

# The Linguistics of Discourse Structure

Shallow Discourse Parsing

ESSLLI 2019, Riga

Tatjana Scheffler

[tatjana.scheffler@uni-potsdam.de](mailto:tatjana.scheffler@uni-potsdam.de)

Day 1, August 12, 2019

# This course

topic & organization

# Me

Tatjana Scheffler

[tatjana.scheffler@uni-potsdam.de](mailto:tatjana.scheffler@uni-potsdam.de)

@tschfflr



# Shallow Discourse Parsing

According to Lawrence Eckenfelder, a securities industry analyst at Prudential-Bache Securities Inc., "Kemper is the first firm to make a major statement with program trading." He added that "having just one firm do this isn't going to mean a hill of beans. But if this prompts others to consider the same thing, then it may become much more important."

# Shallow Discourse Parsing

According to Lawrence Eckenfelder, a securities industry analyst at Prudential-Bache Securities Inc., "Kemper is the first firm to make a major statement with program trading." He added that "**having just one firm do this isn't going to mean a hill of beans. But if this prompts others to consider the same thing, then it may become much more important.**"

Connective: **but**

First argument (Arg1); second argument (Arg2)

Sense (discourse relation): Comparison.Concession

# Shallow Discourse Parsing

According to Lawrence Eckenfelder, a securities industry analyst at Prudential-Bache Securities Inc., "**Kemper is the first firm to make a major statement with program trading.**" He added that "**having just one firm do this isn't going to mean a hill of beans.** But if this prompts others to consider the same thing, then it may become much more important."

Connective: (implicit)

First argument (Arg1); second argument (Arg2)

Sense (discourse relation): Comparison.Congrue

# organization of the course

- theoretical and practical parts
- theory: discourse structure, (shallow) discourse parsing
- praxis: exploring data, playing with parsers
- you:
  - Ask me questions!
  - Participate

# Why discourse structure?

# Why discourse?

- (1) Max fell. Peter helped him get up.
- (2) Max fell. Peter pushed him.
- (3) Max fell. He likes spinach.

Narration / Elaboration

Explanation / Cause

Definition:

A discourse is a coherent sequence of sentences/utterances.

# coherence exercise

- a) Only in public TV programmes can viewers get information that is not profit-oriented.
- b) But to therefore offer TV according to audience rates only would be dangerous.
- c) The thought of television exclusively made up of private channels scares me.
- d) And since it is the duty of society to collectively finance certain areas that cannot hold their own in a market economy, everyone should chip in.
- e) Admittedly trash and turmoil are exactly what the majority of people want to see.

# coherence exercise (solution)

1. The thought of television exclusively made up of private channels scares me.
2. Admittedly trash and turmoil are exactly what the majority of people want to see.
3. But to therefore offer TV according to audience rates only would be dangerous.
4. Only in public TV programmes can viewers get information that is not profit-oriented.
5. And since it is the duty of society to collectively finance certain areas that cannot hold their own in a market economy, everyone should chip in.

(Potsdam Microtext corpus)

# coherence

- Temporal sequence is not sufficient:

At 5am, a train arrived in Munich.

At 6am, Angela Merkel gave a press conferences.

- Thematic connection not sufficient:

Like most bears, polar bears have 42 teeth.

Polar bears' size is adapted optimally to their polar habitat.

At the beginning of June, Knut turned one year old.

# discourse

- many different approaches
- discourse consists of:
  - segments
  - connections/relations between segments (coherence relations)
- discourse is structured hierarchically:
  - $\forall w, e \text{ minimal segment}(w, e) \Rightarrow \text{segment}(w, e)$
  - $\forall w_1, w_2, e_1, e_2, e \text{ segment}(w_1, e_1) \wedge \text{segment}(w_2, e_2) \wedge \text{cohRel}(e_1, e_2, e) \Rightarrow \text{segment}(w_1 + w_2, e)$   
( $w$  = a sequence of words;  $e$  = a described event)

# ling. reality of segments (1)

Sue mailed a package at the post office.

Then, she took the bus to Mary's car dealership.

She wanted to buy a new car.

Her new office can't be reached easily by public transport.

She also wanted to talk with Mary about soccer practice.

# ling. reality of segments (2)

- referring to discourse entities (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.

# ling. reality of segments (2)

- ❑ Referring to discourse entities (Webber, 1991):

According to Kim, Sue just bought a 1962 Ford Falcon.

- ❑ ... but **that** turned out to be a lie
- ❑ ... but **that** was false
- ❑ **That** struck me as a funny way to describe the situation
- ❑ **That** caused Sue to become rather poor

# coherence relations

- (1) John hid Peter's car keys. He was drunk.
- (2) Lisa fell. Lara helped her up.
- (3) Lina likes to eat chocolate. Julia prefers chips.

# ling. evidence for coherence relations

- coherence relations influence interpretation:
  - (1) Indira can open Elsa's safe. She knows the combination.
  - (2) Indira can open Elsa's safe. She has to change the combination.
  - (3) Max fell. Peter helped him.
  - (4) Max fell. Peter pushed him.

fall <<sub>t</sub> help

push <<sub>t</sub> fall

# coherence signals

- ❑ connectives

- ❑ phrases

- (1) John hid Peter's car keys **because** he was drunk.
- (2) Lisa fell, **and then** Lara helped her up.
- (3) Indira can open Elsa's safe. **That's why** she has to change the code.

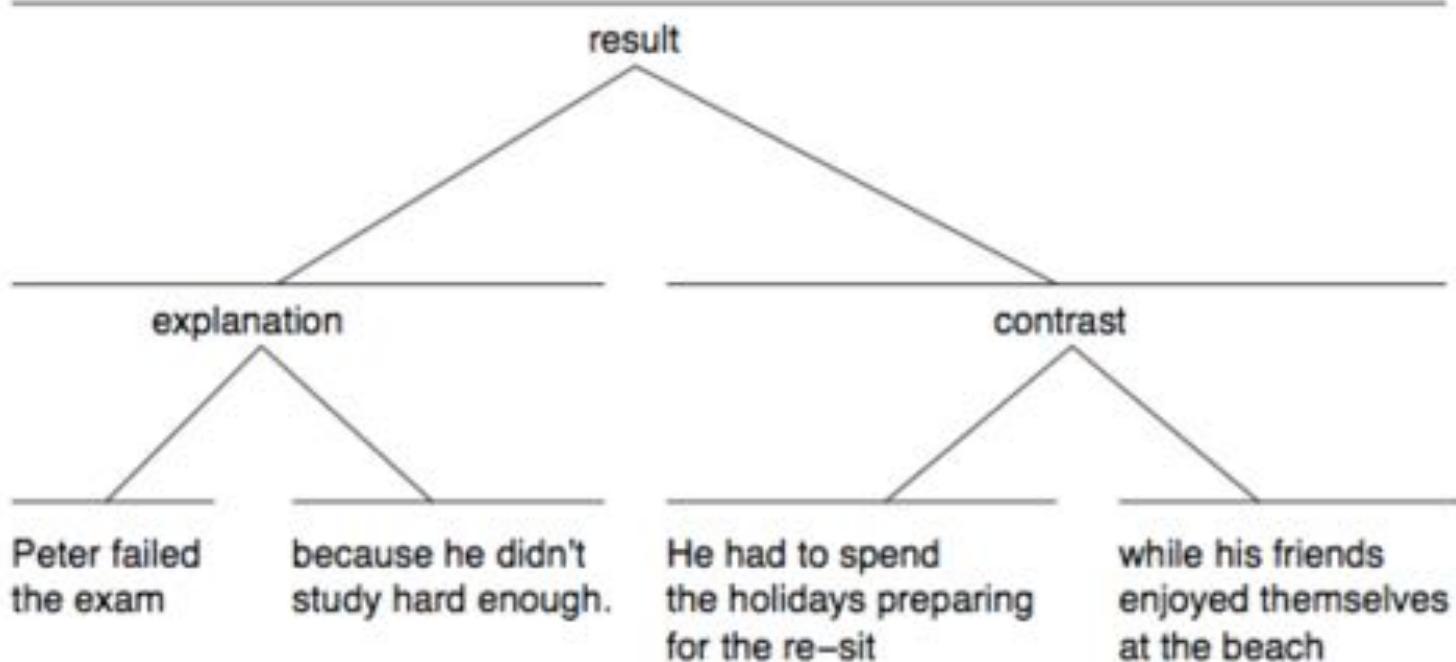
# coherence vs. cohesion

- (Halliday & Hasan 1976)
- coherence: structural relation between discourse segments
- cohesion: non-structural textual relations, e.g. reference (anaphora), ellipsis, lexical cohesion

# coherence: example

1. Peter failed the exam
2. because he didn't study hard enough.
3. He had to spend the holidays preparing for the re-sit
4. while his friends enjoyed themselves at the beach.

# coherence: example



# cohesion: example

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.

# Local Coherence

(Shallow Discourse Structure)

# discourse coherence

- (1) Last night I went for a walk. I love going for a walk at nighttime.
- (2) Last night I went for a walk. I love the fresh air in the park.
- (3) Last night I went for a walk. I love goldfish.

# discourse relations

■ not restricted to intersentential relations:

- (1a) *A jogger* was hit by a car in Palo Alto last night. (Hobbs, 1990)
- (1b) *A farmer* was hit by a car in Palo Alto last night.

(3a) The company fired *the manager who was embezzling money*.  
(Rohde et al, 2011)

- (3b) The company fired *the manager who was hired in 2002*.
- (3c) The company fired *the manager who has a long history of corporate awards*.

# local coherence

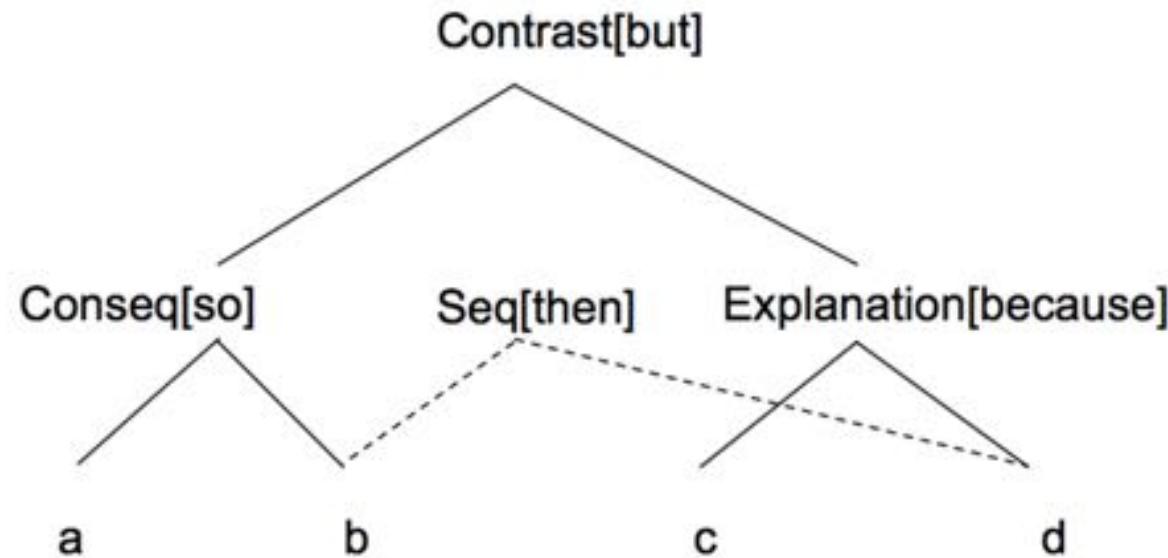
- one view: hierarchical discourse structure
  - ambiguity/vagueness
  - author intentions
  - not all text types contain global coherence
  - no long distance dependencies
- instead: local coherence
  - “theory neutral”
  - lexically bound
  - reliable annotation
  - flat structure

-> Day 4, RST!

“one step up from  
sentence syntax”

# problems with global structure

- a) John loves Barolo. (Webber)
- b) So he ordered three cases of the '97.
- c) But he had to cancel the order.
- d) Because then he discovered that he was broke.



# Penn Discourse Treebank

- <https://www.seas.upenn.edu/~pdtb/>
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi and Bonnie Webber
- Wall Street Journal-articles (Penn Treebank): 1 mio words

# Annotation

- connectives and their arguments
  - conjunctions, adverbials (100 in the PDTB)
  - minimal text spans
- also:
  - implicit connectives
  - alternative lexicalizations (AltLex)
  - entity relations (EntRel)
  - no relation (NoRel)

# discourse ≈ text grammar

## **sentence level**

sentence w/ words

verb

0-5 arguments

## **discourse level**

text w/ sentences

connective

exactly 2 arguments

# explicit connectives

- U.S. Trust, a 136-year-old institution that is one of the earliest high-net worth banks in the U.S., has faced intensifying competition from other firms that have established, and heavily promoted, private-banking businesses of their own. **As a result, U.S. Trust's earnings have been hurt.**
- connective from a well-defined class
- Arg1
- Arg2
- arguments can have different sizes

# implicit connectives

- But a few funds have taken other defensive steps. Some have raised their cash positions to record levels. Implicit = BECAUSE High cash positions help buffer a fund when the market falls.
- between adjacent segments
- inferred relation
- connective can be inserted

# AltLex

- ❑ relation exists
- ❑ inserting a connectives leads to inconsistency/redundancy

Ms. Bartlett's previous work, which earned her an international reputation in the non-horticultural art world, often took gardens as its nominal subject. AltLex Mayhap this metaphorical connection made the BPC Fine Arts Committee think she had a literal green thumb.

# EntRel

- coherence relation based on mere entity coreference

Hale Milgrim, 41 years old, senior vice president, marketing at Elecktra Entertainment Inc., was named president of Capitol Records Inc., a unit of this entertainment concern.  
EntRel Mr. Milgrim succeeds David Berman, who resigned last month.

# NoRel

- ❑ rare
- ❑ no coherence relation between the two sentences

Jacobs is an international engineering and construction concern. **NoRel Total capital investment at the site could be as much as \$400 million, according to Intel.**

# Plan for this course

# Discourse Parsing in 5 Days

**Day 1: The Linguistics of Discourse Structure**

**Day 2: Exploring Discourse Annotated Corpora**

**Day 3: The Computation of Discourse Structure**

**Day 4: Shallow vs. Deep Structure**

**Day 5: Specifics of Shallow Discourse Parsing**

# Further reading:

- CONLL Shared Tasks on Shallow Discourse Parsing:
  - <http://www.cs.brandeis.edu/~clp/conll15st/>
  - <http://www.cs.brandeis.edu/~clp/conll16st/>
- C. Potts' class on working with the PDTB:
  - <http://compprag.christopherpotts.net/pdtb.html>

# Prepare for tomorrow!

- ❑ Please download the PDTB corpus (do not share this!)  
<https://bit.ly/2TjfT5K>
- ❑ Please download the Python notebook and modules for working with the corpus:  
<https://github.com/TScheffler/2019ESSLLI-discparsing>  
(under day2)