Computational Models of Discourse: Introduction to Discourse: Coherence and Cohesion, Lexical Chains

Caroline Sporleder

Universität des Saarlandes

Sommersemester 2009

29.04.2009
New Schedule

29.04.2009  **Introduction Discourse: Coherence and Cohesion**
06.05.2009  **Cohesion and Local Coherence**
  - Lexical Cohesion, Lexical Chains
  - Focus, Centering
13.05.2009  **Text Segmentation**
  - TextTiling
  - Preparatory Meeting “Essay Scoring”
20.05.2009  **Applications (1)**
  - Automatic Essay Scoring
  - Preparatory Meeting “Information Ordering”
27.05.2009  **Applications (2)**
  - Information Ordering for Text Generation
  - Preparatory Meeting “Generating Referring Expressions”
03.06.2009  **Generating Referring Expressions**
  - rule-based
  - machine learning
New Schedule, cont’d

10.06.2009 Co-reference Resolution
• rule-based
• supervised machine learning
• unsupervised machine learning

17.06.2009 Discourse Parsing
• Discourse Parsing with RST
• Machine Learning

24.06.2009 Temporal Ordering

01.07.2009 Text Summarisation
• lexical chains
• RST-based
• multi-document
• argumentative zoning

08.07.2009 Sentiment Analysis

15.07.2009 Dialogue Processing
• classification of dialogue acts
• dialogue planning

22.07.2009 Speech?, Psycholinguistic Models?, Recap
Discourse Structure
Background Reading

- Jurafsky & Martin (2000)
  - Ch. 18 (Discourse)
  - Ch. 19 (Dialogue)
  - Ch. 20 (Generation)
- Jurafsky & Martin (2008)
  - Ch. 21 (Discourse)
  - Ch. 23 (Summarisation)
  - Ch. 24 (Dialogue)
What is a Discourse?

- a sequence of utterances
- **but**: an arbitrary collection of well-formed utterances is not necessarily a “discourse”
- ⇒ sequence of utterances has to be coherent
  - topics which are related
  - events which are connected (e.g. cause-result, temporal succession)
  - utterances have to **fulfil a purpose** in discourse
Coherence

Temporal sequence of events often not enough:

At 5pm a train arrived in Munich
At 6pm Angela Merkel gave a press conference
Coherence

Temporal sequence of events often not enough:

At 5pm a train arrived in Munich
At 6pm Angela Merkel gave a press conference

Topical relatedness often also not enough:

Like most bears, polar bears have 42 teeth.
Polar bears are perfectly adapted to living in the polar regions.
At the beginning of June polar bear Knut turned one and started to discover his predatory side.
Coherence

Temporal sequence of events often not enough:
At 5pm a train arrived in Munich
At 6pm Angela Merkel gave a press conference

Topical relatedness often also not enough:
Like most bears, polar bears have 42 teeth.
Polar bears are perfectly adapted to living in the polar regions.
At the beginning of June polar bear Knut turned one and started to discover his predatory side.

⇒ a discourse is coherent if a plausible discourse structure can be found
⇒ interpreting a discourse means finding the connections between individual sentences (discourse relations, co-reference chains, etc.)
Many different models of discourse. Typically it is assumed that a discourse consists of:

- **segments**
- **relations between segments** (discourse/rhetorical relations)

Discourse is structured hierarchically. A minimal/elementary discourse segment is often a clause/sentence:

\[
\forall w, e \ minimal\_segment(w, e) \Rightarrow segment(w, e)
\]

\[
\forall w_1, w_2, e_1, e_2, e \ segment(w_1, e_1) \land segment(w_2, e_2) \land DiscourseRel(e_1, e_2, e) \Rightarrow segment(w_1, w_2, e)
\]

(w is a sequence of words; e is the described event or state)

To interpret a discourse, one has to show that it is a valid segment: \( \exists e \ Segment(W, e) \)
Example: Simplified RST

```
result
  /
explanation
  /
    Peter failed the exam
  /
contrast
  /
    He had to spend the holidays preparing for the re-sit
  /
    while his friends enjoyed themselves at the beach
```
Example: Real RST

Famington police had to help control traffic recently when hundreds of people lined up to be among the first applying for jobs at the yet-to-open Mariott Hotel. The hotel's help-wanted announcement for 300 openings was a rare opportunity for many unemployed.

The people waiting in line carried a message, a refutation, of the claims that the jobless could be employed if only they showed enough ambition. Every rule has exceptions.

but the tragic and too-common tableaux of hundreds or even thousands of people snake-lining up for any task with a paycheck illustrates a lack of jobs, not laziness.
How do we know that there are segments and relations?

⇒ there are linguistic cues for the existence of both
John went to the bank to cash a cheque. Then he took the bus to his friend Bill who is a car dealer. He had to buy a car. The company for which he had just started working could not be reached by public transport. He also wanted to talk to Bill about the upcoming football match.
John went to the bank to cash a cheque. Then he took the bus to his friend Bill who is a car dealer. He had to buy a car. The company for which he had just started working could not be reached by public transport. He also wanted to talk to Bill about the upcoming football match.
Discourse segments can be referred to (Webber, 1988):

It’s always been presumed that when the glaciers receded, the area got very hot. The Folsum men couldn’t adapt, and they died out. That is what is supposed to have happened.
Discourse segments can be referred to (Webber, 1988):

It’s always been presumed that when the glaciers receded, the area got very hot. The Folsum men couldn’t adapt, and they died out. That is what is supposed to have happened.
Discourse segments can be referred to (Webber, 1988):

It’s always been presumed that when the glaciers receded, the area got very hot. The Folsum men couldn’t adapt, and they died out. That is what is supposed to have happened.
Discourse relations can be signalled by cue words (discourse markers):

- *John hid Peter’s car keys because he was drunk.*
- *Max helped Peter up again after he had fallen.*
- *Tom drinks coffee but Sue prefers tea.*
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- *John can open Bill’s safe. He knows the combination.*

---

Caroline Sporleder  
csporled@coli.uni-sb.de  
Computational Models of Discourse
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- John can open Bill’s safe. He knows the combination.

\[ \Rightarrow \text{Explanation relation (John can open Bill’s safe because John knows the combination.)} \]

John can open Bill’s safe. He will have to change the combination.

\[ \Rightarrow \text{Result relation (John can open Bill’s safe therefore Bill has to change the combination.)} \]
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- **John can open Bill’s safe. He knows the combination.**

  Explanation relation (John can open Bill's safe because John knows the combination.)

- John can open Bill's safe. He will have to change the combination.

  Result relation (John can open Bill's safe therefore Bill has to change the combination.)
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- *John can open Bill’s safe. He knows the combination.*
  ⇒ Explanation relation (John can open Bill’s safe because John knows the combination.)

- John fell. Max pushed him.
  ⇒ Result relation (John fell and therefore broke his leg.)
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- **John can open Bill’s safe. He knows the combination.**
  ⇒ **Explanation** relation (John can open Bill’s safe because John knows the combination.)

- **John can open Bill’s safe. He will have to change the combination.**
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- **John can open Bill’s safe. He knows the combination.**
  ⇒ **Explanation** relation (John can open Bill’s safe because John knows the combination.)

- **John can open Bill’s safe. He will have to change the combination.**
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- *John can open Bill’s safe. He knows the combination.*
  ⇒ *Explanation* relation (John can open Bill’s safe because John knows the combination.)

- *John can open Bill’s safe. He will have to change the combination.*
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- **John can open Bill’s safe. He knows the combination.**
  ⇒ **Explanation** relation (John can open Bill’s safe because John knows the combination.)

- **John can open Bill’s safe. He will have to change the combination.**
  ⇒ **Result** relation (John can open Bill’s safe therefore Bill has to change the combination.)
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- *John can open Bill’s safe. He knows the combination.*  
  ⇒ **Explanation** relation (John can open Bill’s safe because John knows the combination.)

- *John can open Bill’s safe. He will have to change the combination.*  
  ⇒ **Result** relation (John can open Bill’s safe therefore Bill has to change the combination.)

- *John fell. Max pushed him.*
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- **John can open Bill’s safe. He knows the combination.**  
  ⇒ Explanation relation (John can open Bill’s safe because John knows the combination.)

- **John can open Bill’s safe. He will have to change the combination.**  
  ⇒ Result relation (John can open Bill’s safe therefore Bill has to change the combination.)

- **John fell. Max pushed him.**  
  push \( <_t \) fall
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- **John can open Bill’s safe. He knows the combination.**  
  ⇒ **Explanation** relation (John can open Bill’s safe because John knows the combination.)

- **John can open Bill’s safe. He will have to change the combination.**  
  ⇒ **Result** relation (John can open Bill’s safe therefore Bill has to change the combination.)

- **John fell. Max pushed him.**  
  ⇒ **Explanation** relation (John fell because Max pushed him.)

"push \( <_t \) fall"
Linguistic reality of discourse relations

Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- **John can open Bill’s safe. He knows the combination.**
  ⇒ Explanation relation (John can open Bill’s safe because John knows the combination.)

- **John can open Bill’s safe. He will have to change the combination.**
  ⇒ Result relation (John can open Bill’s safe therefore Bill has to change the combination.)

- **John fell. Max pushed him.**
  push $<_t$ fall
  ⇒ Explanation relation (John fell because Max pushed him.)

- **John fell. He broke a leg.**
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- **John** can open **Bill's safe. He knows the combination.**
  ⇒ **Explanation** relation (John can open Bill's safe because John knows the combination.)

- **John** can open **Bill's safe. He will have to change the combination.**
  ⇒ **Result** relation (John can open Bill's safe therefore Bill has to change the combination.)

- **John fell. Max pushed him.**  
  push $<_t$ fall  
  ⇒ **Explanation** relation (John fell because Max pushed him.)

- **John fell. He broke a leg.**  
  fall $<_t$ breaking a leg
Discourse relations influence linguistic interpretation (anaphora resolution, temporal ordering etc.):

- **John can open Bill’s safe. He knows the combination.**
  \(\Rightarrow\) **Explanation** relation (John can open Bill’s safe because John knows the combination.)

- **John can open Bill’s safe. He will have to change the combination.**
  \(\Rightarrow\) **Result** relation (John can open Bill’s safe therefore Bill has to change the combination.)

- **John fell. Max pushed him.** \(\text{push} <_t \text{fall}\)
  \(\Rightarrow\) **Explanation** relation (John fell because Max pushed him.)

- **John fell. He broke a leg.** \(\text{fall} <_t \text{breaking a leg}\)
  \(\Rightarrow\) **Result** relation (John fell and therefore broke his leg.)
If a text is **cohesive** it hangs together well.

If a text is **coherent** it makes sense.
If a text is cohesive it hangs together well.
⇒ Cohesion is about linguistic form.

If a text is coherent it makes sense.
If a text is **cohesive** it hangs together well.  
⇒ Cohesion is about linguistic form.

If a text is **coherent** it makes sense.  
⇒ Coherence is about underlying structure.
Peter failed the exam because he didn’t study hard enough. He had to spend the holidays preparing for the re-sit while his friends enjoyed themselves at the beach.
Peter failed the exam because he didn’t study hard enough. He had to spend the holidays preparing for the re-sit while his friends enjoyed themselves at the beach.
Peter failed the exam because he didn’t study hard enough. He had to spend the holidays preparing for the re-sit while his friends enjoyed themselves at the beach.
Peter failed the exam because he didn’t study hard enough. He had to spend the **holidays** preparing for the re-sit while his friends **enjoyed** themselves at the **beach**.
Example: Coherence with little cohesion

Yesterday Peter passed his driving test. Afterwards Peter went to see Klaus. Klaus was happy about the visit because Klaus hadn’t seen Peter for a while. Later Peter and Klaus went to the pub.
Example: Coherence with little cohesion

Yesterday Peter passed his driving test. Afterwards Peter went to see Klaus. Klaus was happy about the visit because Klaus hadn’t seen Peter for a while. Later Peter and Klaus went to the pub.
Yesterday Peter flew to Australia.
This country is well-known for its kangaroos.
The kangaroos in Cologne Zoo are a favorite of Carla’s.
She likes to travel.
Gnus are nice animals.
Example: Cohesion with little Coherence

Yesterday Peter flew to **Australia**.

**This country** is well-known for its kangaroos.

The kangaroos in Cologne Zoo are a favorite of Carla’s.

She likes to travel.

Gnus are nice animals.
Yesterday Peter flew to Australia.
This country is well-known for its kangaroos.
The kangaroos in Cologne Zoo are a favorite of Carla’s.
She likes to travel.
Gnus are nice animals.
Yesterday Peter flew to Australia.
This country is well-known for its kangaroos.
The kangaroos in Cologne Zoo are a favorite of Carla’s.
She likes to travel.
Gnus are nice animals.
Yesterday Peter flew to Australia.
This country is well-known for its kangaroos.
The kangaroos in Cologne Zoo are a favorite of Carla’s.
She likes to travel.
Gnus are nice animals.
...is not just about coherence and cohesion.
Four interdependent aspects/dimensions of discourse structure:

- **(Para-)Linguistic Structure**: linguistic manifestation of discourse structure, e.g., lexical cohesions, cue words, intonation, gesture, referring expressions etc.

- **Intentional Structure**: each discourse segment fulfils a purpose (why does a speaker/write make a given utterance in a given form?)

- **Informational Structure**: how do the different segments of a discourse relate to each other (which segments are directly related and which discourse relations hold)?

- **Focus/Attentional Structure**: which entities are *salient* at a given point in discourse?
Linguistic structure is about cohesion.

Intentional, informational, and focus structure are about coherence.
Linguistic Structure

Linguistic form
often an indicator of discourse structure:

- **discourse connectives** (*but, because*):
  ⇒ reflect how sentences are related to each other (contrast, explanation etc.)

- **referring expressions** (*she, Mary, a girl, the girl who likes ice-cream . . .*)
  ⇒ reflect the status of an entity in the discourse (salient, not-salient, new, old, inferred etc.)

- **semantically related words** (*flooding . . . torrential rain . . . storm*):
  ⇒ reflect lexical cohesion
John hid Peter’s car keys. He was drunk.
John hid Peter’s car keys. He was drunk.

⇒ The fact that John was drunk explains why he hid Peter’s car keys.
John hid Peter’s car keys. He was drunk.

⇒ The fact that John was drunk explains why he hid Peter’s car keys.

Mary likes chocolate, Maggie likes crisps
John hid Peter’s car keys. He was drunk.

⇒ The fact that John was drunk explains why he hid Peter’s car keys.

Mary likes chocolate, Maggie likes crisps

⇒ The fact that Maggie likes crisps contrasts with Mary’s liking of chocolate.
John hid Peter’s car keys. He was drunk.
John hid Peter’s car keys. He was drunk.

Possible intention: explain to listener why John hid Peter’s keys (and why Peter was consequently late for work)
John hid Peter’s car keys. He was drunk.

Possible intention: explain to listener why John hid Peter’s keys (and why Peter was consequently late for work)

Another Possible intention: outline to listener what consequences John’s drunkenness has (and why something must be done about his binge drinking)
Susan would like to go on a holiday. But she needs to find somebody to do her work while she’s away. She can’t think of anybody to do that. She considered Mike but he’s a bit unreliable. Yesterday he forgot to turn up for an important meeting with a client. The client was very annoyed and said she would never do business with Susan’s company again.
Susan would like to go on a holiday. But she needs to find somebody to do her work while she’s away. She can’t think of anybody to do that. She considered Mike but he’s a bit unreliable. Yesterday he forgot to turn up for an important meeting with a client. The client was very annoyed and said she would never do business with Susan’s company again.
Susan would like to go on a holiday. But she needs to find somebody to do her work while she’s away. She can’t think of anybody to do that. She considered Mike but he’s a bit unreliable. Yesterday he forgot to turn up for an important meeting with a client. The client was very annoyed and said she would never do business with Susan’s company again.
Susan would like to go on a holiday. But she needs to find somebody to do her work while she’s away. She can’t think of anybody to do that. She considered Mike but he’s a bit unreliable. Yesterday he forgot to turn up for an important meeting with a client. The client was very annoyed and said she would never do business with Susan’s company again.
Not all four aspects of discourse structure are equally easy to model
Not all four aspects of discourse structure are equally easy to model

- **linguistic structure**: easy to observe
Not all four aspects of discourse structure are equally easy to model

- **linguistic structure**: easy to observe
- **focus structure**: relatively strong correlation with linguistic form (pronouns, salient positions in a sentence (subject vs. object)), fairly easy to model
Not all four aspects of discourse structure are equally easy to model

- **linguistic structure:** easy to observe
- **focus structure:** relatively strong correlation with linguistic form (pronouns, salient positions in a sentence (subject vs. object)), fairly easy to model
- **informational structure:** relatively weak correlation with linguistic form (discourse connectives), difficult to model
Not all four aspects of discourse structure are equally easy to model

- **linguistic structure**: easy to observe
- **focus structure**: relatively strong correlation with linguistic form (pronouns, salient positions in a sentence (subject vs. object)), fairly easy to model
- **informational structure**: relatively weak correlation with linguistic form (discourse connectives), difficult to model
- **intentional structure**: barely visible in linguistic form, extremely difficult to model on analysis side
How would you go about modelling discourse structure?

On the analysis side all we typically have is linguistic form ⇒ model those aspects of discourse structure which correlate strongest with linguistic form, use linguistic form as cues for structure → focus structure: Centering Theory → linguistic structure, lexical cohesion: Lexical Chains

We'll talk about informational structure (discourse parsing) later.
On the analysis side all we typically have is linguistic form
On the analysis side all we typically have is linguistic form

⇒ model those aspects of discourse structure which correlate strongest with linguistic form, use linguistic form as cues for structure
How would you go about modelling discourse structure?

On the analysis side all we typically have is linguistic form

⇒ model those aspects of discourse structure which correlate strongest with linguistic form, use linguistic form as cues for structure

→ focus structure: Centering Theory

→ linguistic structure, lexical cohesion: Lexical Chains
How would you go about modelling discourse structure?

On the analysis side all we typically have is linguistic form

⇒ model those aspects of discourse structure which correlate strongest with linguistic form, use linguistic form as cues for structure

→ focus structure: Centering Theory

→ linguistic structure, lexical cohesion: Lexical Chains

We’ll talk about informational structure (discourse parsing) later
Lexical Chains
Lexical Chains

... are sequences of semantically related words

*From Nineteen Eighty-Four [abridged]:*

... the book that he had just taken out of the drawer. It was a peculiarly beautiful book. Its smooth creamy paper was of a kind that had not been manufactured for at least forty years past. He could guess, however, that the book was much older than that. He had seen it lying in the window of a frowsy little junk-shop and had been stricken immediately by an overwhelming desire to possess it. Party members were supposed not to go into ordinary shops. He had slipped inside and bought the book for two dollars fifty. Even with nothing written in it, it was a compromising possession. The thing that he was about to do was to open a diary. Winstor fitted a nib into the penholder and sucked it to get the grease off. The pen was an archaic instrument, seldom used even for signatures, and he had procured one, furtively and with some difficulty, simply because of a feeling that the beautiful creamy paper deserved to be written on with a real nib instead of being scratched with an ink-pencil. Actually he was not used to writing by hand. He dipped the pen into the ink.

(Source: Graeme Hirst & Alexander Budanitsky, Eurolan-2001 presentation)
We need:

- a measure of semantic relatedness between words
- a chain building algorithm
Computing Semantic Relatedness

... a hot research topic.

**Two basic methods:**

- concept distance in a hierarchical lexicon (e.g. WordNet)
- distributional similarity computed from a corpus
Structure

- **synsets**: collections of words with the same sense (e.g., {bank, depository financial institution, banking concern, banking company} vs. {bank, river bank})

- **relations** between synsets
  - hyponym (e.g., Federal Reserve Bank)
  - hypernym (e.g., financial organisation)
  - member holonym (e.g., banking industry)
  - antonyms
  - etc.
**Simple approach**

- count path length between two concepts/synsets
- possibly normalise by overall depth of hierarchy etc.

**But:**

not all paths are equal, e.g. changes in direction weaken similarity
Three types of relations:

1. **extra-strong**: literal repetition of a word

2. **strong**:
   - concepts are in the same synset (e.g. *human* and *person*), or
   - the concept’s synsets are connected by a horizontal link (antonymy or similarity relation, e.g. *precursor* and *successor*), or
   - one of the words is a compound that includes the other (e.g. *school* and *private school*)

3. **medium-strong**: there exists an allowable path between the concept’s synsets
   Medium-strong paths are weighted by:
   \[ C - \text{path length} - k \times \text{number of changes of direction} \]
   \((C \text{ and } k \text{ are empirically set constants})\)
An allowable path:

- contains no more than five links and
- conforms to one of eight patterns, definable by the following rules
  - no other direction may precede an upward link
  - at most one change of direction is allowed
  - it is permitted to use a horizontal link to make a transition from an upward to a downward direction
Figure 2: (a) Patterns of paths allowable in medium-strong relations and (b) patterns of paths not allowable. (Each vector denotes one or more links in the same direction.)
Figure 3: Example of a regular relation between two words. (@ = hypernymy, ~ = hyponymy)
Problems/Disadvantages

- availability of hierarchical lexicon
- coverage of lexicon
- sparse and dense areas in hierarchy not comparable (normalisation necessary, not a solved problem)
- only "classical" relations (hypernymy, hyponymy, antonymy etc., can't model fuzzy relations, e.g. fire and coals)
Problems/Disadvantages

- availability of hierarchical lexicon
- coverage of lexicon
- sparse and dense areas in hierarchy not comparable (normalisation necessary, not a solved problem)
- only “classical” relations (hypernymy, hyponymy, antonymy etc., can’t model fuzzy relations, e.g. fire and coals)
Possible measures:

- **Pointwise Mutual Information (PMI):**
  \[ I(x, y) = \log_2 \frac{P(x,y)}{P(x)P(y)} \]

- PMI over-inflates low-frequency events, better:
  \[ I_{corrected}(x, y) = \log_2 \frac{P(x,y)}{P(x)P(y)} \times \frac{\min(\text{freq}(x),\text{freq}(y))}{\min(\text{freq}(x),\text{freq}(y))+1} \]

- Cosine of the angle between the co-occurrence vectors of two words
Problems/Disadvantages
Relatedness based on distributional similarity

Problems/Disadvantages

- conflation of word senses
- sometimes unpredictable results (corpus size, domain, similarity measure etc. play a role)
Basic Idea:
place two words in the same chain if their relatedness is above threshold $t$
Basic Idea:
place two words in the same chain if their relatedness is above threshold $t$

Design decisions

- can a word be placed in several chains?
- does a word have to be related to all other words in the chain or just to one other word? If it has to be related to all words does the avg. similarity have to be above the threshold or the minimum/maximum?
- greedy vs. non-greedy chain building (and its interaction with word-sense disambiguation)
Figure 5: A word starting a new chain. (The word *man* has six synsets.)
Figure 6: Push the same word
Figure 7: Push an antonym
Figure 8: Updated chain after insertion
Lexical Chains. So what are they good for?