Coreference-inspired coherence modeling

Micha Elsner and Eugene Charniak
Brown University

Ordering sentences

Ensures coherent output for:
- Multi-document summarization
- Adding new material to an existing article

Previous work using:
- Word repetition (and similarity)
- Syntactic roles of NPs

Discourse-new model

Discourse-new NPs introduce new entities to the text.
- Marked by well-studied syntactic patterns; can be detected automatically [6].
- Detection can aid coreference.
- Also used for reference rewriting in summarization [4].

Coherence modeling

Guess at coreferential chains.
- NPs with same head corefer.

Label first in chain **new**, rest **old**.

Evaluate probability of labels using a classifier based on syntactic features.

Mike McNulty, the air traffic manager
McNulty

More probable than:

McNulty
Mike McNulty, the air traffic manager

Baseline (entity grid)

Tracks repetition and syntactic roles of head nouns [3]:
- Ignores NP internal structure.
- Feature in a state-of-the-art system [5].

Pronoun coreference model

Pronouns occur close to antecedents:
- Constraints described in centering [2] and other work.

Coherence modeling

Uses joint probability of text and coreference relationships.
- Model from [1]

Probability of a pronominal reference **R** to entity **A** at a given position, given:
- Hobbs distance (**R**, **A**).
- Number of previous references to **A**.
- Number and gender agreement (**R**, **A**).

Jane Smith...
She...

More probable than:

Jane Smith...
He...

She...
Jane Smith...

Evaluation

Assumption:
- Human-authored documents use the most coherent ordering of their sentences.

Discrimination:
- Choose between original document and random permutation.

Insertion:
- Find the original location of a sentence removed from the document.

Results

	Discrim.	Insert
Chance	50.0	12.6
Entity grid	77.6	19.6
Discourse new	73.5	16.3
Pronoun	62.3	14.0
All combined	**81.0**	**23.0**

All differences significant.

Selected References

[1] Niyu Ge, John Hale and Eugene Charniak. "A statistical approach to anaphora resolution". 1998.
[2] Barbara Grosz, Aravind Joshi, Scott Weinstein. "Centering: a framework for measuring the local coherence of discourse". 1995.
[3] Mirella Lapata and Regina Barzilay. "Automatic evaluation of text coherence: models and representations". 2005.
[4] Ani Nenkova and Kathleen McKeown. "References to named entities: a corpus study". 2003.
[5] Radu Soricut and Daniel Marcu. "Discourse generation using utility-trained coherence models." 2006.
[6] Olga Uryupina. "High-precision identification of discourse-new and unique noun phrases". 2003.