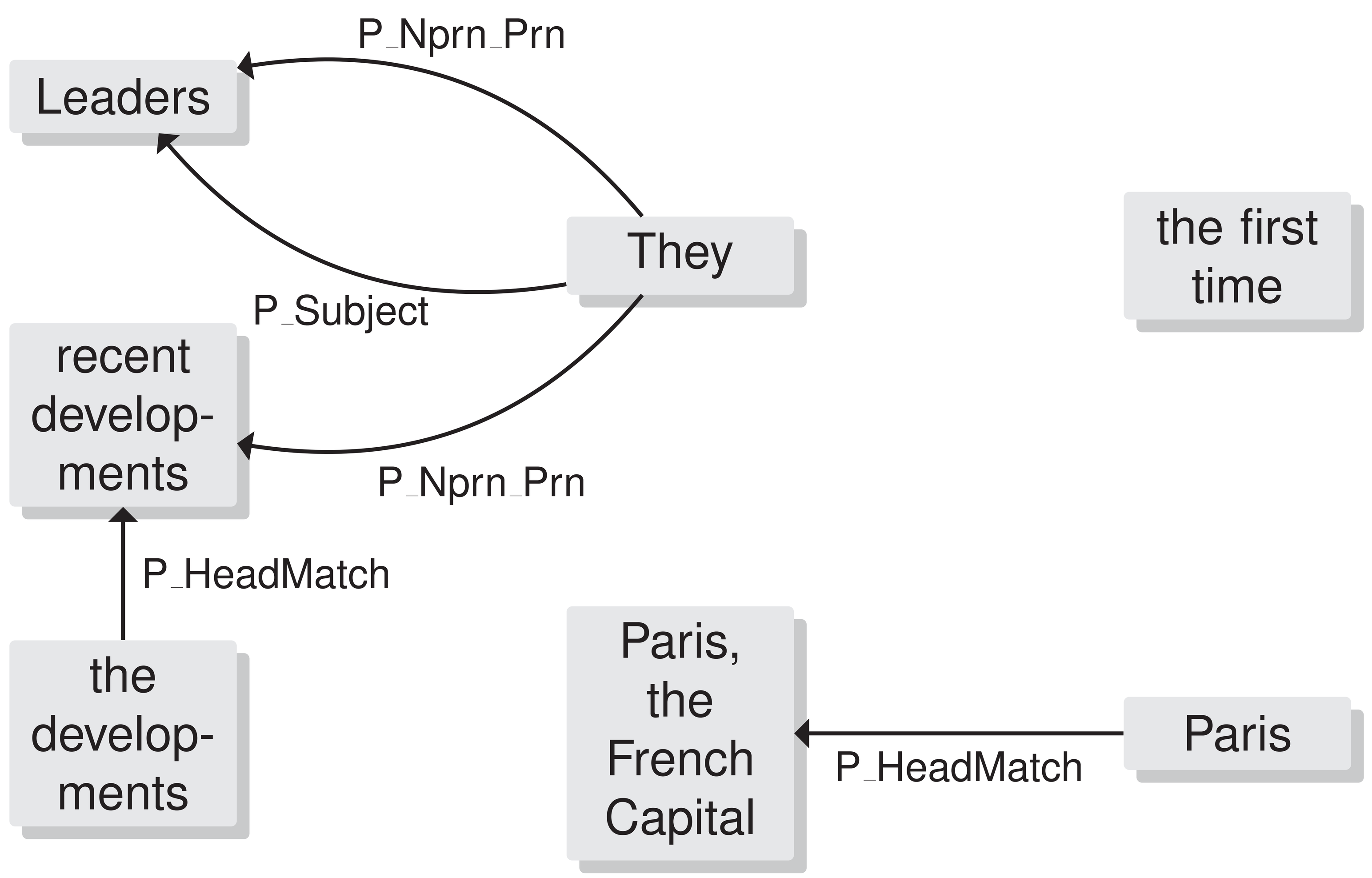
Relations → templates for edges

Multigraph Model

Graph construction. For every document we construct a weighted directed multigraph according to the following steps.

- each mention is a node
- if a negative relation holds for two mentions, we disallow edges between the mentions
- if a positive relation holds, we add an edge from the anaphor to the candidate antecedent
- weight for an edge: weight for relation adjusted according to distance

Example (made-up). *Leaders met in Paris, the French capital, to discuss recent developments. They met for the first time in Paris. The developments are alarming.*

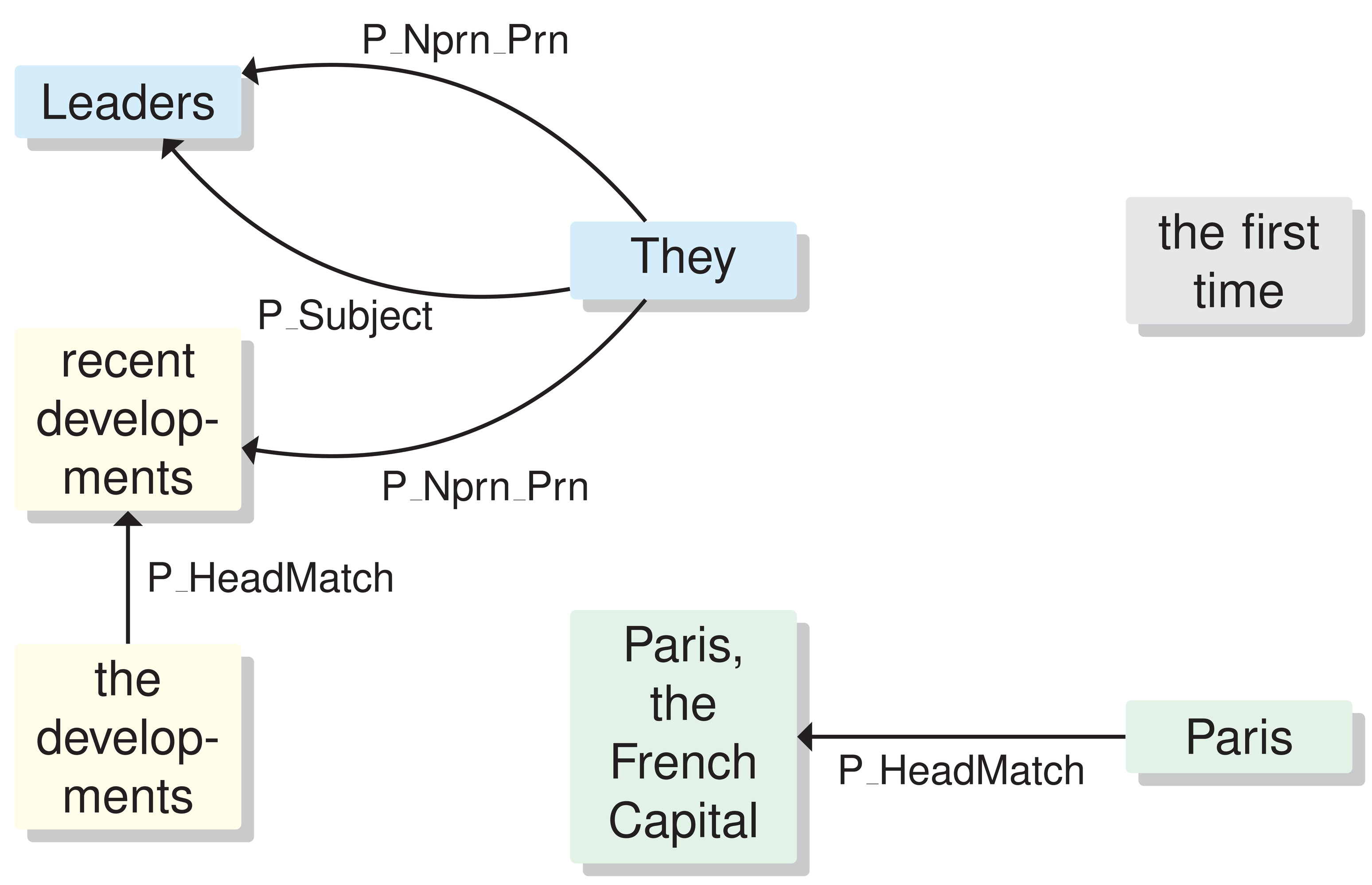


Clustering

Greedy Clustering. We go through the mentions from left to right and perform the following clustering steps:

- **pronouns** are assigned to same entity as child where the sum of edge weights is maximized
- for English, **definites and demonstratives** are assigned to same entity as child that is closest in document
- **all other noun phrases** are assigned to same entity as children
- negative relations can be used as constraints

Spectral Clustering. For Chinese, we apply spectral clustering before greedy clustering to reduce the number of possible antecedents for the mentions in the graph. We then only consider a mention's children in the same cluster as candidate antecedents.



Results and Discussion

	English		Chinese	
dataset	md	score	md	score
dev	74.81	**62.06**	60.51	**54.92**
test	75.15	**61.31**	58.60	**53.15**

- second place in English closed task
- improvement of over 5% overall score on CoNLL '11 dev set compared to last year's system: 55.63 F1 to 60.99 F1
- used training data only for computing weights for relations, only 20% of data used

Acknowledgments

This work has been funded by the Klaus Tschira Foundation, Heidelberg, Germany. The first and the second authors have been supported by a HITS PhD. scholarship.

Second Place in English Closed Task

{sebastian.martschat|jie.cai|michael.strube}@h-its.org