Overview of Natural Language Processing Part II & ACS L90

Lecture 11: Discourse

Simone Teufel

Department of Computer Science and Technology
University of Cambridge

Michaelmas 2021/22

If we consider a discourse relation as a relationship between two phrases, we get a binary branching tree structure for the discourse. In many relationships, such as Explanation, one phrase depends on the other: e.g., the phrase being explained is the main one and the other is subsidiary. In fact we can get rid of the subsidiary phrases and still have a reasonably coherent discourse.

putting sentences together (in text)

Lecture 11: Discourse

1. Coherence
2. Anaphora (pronouns etc)
3. Algorithms for anaphora resolution

most slides by
Ann Copestake

Coherence

Document structure and discourse structure

Most types of document are highly structured, implicitly or explicitly:

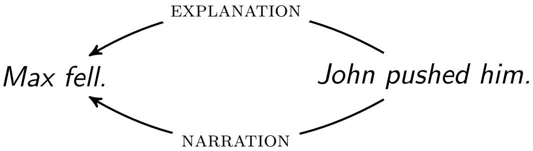
- Scientific papers: conventional structure (differences between disciplines).
- News stories: first sentence is a summary.
- Blogs, etc etc

Three Generative, Lexicalised Models for Statistical Parsing

Abstract: *In this paper we first propose [...] We then [...] Results on Wall Street Journal text show [...]*

1. **Introduction**: *[...] In the remainder of this paper we describe the 3 models in section 2, discuss practical issues in section 3, give results in section 4, and give conclusions in section 5.*
2. **The Three Parsing Models**
3. **Practical Issues**
4. **Results**
5. **Conclusions**

www.aclweb.org/anthology/P97-1003.pdf



1. Max fell *because* John pushed him. ▷ EXPLANATION
2. Max fell *and then* John pushed him. ▷ NARRATION

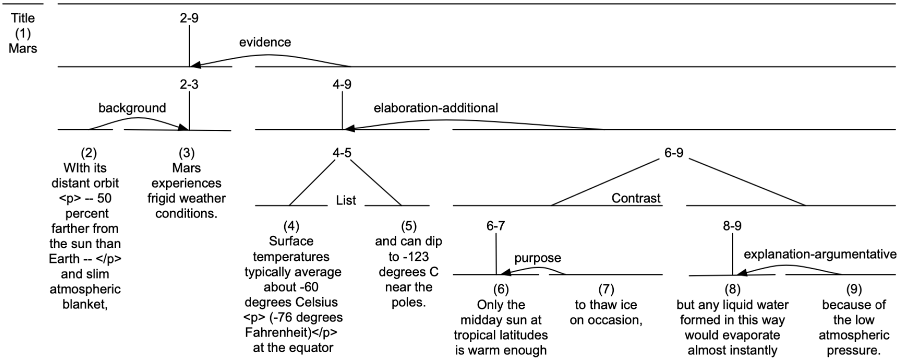
Rhetorical Structure Theory

Analysis of text with rhetorical relations generally gives a binary branching structure. Relations are defined between *elementary discourse unit (EDU)*/*discourse segments*, which can be a sentence, clause or phrase.

- *nucleus*: more central unit to the writer's purpose that is interpretable independently
- *satellite*: less central unit that is only interpretable with respect to the nucleus. ▷$\mathcal{R}$: nucleus←satellite
- *equal weight*: e.g., NARRATION, LIST, CONTRAST

A large text is organized with a hierarchical structure

With its distant orbit–50 percent farther from the sun than Earth–and slim atmospheric blanket, Mars experiences frigid weather conditions. Surface temperatures typically average about -60 degrees Celsius (-76 degrees Fahrenheit) at the equator and can dip to -123 degrees C near the poles. Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, but any liquid water formed in this way would evaporate almost instantly because of the low atmospheric pressure.



Title
(1)
Mars
2-9
evidence
2-3
background
4-9
elaboration-additional
(2) WIth its distant orbit <p> -- 50 percent farther from the sun than Earth -- </p> and slim atmospheric blanket,
(3) Mars experiences frigid weather conditions.
4-5
List
(4) Surface temperatures typically average about -60 degrees Celsius <p> (-76 degrees Fahrenheit)</p> at the equator
(5) and can dip to -123 degrees C near the poles.
6-9
Contrast
6-7
purpose
(6) Only the midday sun at tropical latitudes is warm enough
(7) to thaw ice on occasion,
8-9
explanation-argumentative
(8) but any liquid water formed in this way would evaporate almost instantly
(9) because of the low atmospheric pressure.

Coherence in interpretation

Discourse coherence assumptions can affect interpretation

Kim's bike got a puncture. She phoned the AA.

Assumption of coherence (and knowledge about the AA) leads to bike interpreted as motorbike rather than pedal cycle.

John likes Bill. He gave him an expensive Christmas present.

- If EXPLANATION - *he* is probably *Bill*.
- If JUSTIFICATION (supplying evidence for first sentence), *he* is *John*.

Summarisation by satellite removal

Goal: shortening long pieces of text

If we consider a discourse relation as a relationship between two phrases, we get a binary branching tree structure for the discourse. In many relationships, such as EXPLANATION*, one phrase depends on the other: e.g., the phrase being explained is the main one and the other is subsidiary. In fact we can get rid of the subsidiary phrases and still have a reasonably coherent discourse.*

Coherence in generation

Language generation needs to maintain coherence

In trading yesterday: Dell was up 4.2%, Safeway was down 3.2%, HP was up 3.1%.

Better:

Computer manufacturers gained in trading yesterday: Dell was up 4.2% and HP was up 3.1%. But retail stocks suffered: Safeway was down 3.2%.

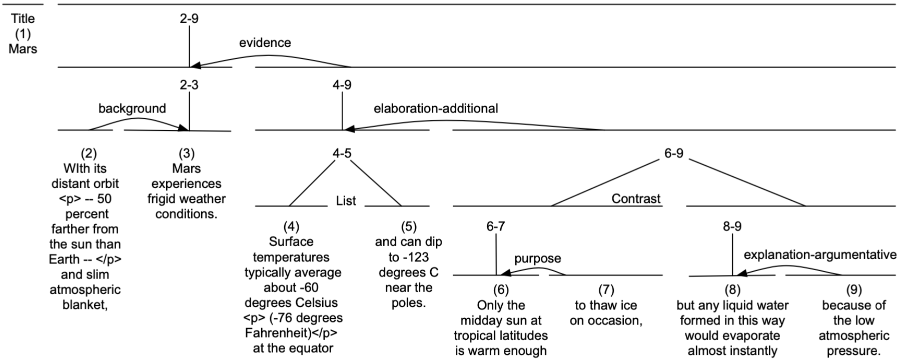
More about generation in the next lecture.

Factors influencing discourse interpretation

1. Cue phrases
   - *because*, *in order to*
2. Punctuation (also prosody) and text structure.
   - *Max fell (John pushed him) and Kim laughed.*
   - *Max fell, John pushed him and Kim laughed.*
3. Real world content:
   - *Max fell. John pushed him as he lay on the ground.*
4. Tense and aspect.
   - *Max fell. John had pushed him.*
   - *Max was falling. John pushed him.*

Hard problem, but ‘surfacy techniques’ (punctuation and cue phrases) work to some extent.

Discourse structure parsing


Title
(1)
Mars
2-9
evidence
2-3
background
4-9
elaboration-additional
(2)
WIth its distant orbit <p> -- 50 percent farther from the sun than Earth -- </p> and slim atmospheric blanket,
(3)
Mars experiences frigid weather conditions.
4-5
List
6-9
Contrast
(4)
Surface temperatures typically average about -60 degrees Celsius <p> (-76 degrees Fahrenheit)</p> at the equator
(5)
and can dip to -123 degrees C near the poles.
6-7
purpose
8-9
explanation-argumentative
(6)
Only the midday sun at tropical latitudes is warm enough
(7)
to thaw ice on occasion,
(8)
but any liquid water formed in this way would evaporate almost instantly
(9)
because of the low atmospheric pressure.

Two subtasks

- EDU segmentation
- RST parsing

Anaphora (pronouns etc)

Referring expressions

Niall Ferguson is prolific, well-paid and a snappy dresser. Stephen Moss hated him — at least until he spent an hour being charmed in the historian's Oxford study.

referent: a real world entity that some piece of text (or speech) refers to, e.g. the actual Prof. Ferguson

referring expressions/mentions: bits of language used to perform reference by a speaker, e.g. *Niall Ferguson*, *he*, *him*, *the historian*

corefer: two or more referring expressions refer to the same referent

antecedent: the text initially evoking a referent, e.g. *Niall Ferguson*

anaphora: the phenomenon of referring to an antecedent.

coreference resolution: the task of determining whether two mentions corefer

entity linking/entity resolution: the task of mapping a mention in text to the representation of some real-world entity in an *ontology*, e.g. Wikipedia pages (wikification)

Event coreference

Event variable

(1) a. *Rover barked loudly*

b. $\exists e[\text{bark'}(e, r) \wedge \text{loud'}(e)]$

AMD agreed to buy Markham, Ontario-based ATI for around $5.4 billion in cash and stock, the companies announced Monday.
The acquisition would turn AMD into one of the world's largest providers of graphics chips.

event coreference: the task of deciding whether two event mentions refer to the same event

Hard constraints

Pronoun agreement (PERSON/NUMBER/GENDER)

(2) a. *A little girl* is at the door — see what *she* wants, please?

b. My dog has hurt *his* foot — *he* is in a lot of pain.

c. *My dog has hurt *his* foot — *it* is in a lot of pain.

Complications

(3) a. *The team* played really well, but now *they* are all very tired.

b. *Kim and Sandy* are asleep: *they* are very tired.

c. *Kim* is snoring and *Sandy* can't keep her eyes open: *they* are both exhausted.

Reflexives

(4) a. John$_i$ cut himself$_i$ shaving. (*himself*=*John*, subscript as ID)

b. #John$_i$ cut him$_j$ shaving. ($i \neq j$ — a very odd sentence)

Reflexive pronouns must be coreferential with a preceeding argument of the same verb, non-reflexive pronouns cannot be.

Non-referring expressions

pleonastic/expletive pronouns: semantically empty, don't refer

(5) a. It is snowing

b. It is not easy to think of good examples.

c. It is obvious that Kim snores.

d. It bothers Sandy that Kim snores.

predicative NPs: properties of the head nouns

(6) the 38-year-old became *the company's president*

appositional NPs: supplementary parenthetical descriptions

(7) Victoria Chen, *CFO of Megabucks Banking*, saw . . .

generics: classes rather than individuals or sets of individuals

(8) I love *mangos*. *They* are very tasty.

Soft preferences: Salience

Recency
Kim has a big car. Sandy has a smaller one. Lee likes to drive it.

Grammatical role: SUBJECTS $\succ$ OBJECTS $\succ$ everything else
Fred went to the Grafton Centre with Bill. He bought a hat.

Repeated mention: Entities that have been mentioned more frequently are preferred.

Parallelism: Entities which share the same role as the pronoun in the same sort of sentence are preferred
Bill went with Fred to the Grafton Centre. Kim went with him to Lion Yard.

World knowledge and common sense reasoning

(9) a. The city council denied the demonstrators a permit because

b. they feared violence. ▷*they*=*the city council*

c. they advocated violence. ▷*they*=*the demonstrators*

Sometimes inference will override soft preferences:

Andrew Strauss again blamed the batting after England lost to Australia last night. They now lead the series three-nil.

they=*Australia*

But a discourse can be odd if strong salience effects are violated:

The England football team won last night. Scotland lost.
?*They have qualified for the World Cup with a 100% record.*

Algorithms for Anaphora Resolution

Anaphora resolution as supervised classification

- *Classification*: training data labelled with class and features, derive class for test data based on features.
- For *potential pronoun/antecedent pairings*, class is TRUE/FALSE.
- Assume candidate antecedents are all NPs in current sentence and preceeding 5 sentences (excluding pleonastic pronouns)

Example

Niall Ferguson is prolific, well-paid and a snappy dresser. Stephen Moss hated him — at least until he spent an hour being charmed in the historian's Oxford study.

Issues: detecting pleonastic pronouns and predicative NPs, deciding on treatment of possessives (*the historian* and *the historian's Oxford study*), named entities (e.g., *Stephen Moss*, not *Stephen* and *Moss*), allowing for cataphora, . . .

Features

Cataphoric: `true` if pronoun before antecedent.

Number agreement: `true` if pronoun compatible with antecedent.

Gender agreement: `true` if gender agreement.

Same verb: `true` if the pronoun and the candidate antecedent are arguments of the same verb.

Sentence distance: $\{0, 1, 2, \ldots\}$ (discrete)

Grammatical role (the role of the potential antecedent): $\{$ SUBJECT, OBJECT, other $\}$

Parallel: `true` if the potential antecedent and the pronoun share the same grammatical role.

Linguistic form: $\{\text{PROPER}, \text{DEFINITE}, \text{INDEFINITE}, \text{PRONOUN}\}$

Problems with simple classification model

- Cannot implement ‘repeated mention’ effect.
- Cannot use information from previous links (not a structured prediction solution)
- Not really pairwise: really need discourse model with real world entities corresponding to clusters of referring expressions.

Zero anaphors

zero anaphor/zero pronoun: in some languages (including Chinese, Japanese, and Italian) it is possible to have an anaphor that has no lexical realization at all.

EN John$_i$ went to visit some friends. On the way he$_i$ bought some wine.

IT Giovanni$_i$ andò a far visita a degli amici. Per via $\emptyset_i$ comprò del vino.

JA ジョン$_i$-は 友人を 訪問 した. 途中 $\emptyset_i$ ワインを買った。

zero anaphor resolution

Readings

- Ann's lecture notes.
  `https://www.cl.cam.ac.uk/teaching/1920/NLP/materials.html`
- D Jurafsky and J Martin. *Speech and Language Processing*.
  Chapter 22. Coreference resolution.
  `https://web.stanford.edu/~jurafsky/slp3/22.pdf`