Anaphora Resolution in Biomedical Literature: A Hybrid Approach

Jennifer D'Souza and Vincent Ng

Human Language Technology Research Institute

The University of Texas at Dallas

What is Anaphora Resolution?

FK506 suppressed the transcriptions through the AP-1 or kappa B-like sites induced by PMA plus Ca(2+)-mobilizing agents, but not those induced by Ca(2+)-independent stimuli.

- Task: identify an antecedent for each anaphor
- 3 subtasks
  1. Identify all the anaphors
  2. Identify all the candidate antecedents for each anaphor
  3. Determine which of these candidate antecedents is the correct antecedent for each anaphor

Our Evaluation Data-set

- from BioNLP 2011 Coreference Task

Why Coreference?

Useful for Event Extraction

BioNLP Event Extraction

Event Cause

A mutant of KBF1, p50 (delta SP), unable to bind to DNA but able to form homo-

Negative Regulation Event

or heterodimers, has been constructed. This protein reduces or abolishes in vitro

the DNA binding activity of wild-type proteins of the same family…

Previous Approaches to Coreference

- Rule-Based or Learning-Based

Our Approach: Hybrid Approach

- Use different approaches to resolve different classes of anaphors.

Different classes of anaphors?

Anaphor Type	Examples	Training	Development
Relative Pronoun	that, which, who, where, etc.	54.3%	56.9%
Personal Pronoun	it, they	26.6%	26.0%
Definite Noun Phrase	the genes, this protein, etc.	15.4%	14.0%
Demonstrative & Indefinite Pronoun	this, those, both, etc.	2.4%	2.1%
Others		1.3%	1.1%

- Why no statistics on the test set?
- How then do we evaluate?

The test set is not available to system developers.

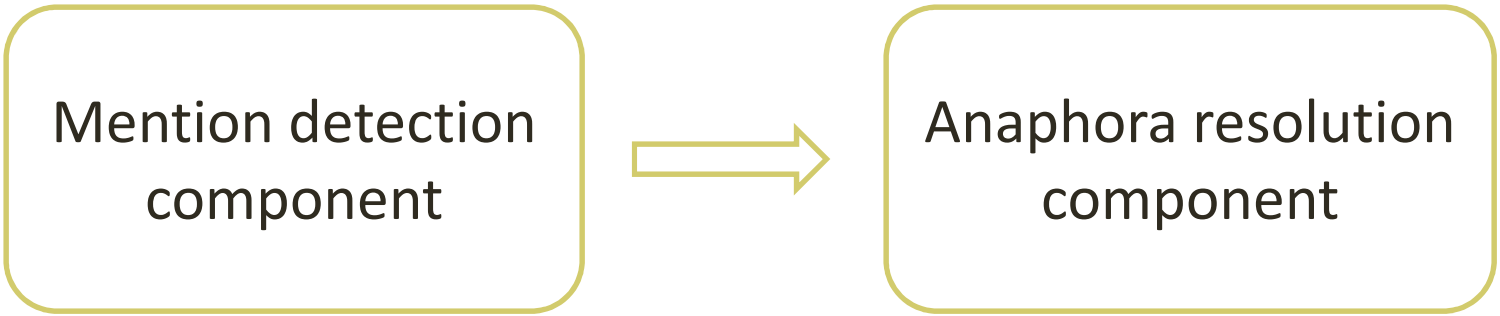
Motivation for Hybrid System

- Hypothesis: Different classes of anaphors might be better resolved using different approaches.

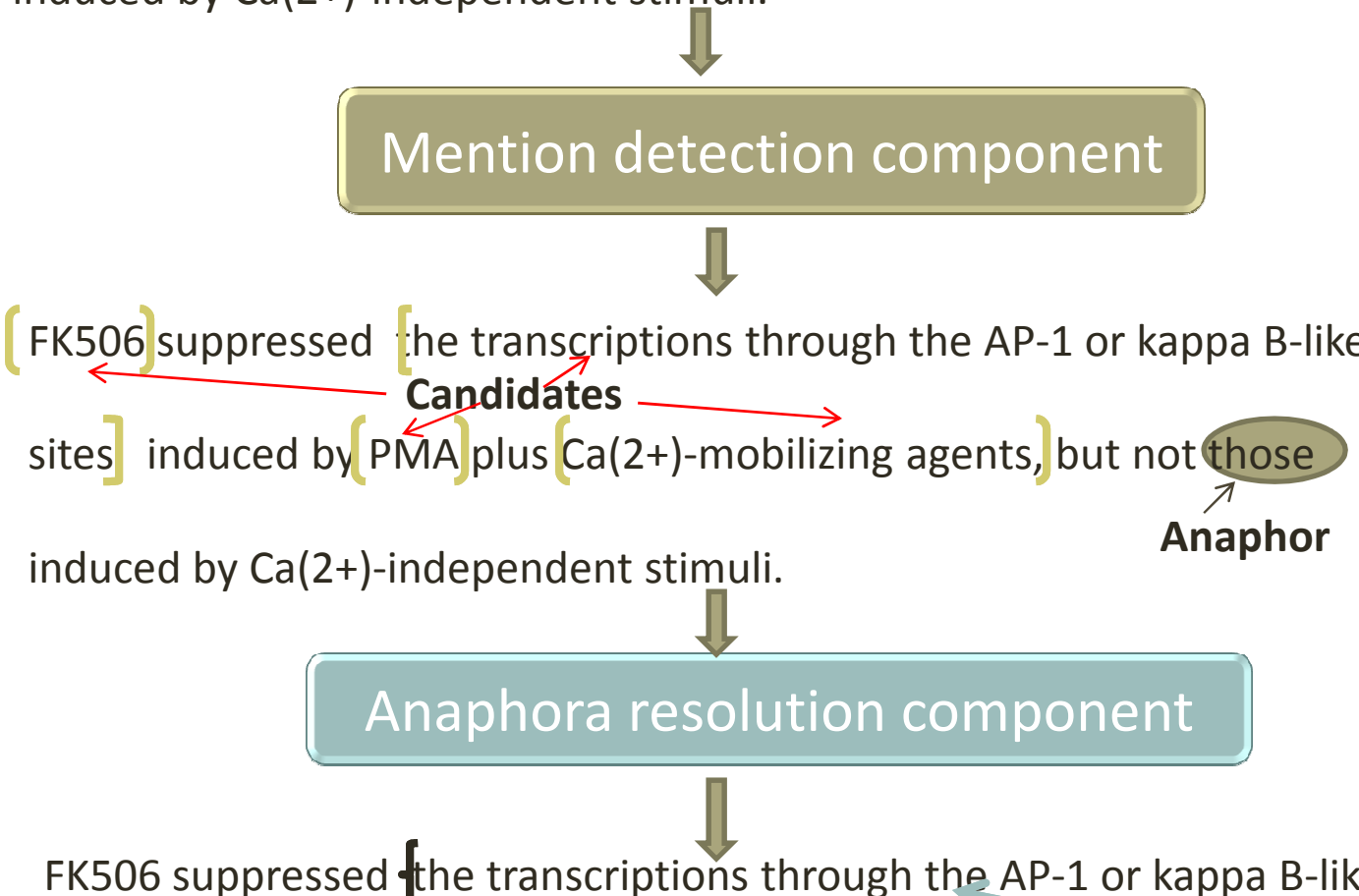
- Basis of Hypothesis?
  - Linguistic properties
    - Different features for different anaphor types?
  - Data-set distributions
    - Rule-based versus learning-based approaches?

System Architecture

- A pipeline architecture

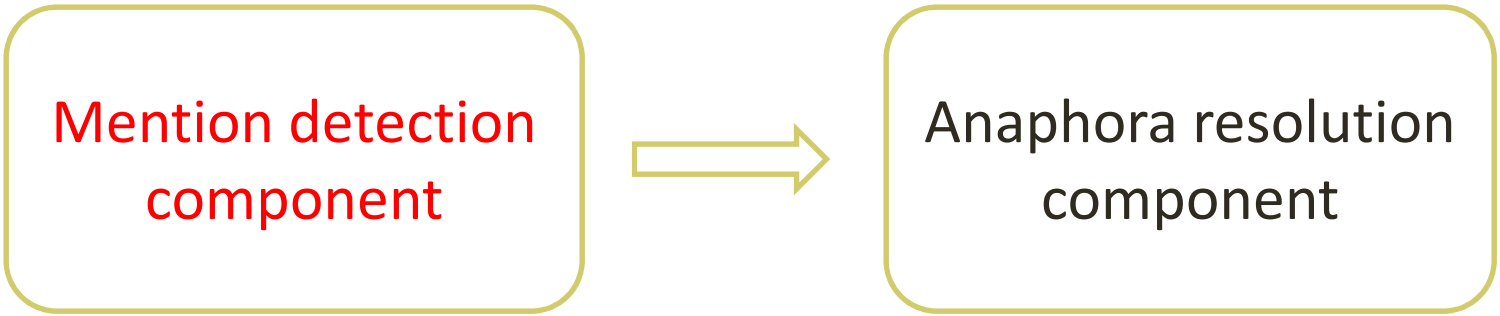


FK506 suppressed the transcriptions through the AP-1 or kappa B-like sites induced by PMA plus Ca(2+)-mobilizing agents, but not those induced by Ca(2+)-independent stimuli.



System Architecture

- A pipeline architecture



Goal: Extract Anaphors & Candidate Antecedents

2 Approaches to Mention Detection

1. Learning-Based Approach
2. Heuristic-Based Approach

Learning-Based Mention Detection

- Sequential Labeling Task – CRF
- Class Values: given a sentence token, does it begin the mention (B), or is it inside the mention (I), or is it outside a mention (O)?
- Features: Token, POS, word shape information, etc.
- Separate Anaphor & Candidate Antecedent Classifiers *[Kim et al., 2011]*
- Limitation:
  - Insufficient training instances for sparse anaphor classes

Heuristic-Based Mention Detection

- Anaphor Extractor
  - Step1: List-Based Extraction
    - Use pre-created lists to extract anaphors
  - Step 2: Prune Extracted Non-Anaphors with Heuristics
    - E.gs. of non-anaphors are complementizers as in “found that”, “suggests that”, or pleonastic pronouns as in “It is found that”, “It was possible that”, etc.
- Antec
  - List
  - synt

Anaphor type	Before Pruning TP/FP	After Pruning TP/FP
Relative pronoun	269/313	262/22
Personal pronoun	123/235	120/5
D&I pronoun	32/19	32/13
Definite NP	10/12	10/2

from the

nention)

Combinations of Mention Extraction Methods

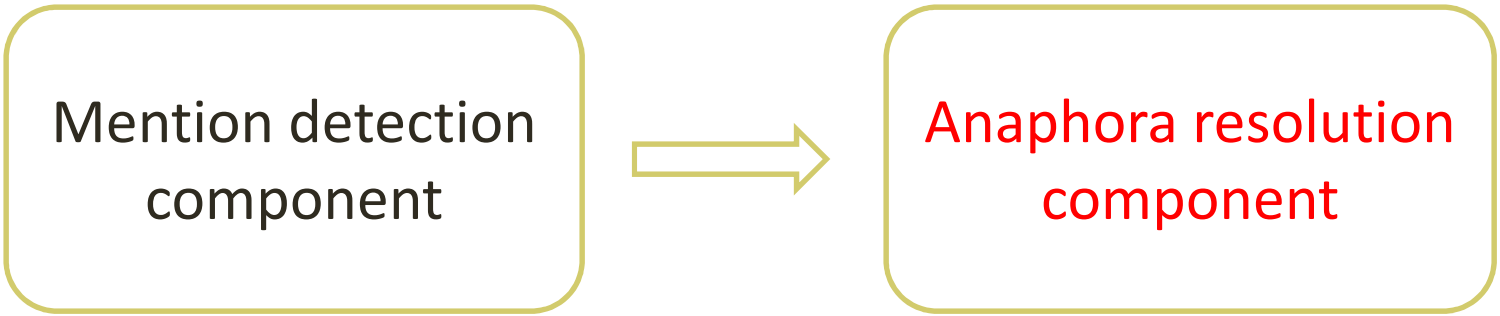
- We now have 2 methods for extracting candidate antecedents (1 learning-based, 1 heuristic-based)
- We now have 2 methods for extracting anaphors (1 learning-based, 1 heuristic-based)
- We can mix learning-based and heuristic-based methods for extracting anaphors and candidate antecedents
  - 4 possible ways:
    - CRF Anaphors + CRF Antecedents
    - CRF Anaphors + Heuristic Antecedents
    - Heuristic Anaphors + Heuristic Antecedents
    - Heuristic Anaphors + CRF Antecedents

Which combination should we use?

- Development data helps us decide…

System Architecture

- A pipeline architecture



Goal: To find the antecedent for an anaphor

6 Anaphora Resolution Methods

1. Reconcile Features
2. Sentence-Based Flat Parse Features
3. Document-Based Flat Parse Features
4. Sentence-Based Structured Parse Feature
5. Document-Based Structured Parse Feature
6. Rule-Based Method

Learning-Based Methods

- Why 6 methods?
  - Hypothesis: Different methods may work well for different anaphor types

Resolution Method 1

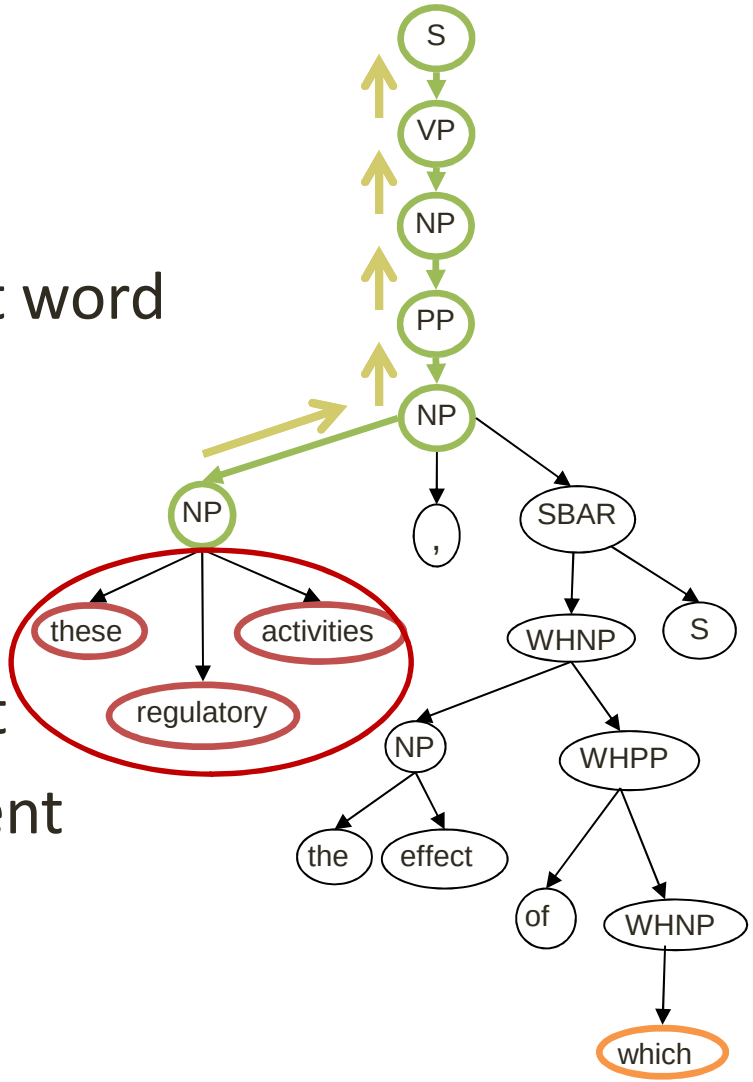
- Goal
  - using a ranker trained on Reconcile features to obtain the correct antecedent for an anaphor
    - 66 string-matching, grammatical, positional, and semantic features from Reconcile
    - ranker aims to rank the candidate so the correct one has highest rank
- How do we train this ranker?
  - generate a feature vector for anaphor paired with a candidate from the list

Resolution Method 2

- Weakness of Method 1
  - need to design potentially complex heuristics for encoding parse tree information as features
- Solution
  - train a ranker on path-based features extracted from sentence parse trees (i.e. features derived from paths in a parse tree)
- 6 path-based features

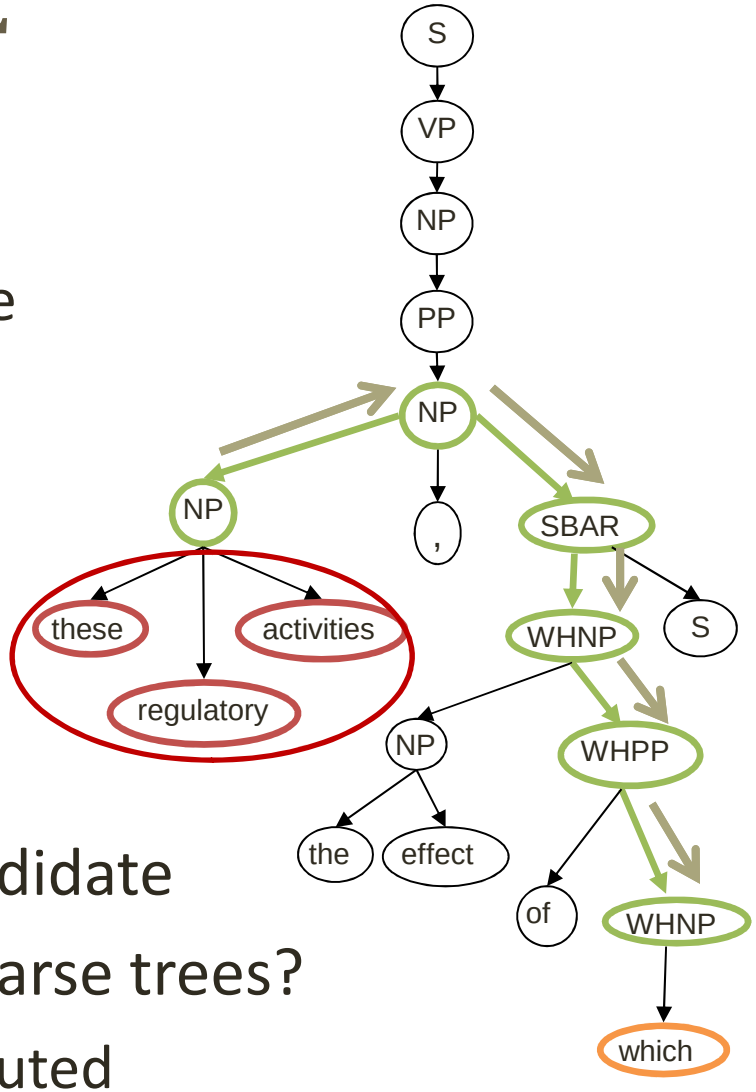
Resolution Method 2

- Feature 1
  - Path from the parent of first candidate antecedent word to the root of the tree
- Motivation
  - Captures syntactic context of the candidate antecedent



Resolution Method 2

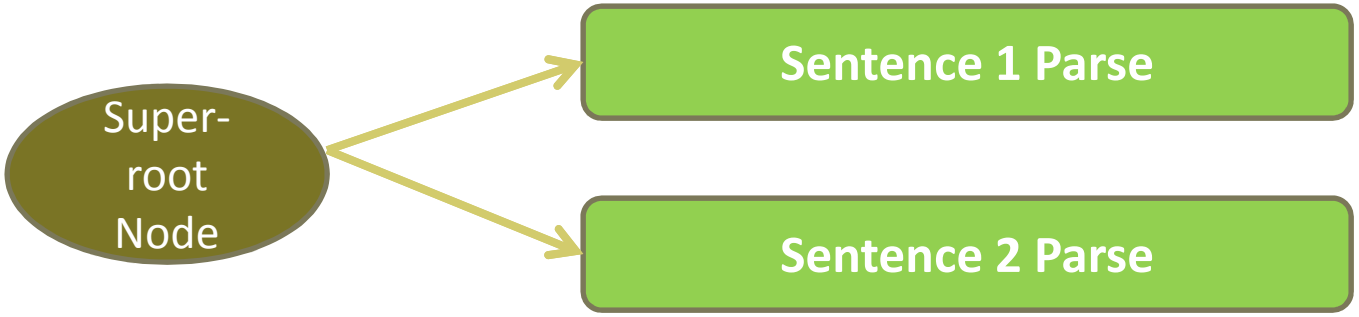
- Feature 6
  - Directed path from candidate antecedent to anaphor
- Motivation
  - Captures syntactic context
- What if the anaphor and candidate antecedent are in different parse trees?
  - This feature cannot be computed



Resolution Method 3

- Addresses this problem by using document based rather than sentence based parse trees

- What are document based parse trees?
  - sentence parses are connected by a pseudo link



- Ranker trained on the same 6 features as in method 2 except that they are computed on document parse trees

Resolution Method 4

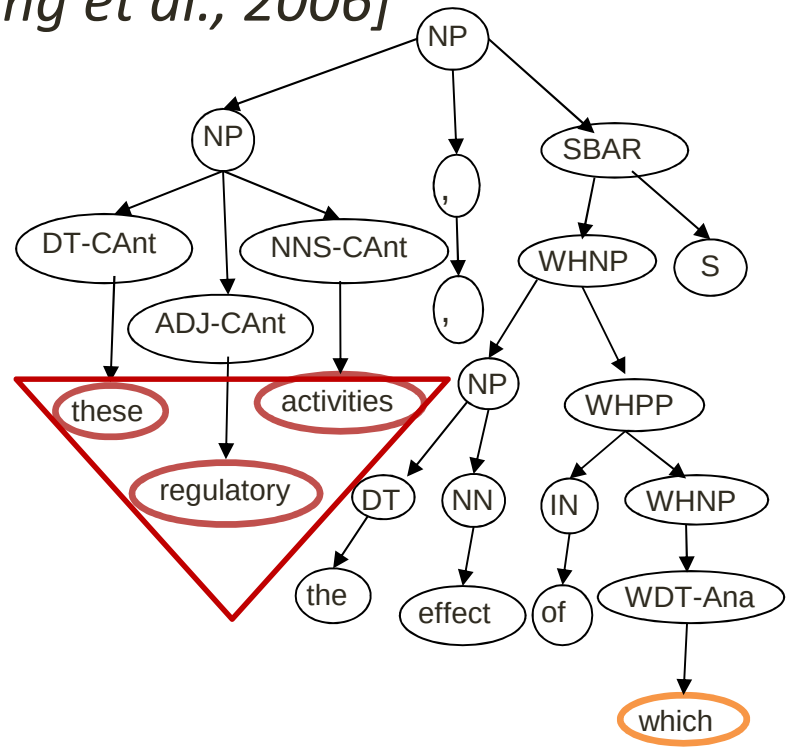
- Weakness of methods 2 & 3
  - Need to manually determine which paths in a parse tree to use as features
- Solution
  - Use a sentence-based parse tree as a structured feature
- What is a structured feature?
  - A feature whose value is a linear or hierarchical structure, as opposed to a flat feature, which has a discrete or real value

Resolution Method 4

- But we cannot use the entire parse tree…
  - the learner cannot generalize well
  - so we extract a parse substructure (i.e. subtree) and use as a structured feature
  - But which parse substructure do we extract?

Structured Tree Feature

- Simple Expansion Tree *[Yang et al., 2006]*
  - includes all nodes in path from candidate antecedent to anaphor and the nodes first level children

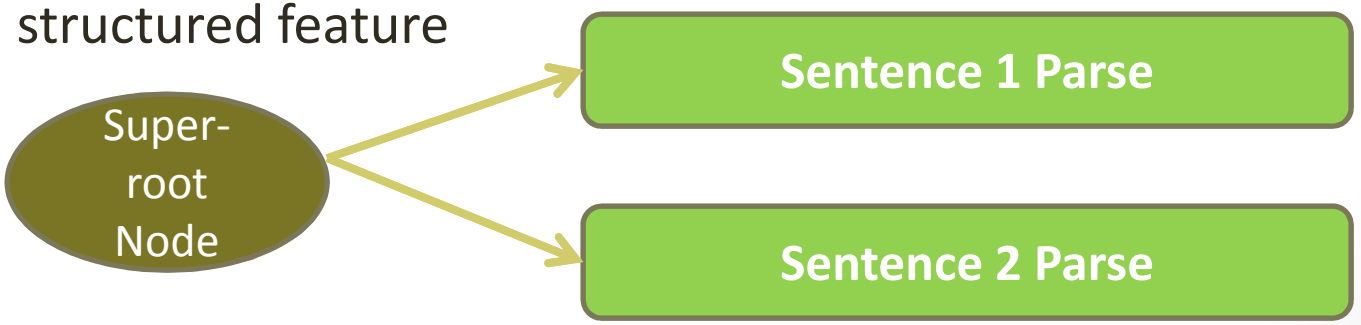


Resolution Method 4

- Use this sentence-based structured feature to train a classifier

Resolution Method 5

- Weakness of method 4
  - The sentence-based structured feature cannot be computed if the candidate antecedent and the anaphor are not in the same sentence
- Solution
  - Same as method 4 except that we connect sentence-based parse trees by a pseudo link to create a document-based structured feature



Resolution Method 6

- Rule-based method
- Each rule specifies which candidate antecedent an anaphor should be resolved to.
- Each type of anaphors has its own set of resolution rules.
  - Each set of resolution rules is ordered
    - So that the second rule is applied only if the first rule is not applicable

Rules for Resolving Personal Pronouns

Rule 1: Resolve anaphor to candidate if (1) the two agree in number and are in the same sentence; and (2) candidate contains a protein name or one of its words satisfies the three conditions in the Pattern rule.
Rule 2: Resolve anaphor to candidate if the two agree in number and are in the same sentence.
Rule 3: Resolve anaphor to candidate if candidate contains a protein name or one of its words satisfies the three conditions in the Pattern rule.
Rule 4: Resolve anaphor to candidate if the two are in the same sentence.
Rule 5: Resolve anaphor to candidate if the two agree in number.

Rule for Resolving Relative Pronouns

Resolve anaphor to the closest candidate.

- For each type of anaphors, we have 24 method combinations, because we have:
  - 2 candidate antecedent extraction methods
  - 2 anaphor extraction methods
  - 6 resolution methods

- Which combination should we use?
  - We use the development set to determine the best combination of anaphor extraction method, antecedent extraction method, and resolution method for each of the 4 types of anaphors.

Relative Pronoun Resolution Results on Development Set

Resolution Method	CRF anaphors						Heuristic anaphors					
	CRF candidates			Heuristic candidates			CRF candidates			Heuristic candidates		
	R	P	F	R	P	F	R	P	F	R	P	F
Ranking-based Reconcile	21.3	60.6	31.5	13.4	47.4	20.8	21.3	62.3	31.7	14.9	53.6	23.3
Sentence-based flat	19.8	83.3	32.0	28.2	83.8	**42.2**	18.8	84.4	30.8	25.2	91.1	39.5
Document-based flat	19.3	83.0	31.3	28.2	78.0	41.4	19.3	84.8	31.5	24.3	90.7	38.3
Sentence-based structured	21.3	75.4	33.2	22.8	79.3	35.4	20.8	77.8	32.8	22.3	78.9	34.7
Document-based structured	21.3	69.4	32.6	22.3	77.6	34.6	20.8	72.4	32.3	22.3	81.8	35.0
Rule-based	—	—	—	27.2	75.3	40.0	—	—	—	27.7	77.8	40.8

- Best combination for relative pronouns:
  - CRF anaphors, heuristic candidates and learning method using sentence-based flat features.

Personal Pronoun Resolution Results on Development Set

Resolution Method	CRF anaphors						Heuristic anaphors					
	CRF candidates			Heuristic candidates			CRF candidates			Heuristic candidates		
	R	P	F	R	P	F	R	P	F	R	P	F
Ranking-based Reconcile	3.5	24.1	6.1	19.3	63.9	29.7	5.0	40.0	8.8	19.8	59.7	29.7
Sentence-based flat	3.5	53.8	6.5	21.8	74.6	33.7	3.5	63.6	6.6	21.3	76.8	33.3
Document-based flat	3.0	54.5	5.6	19.8	80.0	31.7	3.5	63.6	6.6	19.8	81.6	31.9
Sentence-based structured	3.5	53.8	6.5	24.3	73.1	36.4	5.0	66.7	9.2	26.3	77.9	**39.3**
Document-based structured	3.5	26.9	6.1	21.8	75.9	33.8	5.0	34.5	8.7	23.8	76.2	36.2
Rule-based	—	—	—	13.9	75.7	23.4	—	—	—	16.3	71.7	26.6

- Best combination for personal pronouns:
  - Heuristic anaphors, heuristic candidates and learning method using sentence-based structured feature.

Demonstrative & Indefinite Pronoun Resolution Results on Development Set

Resolution Method	CRF anaphors						Heuristic anaphors					
	CRF candidates			Heuristic candidates			CRF candidates			Heuristic candidates		
	R	P	F	R	P	F	R	P	F	R	P	F
Ranking-based Reconcile	0.0	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN
Sentence-based flat	0.0	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	2.0	12.9	**3.4**
Document-based flat	0.0	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	0.0	0.0	NaN
Sentence-based structured	0.0	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	0.0	0.0	NaN
Document-based structured	0.0	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN
Rule-based	—	—	—	0.0	NaN	NaN	—	—	—	1.0	100	2.0

- Best combination for demonstrative and indefinite pronouns:
  - Heuristic anaphors, heuristic candidates and learning method using sentence-based flat features.

Definite Noun Phrase Resolution Results on Development Set

Resolution Method	CRF anaphors						Heuristic anaphors					
	CRF candidates			Heuristic candidates			CRF candidates			Heuristic candidates		
	R	P	F	R	P	F	R	P	F	R	P	F
Ranking-based Reconcile	0.0	NaN	NaN	0.5	100	1.0	0.5	11.1	0.9	1.0	50.0	1.9
Sentence-based flat	0.0	NaN	NaN	0.5	7.1	0.9	0.0	NaN	NaN	2.5	14.7	4.2
Document-based flat	0.0	NaN	NaN	1.0	12.5	1.8	0.0	NaN	NaN	0.0	0.0	NaN
Sentence-based structured	0.0	NaN	NaN	0.0	0.0	NaN	0.0	NaN	NaN	0.0	NaN	NaN
Document-based structured	0.0	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN
Rule-based	—	—	—	5.0	38.5	8.8	—	—	—	6.9	58.3	**12.4**

- Best combination for definite noun phrases:
  - Heuristic anaphors, heuristic candidates and rule-based method.

Observation

- Different combination methods work best for different types of anaphors on development set
  - Provides empirical support for a hybrid approach to anaphora resolution

- We employ the best combination learned for each anaphor type from the development set to resolve the anaphors in the test documents.

Results Using the Best Combination on Development and Test Sets

System	Development Set			Test Set		
	R	P	F	R	P	F
Reconcile	26.7	74.0	39.3	22.2	73.3	34.1
EventMine	53.5	69.8	60.5	50.4	62.7	55.9
Our system	59.9	77.1	**67.4**	55.6	67.2	**60.9**

Error Analysis

- Definite Noun Phrases:
  - Our mention detection method is constrained to only extract the seen anaphors in the training set.
- Personal Pronouns:
  - Our system only accounts for intra-sentential pronouns. This affects both precision and recall.

Conclusion

- Substantiated our hypothesis that different methods are needed for resolving different types of anaphors.
- Proposed a hybrid approach to coreference resolution.