Discourse Coherence

COMP-599

Nov 14, 2016

Outline

Machine learning for Coreference resolution

Cohesion and Coherence

Theories of coherence

- Rhetorical Structure Theory
- Local coherence modelling

Other Heuristic Approaches

Lappin and Leass (1994)

- Assigns weights to the various factors that we have discussed by hand.

Centering Theory (Grosz et al., 1995)

- A theory of entity transitions in a discourse, looking at their syntactic positions in sentences.
- Brennan et al. (1987) used this for pronominal anaphor resolution. Antecedent is selected in order to yield a series of entity transitions that are preferred, according to Centering Theory.

Coreference Resolution by ML

How would you set up the problem?

Soon et al., 2001

They frame coreference resolution as follows:

1. Find all **markables** in a document (various kinds of noun phrases)
2. For each target markable, m_i:
    1. Go backwards from the current markable, run a binary classifier to determine whether a previous markable m_j is coreferent with m_i
    2. If the start of the document is reached, then the current markable is a singleton

Features for Binary Classification

12 features for NP coreference resolution (not just pronominal):

Feature Type	Feature	Description
Lexical	SOON_STR	C if, after discarding determiners, the string denoting NP_i matches that of NP_j; else I.
Grammatical	PRONOUN_1*	Y if NP_i is a pronoun; else N.
	PRONOUN_2*	Y if NP_j is a pronoun; else N.
	DEFINITE_2	Y if NP_j starts with the word "the;" else N.
	DEMONSTRATIVE_2	Y if NP_j starts with a demonstrative such as "this," "that," "these," or "those;" else N.
	NUMBER*	C if the NP pair agree in number; I if they disagree; NA if number information for one or both NPs cannot be determined.
	GENDER*	C if the NP pair agree in gender; I if they disagree; NA if gender information for one or both NPs cannot be determined.
	BOTH_PROPER_NOUNS*	C if both NPs are proper names; NA if exactly one NP is a proper name; else I.
	APPOSITIVE*	C if the NPs are in an appositive relationship; else I.
Semantic	WNCLASS*	C if the NPs have the same WordNet semantic class; I if they don't; NA if the semantic class information for one or both NPs cannot be determined.
	ALIAS*	C if one NP is an alias of the other; else I.
Positional	SENTNUM*	Distance between the NPs in terms of the number of sentences.

Table from (Ng and Cardie, 2002)

Experimental Results

They trained a supervised decision tree classifier using these features.

Results on MUC-6 data set:

58.6/67.3/62.6 in terms of R/P/F1

Ng and Cardie, (2002) extended the feature set.

62.4/73.5/67.5

Durrett and Klein (2013) incorporated many features into a log-linear model (~3M).

Word-level features + simple recency, syntax, and gender/number features actually work very well.

Other Types of Reference Resolution

Event coreference resolution

Anaphoric shell nouns

Cross-document coreference resolution

Event Coreference Resolution

s1: Hewlett-Packard is negotiating to **[buy]** technology services provider Electronic Data Systems.

s8: With a market value of about $115 billion, HP could easily use its own stock to finance the **[purchase]**.

s9: If the **[deal]** is completed, it would be HP's biggest **[acquisition]** since it *[bought]* Compaq Computer Corp. for $19 billion in 2002.

(Bejan and Harabagiu, 2010)

Event Coreference Resolution

What does it mean for events to corefer?

- Same causes and effects (Davidson, 1969)
- **Happen in same time and place** (Quine, 1985)

Cues for event coreference (Bejan and Harabagiu, 2010):

- Share same event properties
- Share same participants

“*This*”-anaphora

Anaphoric shell nouns provide nominal shells for complex chunks of information (Kolhatkar et al., 2013).

Despite decades of education and widespread course offerings, the survival rate for out-of-hospital cardiac arrest remains a dismal 6 percent or less worldwide.

<u>This fact</u> prompted the American Heart Association last November to simplify the steps of CPR to make it easier for lay people to remember and to encourage even those who have not been formally trained to try it when needed.

Cross-document Alignment

Even trickier: multiple documents discussing an overlapping set of events and entities

View as an alignment problem (Wolfe et al., 2013; Roth and Frank, 2012)

(2) a. The regulator ruled on September 27 that Nasdaq too was qualified to bid for OMX [. . .][3]

b. The authority [. . .] had already approved a similar application by Nasdaq.[4]

Sure alignment

(3) a. Myanmar's military government said earlier this year it has released some 220 political prisoners [. . .][5]

b. The government has been regularly releasing members of Suu Kyi's National League for Democracy party [. . .][6]

Possible alignment

(Roth and Frank, 2012)

Coherence Modelling

Coherence

A property of a discourse that "makes sense" – there is some logical structure or meaning in the discourse that causes it to hang together.

Coherent:

Indoor climbing is a good form of exercise.

It gives you a whole-body workout.

Incoherent:

Indoor climbing is a good form of exercise.

Rabbits are cute and fluffy.

Cohesion

The use of linguistic devices to tie together text units

Ontario's Liberal government is proposing new regulations that would ban the random stopping of citizens by police.

The new rules say police officers cannot arbitrarily or randomly stop and question citizens.

Officers must also inform a citizen that a stop is voluntary and they have the right to walk away.

http://www.cbc.ca/news/canada/toronto/carding-regulations-ontario-1.3292277

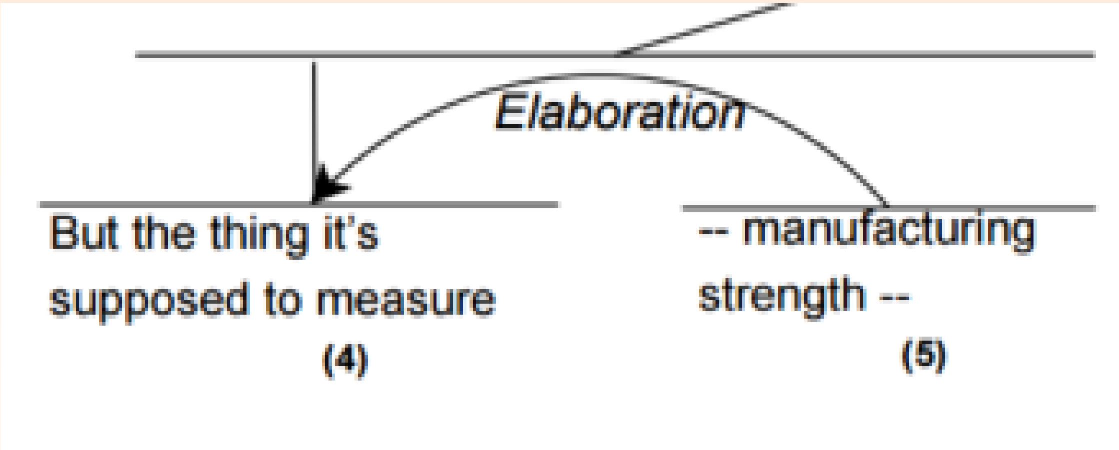
Rhetorical Structure Theory

(Mann and Thomson, 1988)

Describes the structure of a discourse by:

1. Segmenting text into **elementary discourse units** (**EDU**s)
2. Relating spans of text to each other according to a set of **rhetorical relations**
    - Elaboration
    - Attribution
    - Contrast
    - List
    - Background
    - …

Elaboration

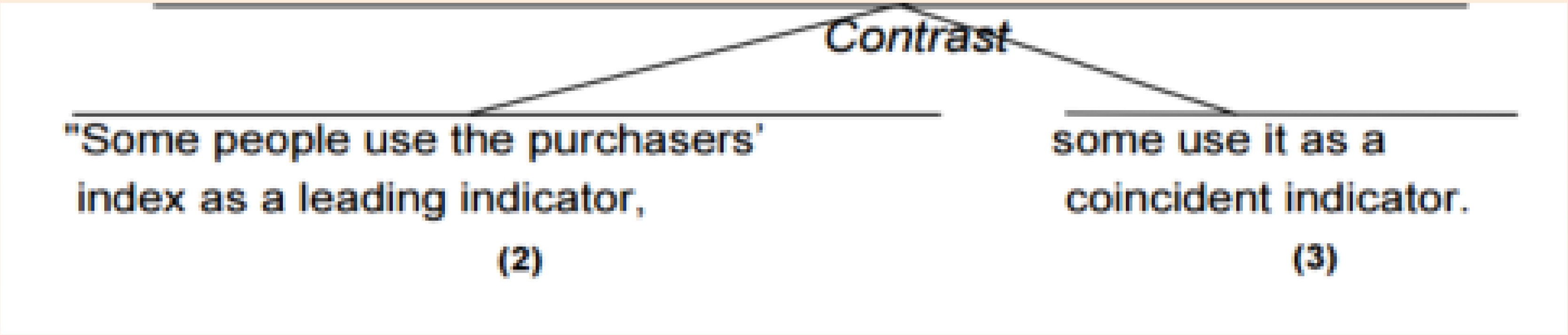


Note that the relation is **asymmetric**:

There is a **nucleus** (4)

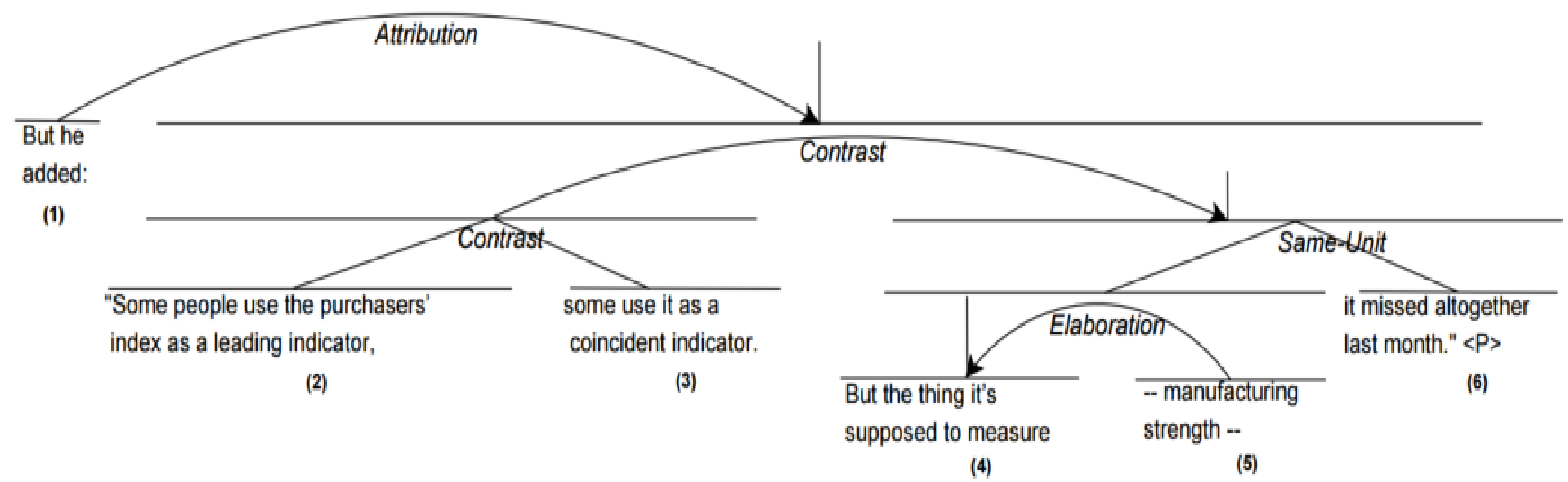
and a **satellite** (5)

Contrast



This particular instance is **symmetric**.

Example of a Full RST Tree



From Joty et al., (2013)

RST Parsing

Usually decomposed into the following steps:

Segmentation of text into EDUs

Recovering the parse structure, with labels

How would you solve this problem?

- How would you decompose the steps?
- What models and algorithms would you use to solve each step?

Applications of RST

People often extract features from RST parse trees to use in downstream applications.

e.g., automatic essay grading

RST is also helpful in automatic summarization.

Marcu (2000) defined heuristics that exploit the asymmetrical structure of RST parse trees to determine the summary content.

Local Coherence Modelling (LCM)

RST builds up a global parse tree, representing the coherence of an entire passage.

Local coherence modelling (Barzilay and Lapata, 2005) emphasizes local cohesive devices that are used to capture coherence between adjacent sentences.

The Life and Death of an Entity

Mentions of entities in a document tend to follow certain patterns (Centering Theory also relies on this):

- First mention: often the subject
  - ***Justin Pierre James Trudeau*** *MP (born December 25, 1971) is a Canadian politician and the prime minister of Canada. …*
- Mention clusters – an entity will often appear multiple times within one part of an article, then disappear
  - *… The others are* ***Ben Mulroney*** *(son of Brian Mulroney), Catherine Clark (daughter of Joe Clark), and Trudeau's younger brother, Alexandre.* ***Ben Mulroney*** *was a guest at Trudeau's wedding. …*
  - Ben Mulroney does not appear anywhere else in the article

Entity Grid Model

Make an **entity grid** that plots entity mentions, indicate the syntactic role in which that entity appears:

Table 1
A fragment of the entity grid. Noun phrases are represented by their head nouns. Grid cells correspond to grammatical roles: subjects (S), objects (O), or neither (X).

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	S	O	S	X	O	–	–	–	–	–	–	–	–	–	–	1
2	–	–	O	–	–	X	S	O	–	–	–	–	–	–	–	2
3	–	–	S	O	–	–	–	–	S	O	O	–	–	–	–	3
4	–	–	S	–	–	–	–	–	–	–	–	S	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	S	O	–	5
6	–	X	S	–	–	–	–	–	–	–	–	–	–	–	O	6

(Barzilay and Lapata, 2008)

Document Representation

Extract a feature vector representation of a document by taking the relative frequencies of entity mention transitions (i.e., in the style of N-gram models)

Table 3
Example of a feature-vector document representation using all transitions of length two given syntactic categories s, o, x, and –.

	s s	s o	s x	s –	o s	o o	o x	o –	x s	x o	x x	x –	– s	– o	– x	– –
d_1	.01	.01	0	.08	.01	0	0	.09	0	0	0	.03	.05	.07	.03	.59
d_2	.02	.01	.01	.02	0	.07	0	.02	.14	.14	.06	.04	.03	.07	0.1	.36
d_3	.02	0	0	.03	.09	0	.09	.06	0	0	0	.05	.03	.07	.17	.39

Exercise

Build an entity grid and document representation (up to bigram transitions) for the following short passage.

Demonstrators were gathering again Saturday in cities across the United States for protests against Donald Trump.

They say he will threaten their civil and human rights.

Rallies were scheduled throughout the day in New York, Los Angeles and Chicago.

Evaluations

1. Distinguish original ordering of a document from a version where the ordering of the sentences has been randomly permutated:
   - ~90% accuracy (What is the expected random accuracy?)
2. Evaluate whether one summary is more coherent than another summary
   - ~83% pairwise accuracy
3. Readability assessment: Distinguish *Encyclopedia Britannica* from *Britannica Elementary*.
   - 88.79% accuracy, in conjunction with a pre-existing model for this task.

Extensions

Combining global and local coherence

(Elsner et al., 2007)

Modelling entity relatedness

(Filippova and Strube, 2007)

Other languages than English

(Cheung and Penn, 2010)

Rest of the Term

Applications

- Automatic summarization
- Natural language generation
- Machine translation

Current and future trends and topics

- Word embeddings
- Evaluation issues in NLP