

# Shallow Discourse Structure – Ressources & Annotation

Shallow Discourse Parsing

ESSLI 2019, Riga

Tatjana Scheffler

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

Day 2, August 13, 2019

# Penn Discourse Treebank

(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*.

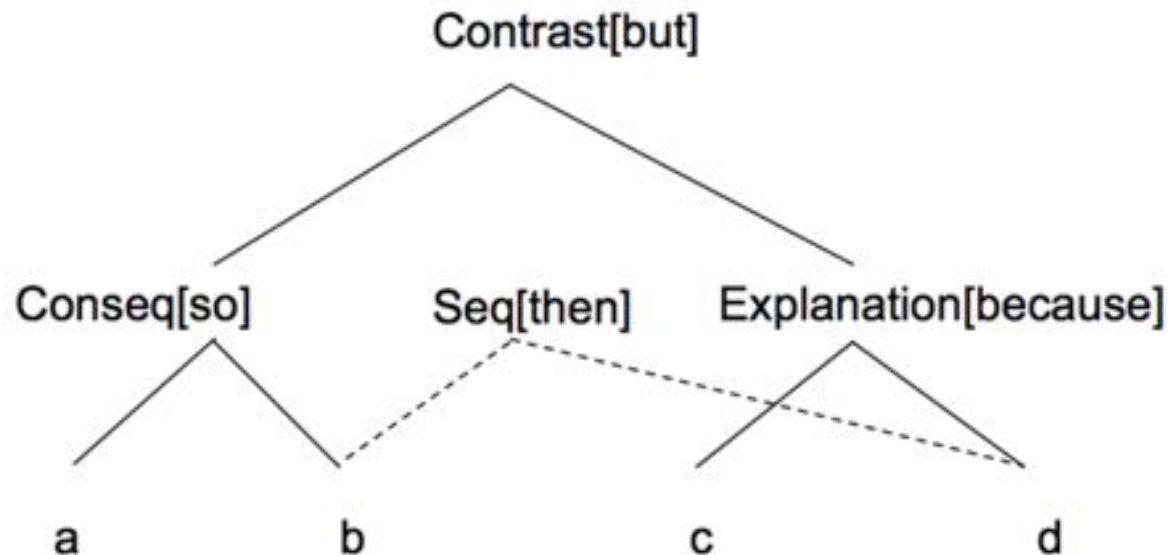
(3c) The company fired *the manager who has a long history of corporate awards*.

# local coherence

- one view: hierarchical discourse structure
  - ambiguity/vagueness → Day 4, RST!
  - author intentions
  - not all text types contain global coherence
  - no long distance dependencies
- instead: local coherence
  - “theory neutral” → “one step up from sentence syntax”
  - lexically bound
  - reliable annotation
  - flat structure

# 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 $\approx$ 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**
- *sense(s)*

# discourse connectives

- (1) Tom's car broke down. **Fortunately**, he was near a garage when it happened.
- (2) Tom's car broke down. **In addition**, his cellphone didn't work when he tried to call the AAA.
- Discourse connective in the PDTB:
  - adverbials or conjunctions (subordinating/coordinating)
  - denoting a semantic relation between the two arguments
  - arguments must be (approximately) clauses (or bigger)

# connectives in English

- Coordinating conjunctions:

(1) John likes apples, **but** Mary prefers pears.

- Subordinating conjunctions:

(2) John likes apples, **because** they are juicy.

- Adverbials:

(3) John likes apples. **However**, these ones are sour.

# Arguments

- ▣ Arguments can be of different sizes
  - ▣ clause
  - ▣ sentence
  - ▣ multiple sentences or clauses
  - ▣ (in rare cases, nominal phrase or discourse deictic referring to an abstract entity)
  - ▣ minimality requirement
- ▣ Arg2 – the one syntactically closer to the connective, Arg1 – the other (external) one
- ▣ (for implicits, Arg1 then Arg2)

# 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
- sense(s)

# 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.**

# PDTB 2.0

PDTB Relations	No. of tokens
Explicit	18459
Implicit	16224
AltLex	624
EntRel	5210
NoRel	254
Total	40600

Table 2: Distribution of Relations in PDTB-2.0

# connective ambiguity

- (1) Yesterday I sat on the balcony and read a book.
  - (2) Yesterday I had a nice dinner and took a walk.
  - (3) Yesterday I crossed the street on a red light and was almost hit by a car.
- 
- sense annotation (disambiguation) for explicit and implicit connectives

Level-1	Level-2	Level-3
TEMPORAL	SYNCHRONOUS	–
	ASYNCHRONOUS	PRECEDENCE
		SUCCESSION
CONTINGENCY	CAUSE	REASON
		RESULT
		NEGRESULT
	CAUSE+BELIEF	REASON+BELIEF
		RESULT+BELIEF
	CAUSE+SPEECHACT	REASON+SPEECHACT
		RESULT+SPEECHACT
	CONDITION	ARG1-AS-COND
		ARG2-AS-COND
	CONDITION+SPEECHACT	–
	NEGATIVE-CONDITION	ARG1-AS-NEGCOND
		ARG2-AS-NEGCOND
	NEGATIVE-CONDITION+SPEECHACT	–
	PURPOSE	ARG1-AS-GOAL
		ARG2-AS-GOAL
COMPARISON	CONCESSION	ARG1-AS-DENIER
		ARG2-AS-DENIER
	CONCESSION+SPEECHACT	ARG2-AS-DENIER+SPEECHACT
	CONTRAST	–
	SIMILARITY	–
EXPANSION	CONJUNCTION	–
	DISJUNCTION	–
	EQUIVALENCE	–
	EXCEPTION	ARG1-AS-EXCPT
		ARG2-AS-EXCPT
	INSTANTIATION	ARG1-AS-INSTANCE
		ARG2-AS-INSTANCE
	LEVEL-OF-DETAIL	ARG1-AS-DETAIL
		ARG2-AS-DETAIL
	MANNER	ARG1-AS-MANNER
		ARG2-AS-MANNER
	SUBSTITUTION	ARG1-AS-SUBST
		ARG2-AS-SUBST

# ambiguity

- 
- (10) *The Mountain View, Calif., company has been receiving 1,000 calls a day about the product since it was demonstrated at a computer publishing conference several weeks ago.*
- (11) *It was a far safer deal for lenders since NWA had a healthier cash flow and more collateral on hand.*
- (12) *Domestic car sales have plunged 19% since the Big Three ended many of their programs Sept. 30.*
-

# attribution

- association of quotes with their sources

When **Mr. Green won a \$240,000 verdict in a land condemnation case against the state in 1983**, he says Judge O'Kicki unexpectedly awarded him an additional \$100,000.

Advocates said *the 90cent-an-hour rise, to \$4.25 an hour, is too small for the working poor*, while opponents argued **that the increase will still hurt small business and cost many thousands of jobs.**

# 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**



# Class material

- 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)