

Coreference & Coherence

Ling571
Deep Processing Techniques for NLP
March 7, 2016

Roadmap

- Coreference algorithms:
 - Deterministic sieves
- Discourse structure
 - Cohesion
 - Topic segmentation
 - Coherence
 - Discourse parsing

NP Coreference Examples

- Link all NPs refer to same entity

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

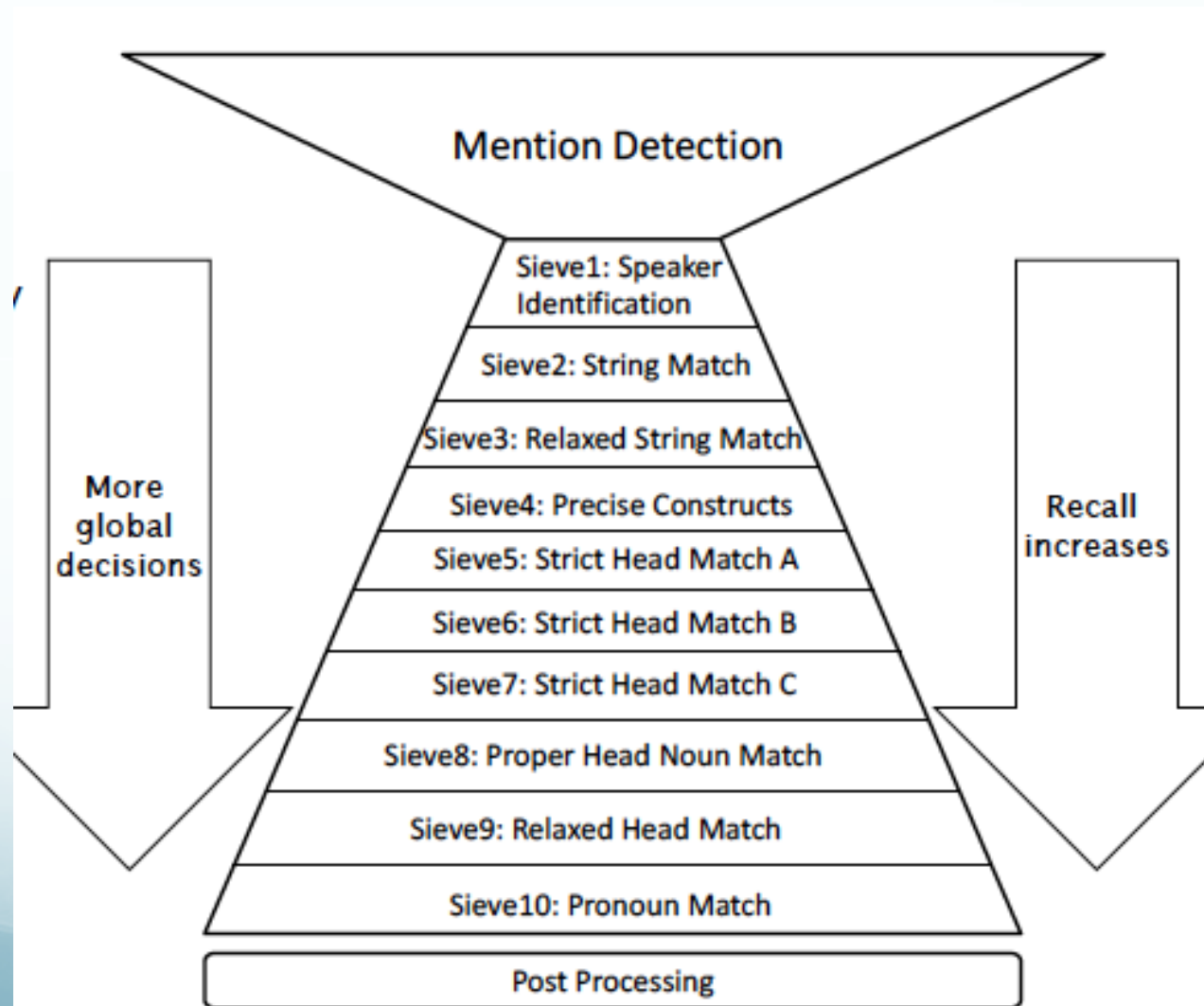
Multi-pass Sieve Approach

- Raghunathan et al., 2010
- Key Issues:
 - Limitations of mention-pair classifier approach
 - Local decisions over large number of features
 - Not really transitive
 - Can't exploit global constraints
 - Low precision features may overwhelm less frequent, high precision ones

Multi-pass Sieve Strategy

- Basic approach:
 - Apply tiers of deterministic coreference modules
 - Ordered highest to lowest precision
 - Aggregate information across mentions in cluster
 - Share attributes based on prior tiers
- Simple, extensible architecture
 - Outperforms many other (un-)supervised approaches

Multi-Pass Sieve



Pre-Processing and Mentions

- Pre-processing:
 - Gold mention boundaries given, parsed, NE tagged
- For each mention, each module can skip or pick best candidate antecedent
 - Antecedents ordered:
 - Same sentence: by Hobbs algorithm
 - Prev. sentence:
 - For Nominal: by right-to-left, breadth first: proximity/recency
 - For Pronoun: left-to-right: salience hierarchy
 - W/in cluster: aggregate attributes, order mentions
 - Prune indefinite mentions: can't have antecedents

Multi-pass Sieve Modules

- Pass 1: Exact match (N): P: 96%
- Pass 2: Precise constructs
 - Predicate nominative, (role) appositive, re;. pronoun, acronym, demonym
- Pass 3: Strict head matching
 - Matches cluster head noun AND all non-stop cluster wds AND modifiers AND non i-within-I (embedded NP)
- Pass 4 & 5: Variants of 3: drop one of above

Multi-pass Sieve Modules

- Pass 6: Relaxed head match
 - Head matches any word in cluster AND all non-stop cluster wds AND non i-within-I (embedded NP)
- Pass 7: Pronouns
 - Enforce constraints on gender, number, person, animacy, and NER labels

Multi-pass Effectiveness

Passes	MUC		
	P	R	F1
{1}	95.9	31.8	47.8
{1,2}	95.4	43.7	59.9
{1,2,3}	92.1	51.3	65.9
{1,2,3,4}	91.7	51.9	66.3
{1,2,3,4,5}	91.1	52.6	66.7
{1,2,3,4,5,6}	89.5	53.6	67.1
{1,2,3,4,5,6,7}	83.7	74.1	78.6

Sieve Effectiveness

- ACE Newswire

This work (sieve)	83.8	73.2	78.1
This work (single pass)	82.2	71.5	76.5
Haghighi and Klein (2009) +S	77.0	75.9	76.5
Poon and Domingos (2008)	71.3	70.5	70.9
Finkel and Manning (2008) +G	78.7	58.5	67.1

Questions

- Good accuracies on (clean) text. What about...
 - Conversational speech?
 - Ill-formed, disfluent
 - Dialogue?
 - Multiple speakers introduce referents
 - Multimodal communication?
 - How else can entities be evoked?
 - Are all equally salient?

More Questions

- Good accuracies on (clean) (English) text: What about..
 - Other languages?
 - Salience hierarchies the same
 - Other factors
 - Syntactic constraints?
 - E.g. reflexives in Chinese, Korean,...
 - Zero anaphora?
 - How do you resolve a pronoun if you can't find it?

Reference Resolution Algorithms

- Many other alternative strategies:
 - Linguistically informed, saliency hierarchy
 - Centering Theory
 - Machine learning approaches:
 - Supervised: Maxent
 - Unsupervised: Clustering
 - Heuristic, high precision:
 - Cogniac

Conclusions

- Co-reference establishes coherence
- Reference resolution depends on coherence
- Variety of approaches:
 - Syntactic constraints, Recency, Frequency, Role
- Similar effectiveness - different requirements
- Co-reference can enable summarization within and across documents (and languages!)

Discourse Structure

Why Model Discourse Structure? (Theoretical)

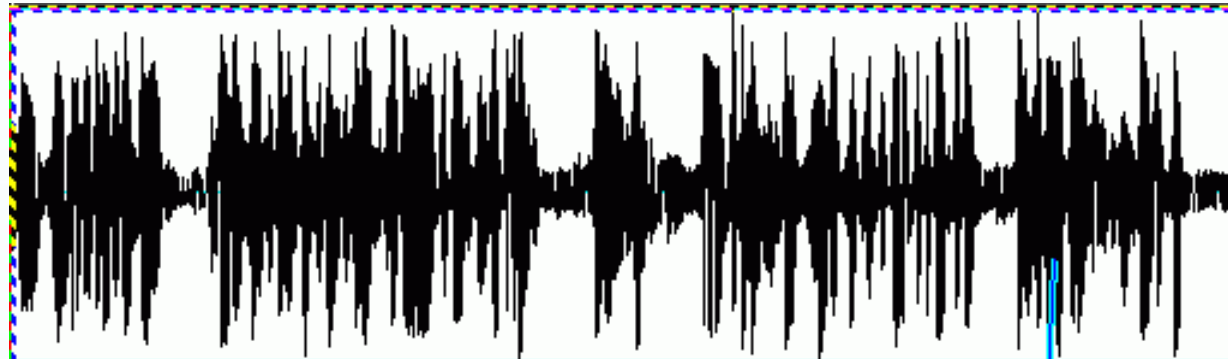
- Discourse: not just constituent utterances
 - Create joint meaning
 - Context guides interpretation of constituents
 - How????
 - What are the units?
 - How do they combine to establish meaning?
 - How can we derive structure from surface forms?
 - What makes discourse coherent vs not?
 - How do they influence reference resolution?

Why Model Discourse Structure?(Applied)

- Design better summarization, understanding
- Improve speech synthesis
 - Influenced by structure
- Develop approach for generation of discourse
- Design dialogue agents for task interaction
- Guide reference resolution

Discourse Topic Segmentation

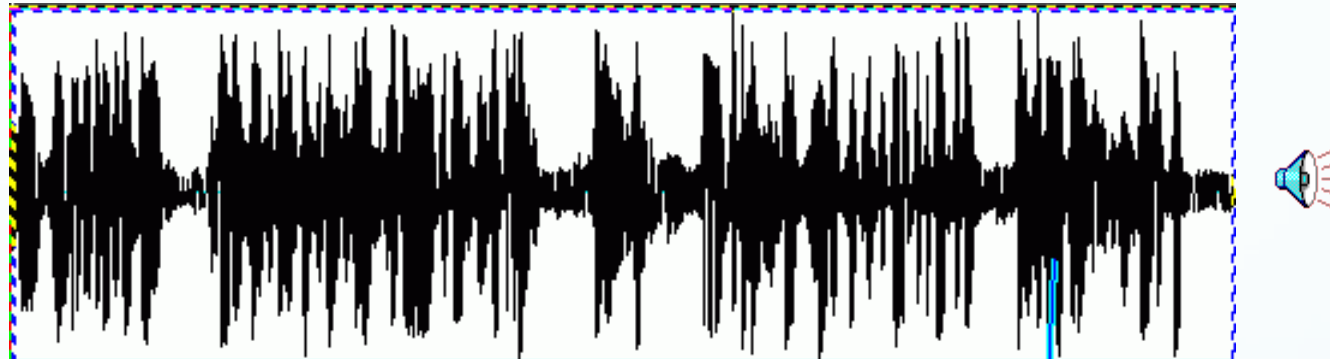
- Separate news broadcast into component stories
 - Necessary for information retrieval



On "World News Tonight" this Thursday, another bad day on stock markets, all over the world global economic anxiety. Another massacre in Kosovo, the U.S. and its allies prepare to do something about it. Very slowly. And the millennium bug, Lubbock Texas prepares for catastrophe, Bangalore in India sees only profit.

Discourse Topic Segmentation

- Separate news broadcast into component stories



On "World News Tonight" this Thursday, another bad day on stock markets, all over the world global economic anxiety. ||

Another massacre in Kosovo, the U.S. and its allies prepare to do something about it. Very slowly. ||

And the millennium bug, Lubbock Texas prepares for catastrophe, Bangalore in India sees only profit. ||

Discourse Segmentation

- Basic form of discourse structure
 - Divide document into linear sequence of subtopics
- Many genres have conventional structures:
 - Academic: Intro, Hypothesis, Methods, Results, Concl.
 - Newspapers: Headline, Byline, Lede, Elaboration
 - Patient Reports: Subjective, Objective, Assessment, Plan
- Can guide: summarization, retrieval

Cohesion

- Use of linguistics devices to link text units
 - Lexical cohesion:
 - Link with relations between words
 - Synonymy, Hypernymy
 - *Peel, core and slice the pears and the apples. Add the fruit to the skillet.*
 - Non-lexical cohesion:
 - E.g. anaphora
 - *Peel, core and slice the pears and the apples. Add them to the skillet.*
- Cohesion chain establishes link through sequence of words
- Segment boundary = dip in cohesion

TextTiling (Hearst '97)

- Lexical cohesion-based segmentation
 - Boundaries at dips in cohesion score
 - Tokenization, Lexical cohesion score, Boundary ID
- Tokenization
 - Units?
 - White-space delimited words
 - Stopped
 - Stemmed
 - 20 words = 1 pseudo sentence

Lexical Cohesion Score

- Similarity between spans of text
 - b = 'Block' of 10 pseudo-sentences before gap
 - a = 'Block' of 10 pseudo-sentences after gap
 - How do we compute similarity?
 - Vectors and cosine similarity (again!)

$$sim_{\cosine}(\vec{b}, \vec{a}) = \frac{\vec{b} \cdot \vec{a}}{|\vec{b}| |\vec{a}|} = \frac{\sum_{i=1}^N b_i \times a_i}{\sqrt{\sum_{i=1}^N b_i^2} \sqrt{\sum_{i=1}^N a_i^2}}$$

Segmentation

- Depth score:
 - Difference between position and adjacent peaks
 - E.g., $(y_{a1} - y_{a2}) + (y_{a3} - y_{a2})$

