Lecture 11: Lexical Chains and Entity-based Coherence

Lexical Semantics and Discourse Processing
MPhil in Advanced Computer Science

Simone Teufel

Natural Language and Information Processing (NLIP) Group
UNIVERSITY OF CAMBRIDGE
Simone.Teufel@cl.cam.ac.uk

Slides after Frank Keller, Regina Barzilay and Mirella Lapata

February 25, 2011

Reading:

- Silber and McCoy, 2002. Efficiently Computed Lexical Chains As an Intermediate Representation for Automatic Text Summarisation. *Computational Linguistics*
- Barzilay and Lapata, 2005. Modelling Local Coherence: An Entity-based approach. *Proceedings of ACL*.

Coherence in Text

Coherence:

- is a property of well-written texts;
- makes them easier to read and understand;
- ensures that sentences are meaningfully related;
- and that the reader can work out what expressions mean;
- the text is thematically organized;
- temporally organized;
- rather than a random concatenation of sentences.

Two solutions discussed in this lecture:

- Lexical Chains (Morris and Hearst, 1991)
- Entity-based Coherence (Barzilay and Lapata, 2005)

Coherence in Text

Summary A

Britain said he did not have diplomatic immunity. The Spanish authorities contend that Pinochet may have committed crimes against Spanish citizens in Chile. Baltasar Garzon filed a request on Wednesday. Chile said, President Fidel Castro said Sunday he disagreed with the arrest in London.

Summary B

Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998. Pinochet, 82, was recovering from surgery. The arrest was in response to an extradition warrant served by a Spanish judge. Pinochet was charged with murdering thousands, including many Spaniards. Pinochet is awaiting a hearing, his fate in the balance. American scholars applauded the arrest.

Coherence in Text

Summary A

Britain said he did not have diplomatic immunity. The Spanish authorities contend that Pinochet may have committed crimes against Spanish citizens in Chile. Baltasar Garzon filed a request on Wednesday. Chile said, President Fidel Castro said Sunday he disagreed with the arrest in London.

Summary B

Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998. Pinochet, 82, was recovering from surgery. The arrest was in response to an extradition warrant served by a Spanish judge. Pinochet was charged with murdering thousands, including many Spaniards. Pinochet is awaiting a hearing, his fate in the balance. American scholars applauded the arrest.

Lexical Chains

	LC1	LC2	LC3	LC4
S1	Pinochet	arrested		
S2	Pinochet			
S3		arrest, extradition	warrant, served	Spanish
S4	Pinochet		charged	Spaniards
S5	Pinochet		hearing	
S6		arrest		

Lexical Chain: Definition

- Sequence of related words in text, spanning short (adjacent sentences) or longer distances (entire text)
- Originally due to Halliday and Hasan (1976) *Cohesion in English*
- First CL definition in Morris and Hearst (1991), *Computational Linguistics*
- Independent of the grammatical structure of the text
- captures (some of) the cohesive structure of the text
- can provide context for the resolution of ambiguous terms
- Wordnet can be used in the identification of lexical chains.

Lexical Chains: Applications

- Summarisation (Barzilay, Elhadad 97)
- IR (Stairmand, 96)
- Detection of malapropisms (St Onge 98)
- Topic Segmentation (Hearst 97, Kazman 96)
- Hypertext link generation (Green 97, 99)
- WSD (Okumura and Honda 94)
- Word Prediction (Fazli and Hirst 2002)

Ambiguity Problem of Lexical Chains

- Allowed lexical relations: identity, synonymy, hypo/hypernymy, siblings
- Each noun occurrence must belong to only one lexical chain (in one particular sense)
- If more than one sense fits, then the overall grouping must be optimal (i.e, result in the strongest chains)

Linear-time algorithm for lexical chain computation

- Due to Silber and McCoy (2002), *Computational Linguistics*
- Barzilay and Elhadad algorithm (2007) is exponential, which makes processing chains for long documents impossible
- Idea: create "meta-chains",
- Number of meta-chains: all noun senses in WN + all nouns in document (as they might not be in WN)
- Meta-chains represent all possible chains that can contain that particular sense of a noun encountered in the text

Silber and McCoy's algorithm

1. Build meta-chains (first pass)
   - For each noun, for each sense, add sense to every metachain for which there is a lexial relation with that sense
2. Find the right interpretation (second pass)
   - For each noun, determine the meta-chain that the noun contributes to the most (based on type of relation and distance factors)
   - Distance factors: 1 sentence, 3 sentences, paragraph, default
   - Delete the word from all other meta-chains; update weights
3. Rank chains with relation to each other (third pass)
   - A strong chain is one which is more than 2 standard deviations above the mean of the chain scores in the document

Silber and McCoy: example

John has a computer. The machine is an IBM.
Meta-chains after first pass:

Index	Synset	Chain
0	person	John – machine/3(sib)
1	computer/1, data processor...	computer – machine(h) – IBM
2	computer/2, calculator/1, reckoner/1	computer
3	machine/1 (device)	machine – IBM
4	machine/2 (organisation; war machine)	machine – IBM
5	machine/4 (car)	machine – IBM
4	IBM	IBM

Evaluation: compare concepts in strong lexical chains of the full text with concepts in human-written summaries.
Evaluation Result: 79% of Strong chains have a corresponding noun in the summary; 80% of noun instances in the summary have a corresponding strong chain in the document.

Entity-based Coherence

- Coherence as a model of sequences of entity types in text
- Assume we know whether two linguistic expression **co-refer**, i.e., talk about the same entity.
- Observations from discourse theory:
  - The way entities are introduced and discussed influences coherence (Grosz et al 1995).
  - Salience of entities is related to where in the sentence they occur (Sidner, 1992).
  - Frequency, syntactic position, pronominalisation are relevant coherence properties.

The Entity Grid

1 Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998.
2 Pinochet, 82, was recovering from surgery.
3 The arrest was in response to an extradition warrant served by a Spanish judge.
4 Pinochet was charged with murdering thousands, including many Spaniards.
5 He is awaiting a hearing, his fate in the balance.
6 American scholars applauded the arrest.

The Entity Grid

1 [Former Chilean dictator Augusto Pinochet]$_S$, was arrested in [London]$_X$ on [14 October]$_X$ 1998.
2 [Pinochet]$_S$, 82, was recovering from [surgery]$_X$.
3 [The arrest]$_S$ was in [response]$_X$ to [an extradition warrant]$_X$ served by [a Spanish judge]$_S$.
4 [Pinochet]$_O$ was charged with murdering [thousands]$_O$, including many [Spaniards]$_O$.
5 [Pinochet]$_S$ is awaiting [a hearing]$_O$, [his fate]$_X$ in [the balance]$_X$.
6 [American scholars]$_S$ applauded the [arrest]$_O$.

The Entity Grid

1	Pinochet$_S$	London$_X$	October$_X$		
2	Pinochet$_S$	surgery$_X$			
3	arrest$_S$	response$_X$	warrant$_X$	judge$_O$	
4	Pinochet$_O$	thousands$_O$	Spaniards$_O$		
5	Pinochet$_S$	hearing$_O$	Pinochet$_X$	fate$_X$	balance$_X$
6	scholars$_S$	arrest$_O$			

The Entity Grid

	Pinochet	London	October	Surgery	Arrest	Extradition	Warrant	Judge	Thousands	Spaniards	Hearing	Fate	Balance	Scholars
1	**S**	**X**	**X**	–	–	–	–	–	–	–	–	–	–	–
2	**S**	–	–	**X**	–	–	–	–	–	–	–	–	–	–
3	–	–	–	–	**S**	**X**	**X**	**O**	–	–	–	–	–	–
4	**O**	–	–	–	–	–	–	–	**O**	**O**	–	–	–	–
5	**S**	–	–	–	–	–	–	–	–	–	**O**	**X**	**X**	–
6	–	–	–	–	**O**	–	–	–	–	–	–	–	–	**S**

Columns: entities; lines: sentences

Entity Transitions

Definition

A local entity transition is a sequence $\{\mathbf{S}, \mathbf{O}, \mathbf{X}, -\}^n$ that represents entity occurrences and their syntactic roles in n adjacent sentences.

Feature Vector Notation

Each grid x_{ij} for document d_i is represented by a feature vector:

$$\Phi(x_{ij}) = (p_1(x_{ij}), p_2(x_{ij}), \ldots, p_m(x_{ij}))$$

m: number of entity transitions (predefined)
$p_t(x_{ij})$: probability of transition t in grid x_{ij}

Entity Transitions

Example (transitions of length 2)

	SS	SO	SX	S–	OS	OO	OX	O–	XS	XO	XX	X–	–S	–O	–X	––
d_1	0	0	0	.03	0	0	0	.02	.07	0	0	.12	.02	.02	.05	.25
d_2	0	0	0	.02	0	.07	0	.02	0	0	.06	.04	0	0	0	.36
d_3	.02	0	0	.03	0	0	0	.06	0	0	0	.05	.03	.07	.07	.29

Linguistic Dimensions

Salience: Are some entities more important than others?

- Discriminate between salient (frequent) entities and the rest.
- Collect statistics separately for each group.

Coreference: Talking about the same entity

- Entities are coreferent if they have (roughly) the same surface form.
- Coreference resolution systems exist (cf. next lecture)

Syntax: Does syntactic knowledge matter?

- Use four categories {**S**, **O**, **X**, –}.
- Or just two {**X**, –}.

Learning a Ranking Function

Training Set

Ordered pairs (x_{ij}, x_{ik}), where x_{ij} and x_{ik} represent the same document d_i, and x_{ij} is more coherent than x_{ik} (assume $j > k$).

- Source document and permutations of its sentences.
- Original order **assumed coherent.**
- Given k documents, with n permutations, obtain $k \cdot n$ pairwise rankings for training and testing.
- Two corpora, Earthquakes and Accidents, 100 texts each.

Learning a Ranking Function

Goal

Find a parameter vector $\vec{w}$ such that:

$$\vec{w} \cdot (\Phi(x_{ij}) - \Phi(x_{ik})) > 0 \ \forall j, i, k \text{ such that } j > k$$

$\vec{w}\Phi(xij)$ is a ranking score, such that the violations of pairwise rankings in the training set are minimised.

Support Vector Machines

Constraint optimization problem can be solved using the search technique described in (Joachims 2002).

Text Ordering

Sentence 1
Sentence 2
Sentence 3
Sentence 4

Sentence 2
Sentence 3
Sentence 4
Sentence 1

Sentence 4
Sentence 3
Sentence 2
Sentence 1

Sentence 2
Sentence 1
Sentence 4
Sentence 3

Results

Model	Earthquakes	Accidents
Coreference+Syntax+Salience+	**87.2**	**90.4**
Coreference+Syntax+Salience−	88.3	90.1
Coreference+Syntax−Salience+	86.6	88.4**
Coreference−Syntax+Salience+	83.0**	89.9
Coreference+Syntax−Salience−	86.1	89.2
Coreference−Syntax+Salience−	82.3**	88.6*
Coreference−Syntax−Salience+	83.0**	86.5**
Coreference−Syntax−Salience−	81.4**	86.0**

Evaluation metric: % correct ranks in test set.
**: sig. different from Coreference+Syntax+Salience+

Results

- Omission of coreference causes performance drop.
- Syntax and Salience have more effect on Accidents corpus.
- Linguistically poor model generally worse.

Summary

Strengths:

- Novel framework for representing and measuring coherence.
- Entity grid and cross-sentential transitions.
- Suited for learning appropriate ranking function.
- Fully automatic and robust, useful for system development.

Weaknesses:

- Entity grid doesn't contain lexical information.
- Doesn't contain a notion of global coherence.
- Can't model multi-paragraph text.