Shallow Discourse Parsing

Grammarly, Kyiv

Tatjana Scheffler

tatjana.scheffler@uni-potsdam.de

February 29, 2020

Me

Tatjana Scheffler

tatjana.scheffler@uni-potsdam.de

@tschfflr

course repository: https://github.com/TScheffler/2020grammarly_ws/

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. <u>But</u> if this prompts others to consider the same thing, then it may become much more important."**

Connective: **<u>but</u>**

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.Contrast

Course schedule

10:15-11:30	Theory of Discourse Structure
11:45-13:30	Discourse Annotations
lunch	
14:30-16:00	Introduction to Shallow Discourse Parsing
16:15-17:30	Discourse Parser Implementations

Please ask me questions at any time!

Why discourse structure?

Why discourse?

(1) Susy fell. Mary helped her get up.

(2) Susy fell. Mary pushed her.

(3) Susy fell. She 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:

  $\forall w,e$ minimal segment$(w,e) \Rightarrow$ segment(w,e)

  $\forall w1, w2, e1, e2, e$ segment$(w1, e1) \wedge$ segment$(w2, e2) \wedge$ cohRel$(e1, e2, e) \Rightarrow$ segment$(w1 + w2, e)$
  (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 Ellie’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 Ellie 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, Anya 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 Anya 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 $<_t$ help push $<_t$ fall

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*.

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.

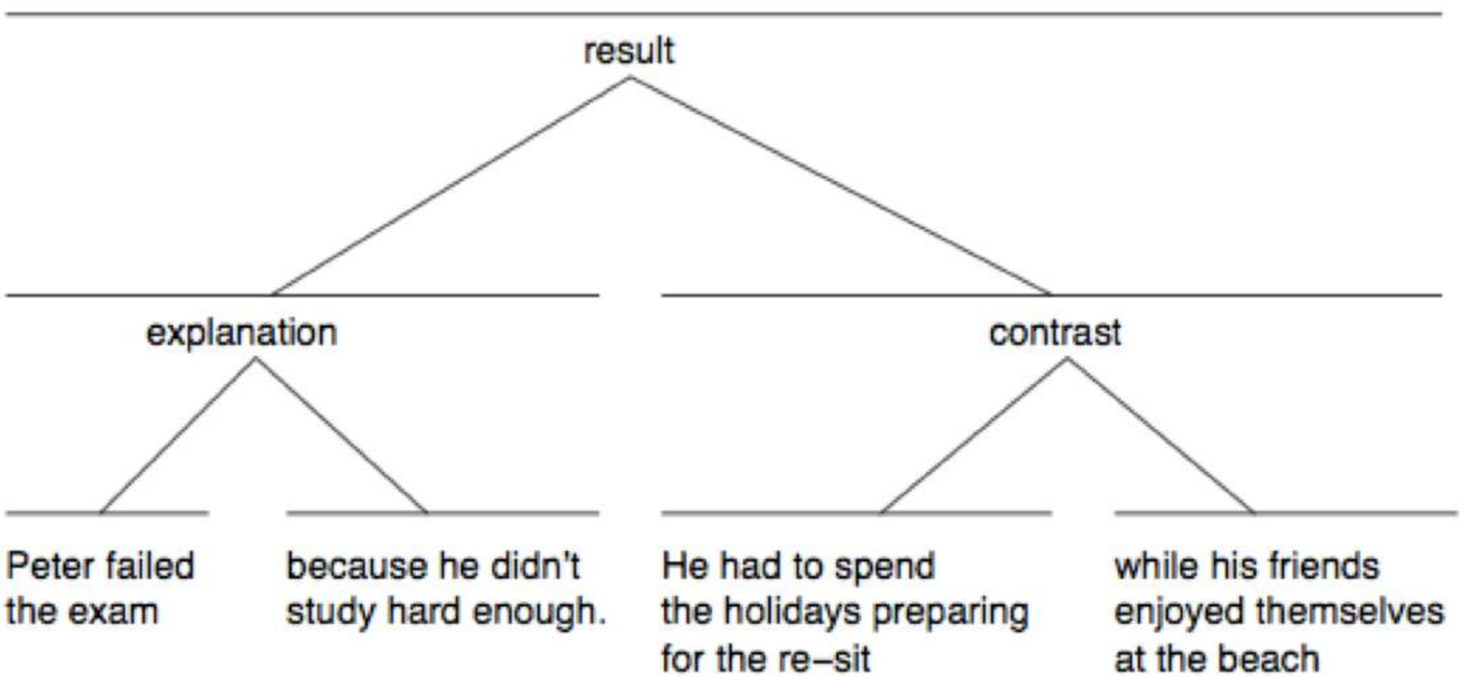
coherence vs. cohesion

- (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)

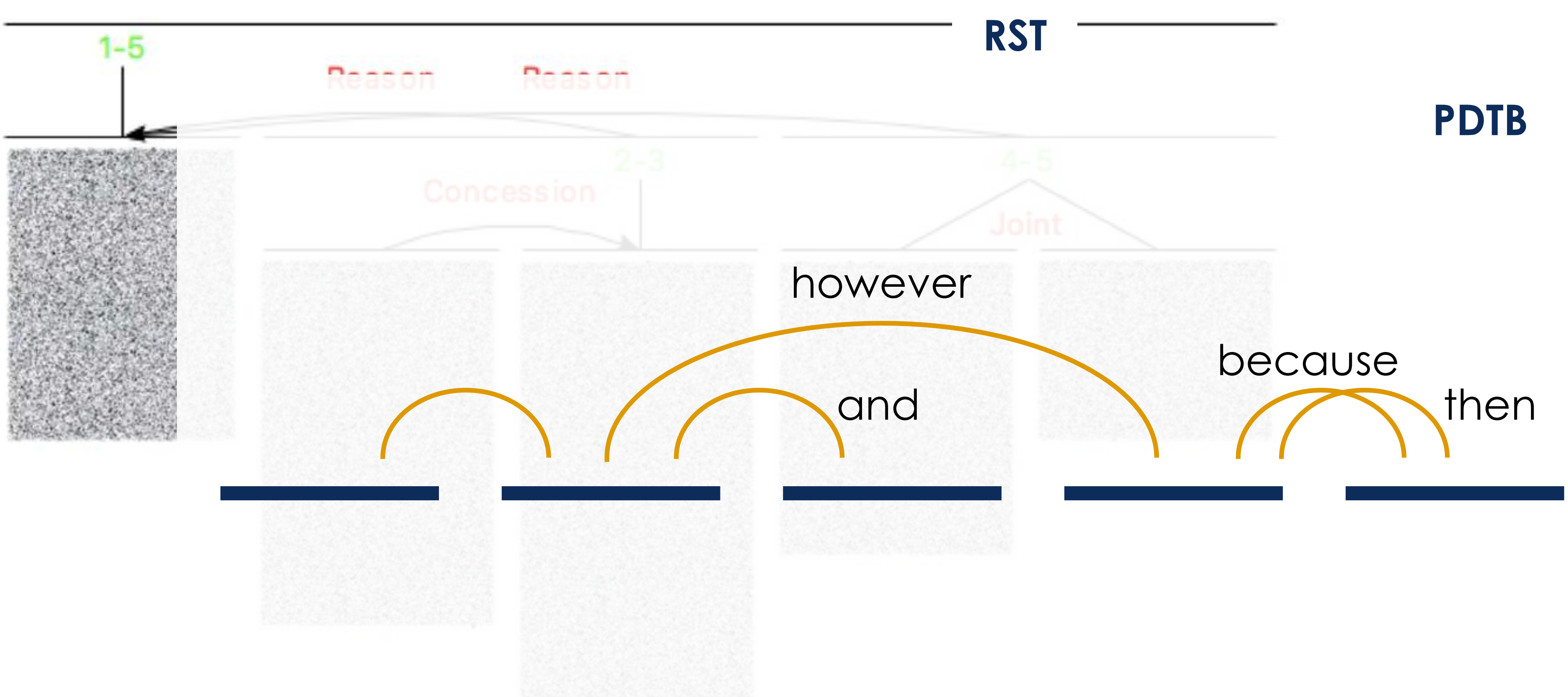
discourse coherence

(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.

Shallow vs. Deep Structure



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

“one step up from sentence syntax”

problems with global structure

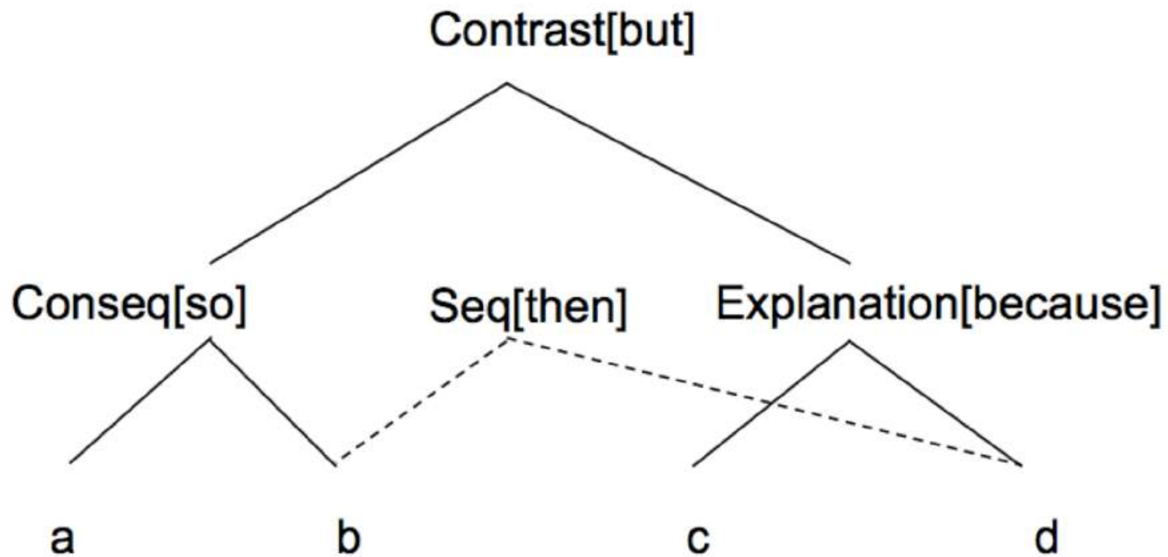
(Webber)

a) John loves Barolo.

b) So he ordered three cases of the '97.

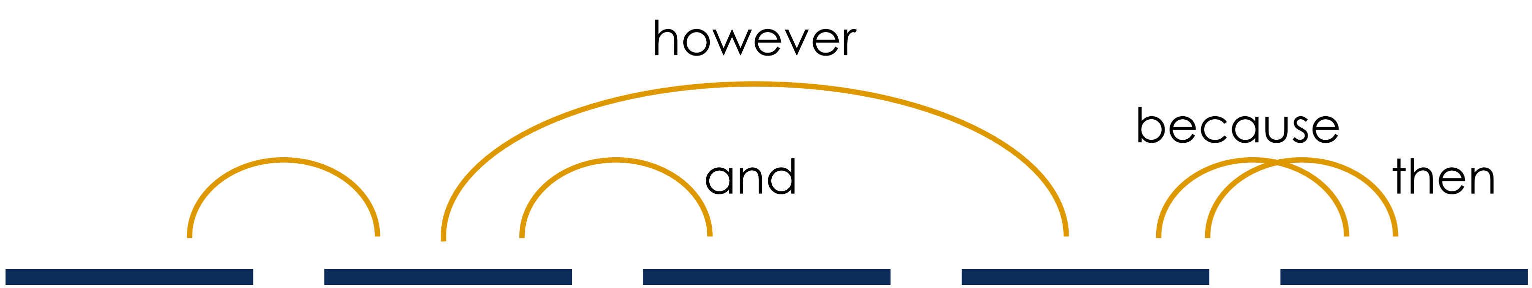
c) But he had to cancel the order.

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	**discourse level**
sentence w/ words	text w/ sentences
verb	connective
0-5 arguments	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

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
  - denoting a semantic relation between the two arguments
  - arguments must be (approximately) clauses

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.

definition of connectives

- Criteria:
  - closed-class words
  - non-inflectable
  - semantics: two-place relation
  - semantic relation links two eventualities that could be expressed as full clauses
- Syntactic categories:
  - conjunctions, coordinating and subordinating
  - certain adverbials
  - certain prepositions: despite, due to, ...
  - NO: free phrases (for this reason)
  - NO: affixes like -wise / -halber, -wegen

Connectives (example)

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)

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 in-ternational reputation in the non-horticultural art world, often took gardens as its nominal subject. <u>AltLex **Mayhap this metaphorical connection made**</u> **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. <u>EntRel</u> **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. <u>NoRel</u> **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* <u>since</u> **it was demonstrated at a computer publishing conference several weeks ago.**

(11) *It was a far safer deal for lenders* <u>since</u> **NWA had a healthier cash flow and more collateral on hand.**

(12) *Domestic car sales have plunged 19%* <u>since</u> **the Big Three ended many of their programs Sept. 30.**

attribution

- association of quotes with their sources

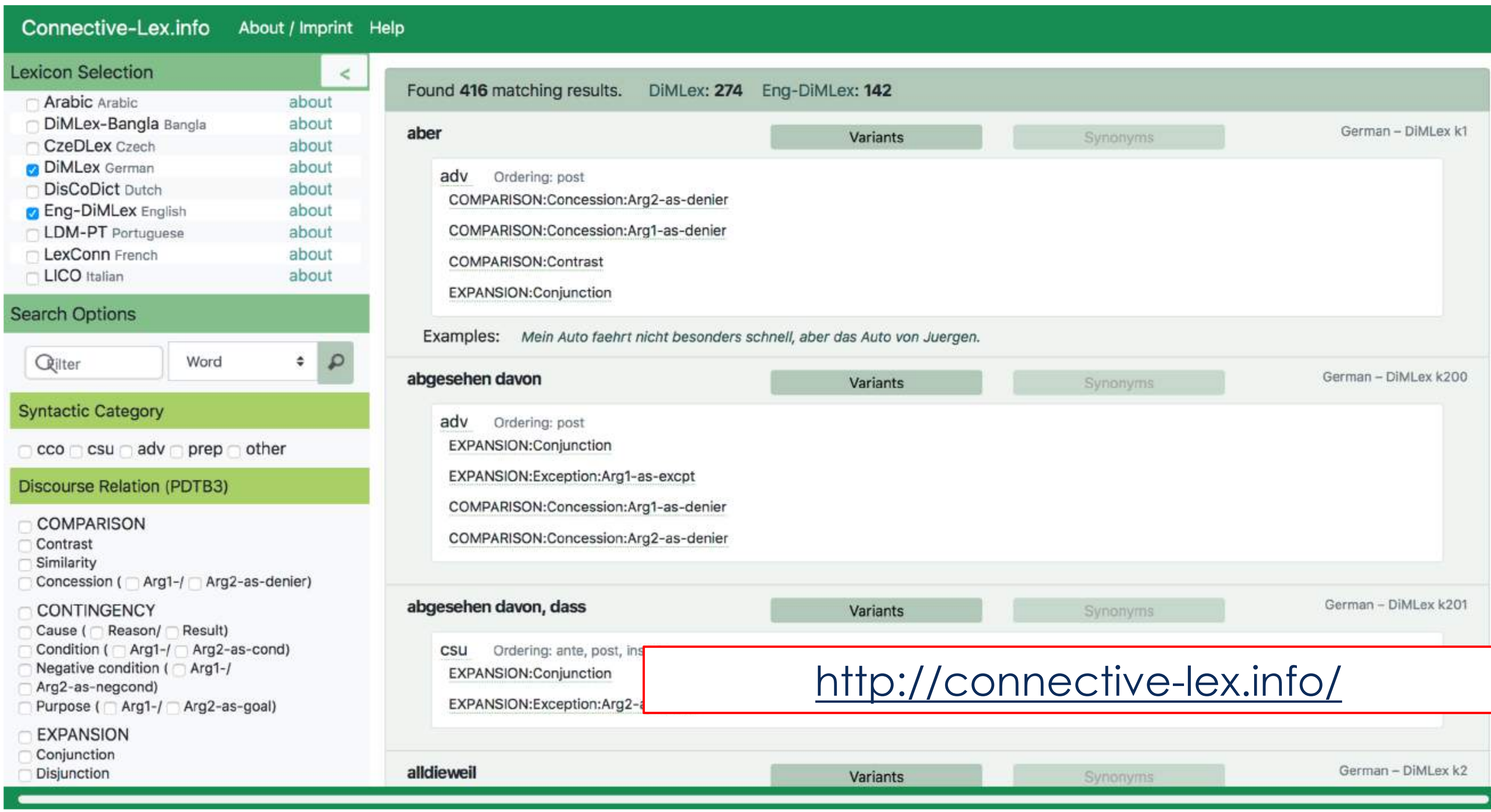
When **Mr. Green won a $240,000 verdict in a land condemnation case against the state in 1983**, he says Judge O'Kicki unexpectedly awarded him an additional $100,000.

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.**

Connectives in Ukrainian

group work

Multilingual connective lexicon



http://connective-lex.info/

(Stede, Scheffler, Mendes 2019)

Task (part 1)

1. Determine whether the given candidate word (or phrase) is a connective:
   - it is a fixed (closed class) expression that cannot be modified or conjugated
   - it semantically relates to arguments
   - the arguments are "abstract objects" (propositions, facts, events, ...)
   - the arguments are in principle expressible as clauses

2. After determining the connectives (from 1.), group them into semantic classes:

   temporal // contingency (causal + conditional) // contrast (contrast and similarity) // expansion (additive relations)

Task (part 2)

3. Group connectives by substitution tests: which connectives can replace each other in the examples?

4. Determine the syntactic and semantic class of each connective and enter it in the Google sheet:

- https://docs.google.com/spreadsheets/d/134_SxwVkoPpMlsyxZ2uThlaYBQ6rK0XxnmVpRKOjeNs/edit?usp=sharing

Explore the PDTB

based on Pott’s Computational Pragmatics class (2011)

Possible data questions

- What is the length of arguments (separately by connective)?
- What are the root node labels for Arg1/2 for each connective?
- What is the difference between Explicit and Implicit connectives?
- What is the difference between Explicit and Implicit senses?

Shallow Discourse Parsing

CONLL Shared Task (2015/16)

- https://www.cs.brandeis.edu/~clp/conll16st/

CoNLL-2016 Shared Task | Shallow Discourse Parsing | Rules | Results | Registration | Data & Resources | Blog | Forum

CoNLL 2016 Shared Task

Multilingual Shallow Discourse Parsing

Task period : January 15 - April 24, 2016

CoNLL Conference : August 11 - 12, 2016 in Berlin, Germany

Multilingual Shallow Discourse Parsing

This is the 2nd edition of the CoNLL Shared Task on Shallow Discourse Parsing, following the first edition in 2015. A participant system is given a piece of newswire text as input and returns discourse relations in the form of a discourse connective (explicit or implicit) taking two arguments (which can be clauses, sentences, or multi-sentence segments). Specifically, the participant system needs to

1. locate both explicit (e.g., "because", "however") discourse connectives in the text
2. identify the spans of text that serve as the two arguments for each discourse connective
3. predict the sense of the discourse connectives (e.g., "Cause", "Contrast")

Recognizing such discourse relations is an important part of natural language understanding, which benefits a wide range of natural language applications. More detail and examples.

What's new this year?

There are a few things. More detail will be provided later

Official blind test sets in English and Chinese

The task has already concluded. We have released the blind test sets used in English and Chinese.

Joining the shared task

The instructions are the same whether you would like to participate in both languages and/or just the supplementary task.

1. Complete the registration form (one per team)
2. Submit the license agreement form to LDC
3. Download the data from the link, which LDC will send to you after a few days
4. Check out the resources that might be useful
5. Login to the evaluation platform on tira.io using your credential, which we will send to you
6. Clone/fork from our github repo and familiarize yourself with the data format
7. Start developing the parser

Stay updated

CONLL Shared Task (2015/16)

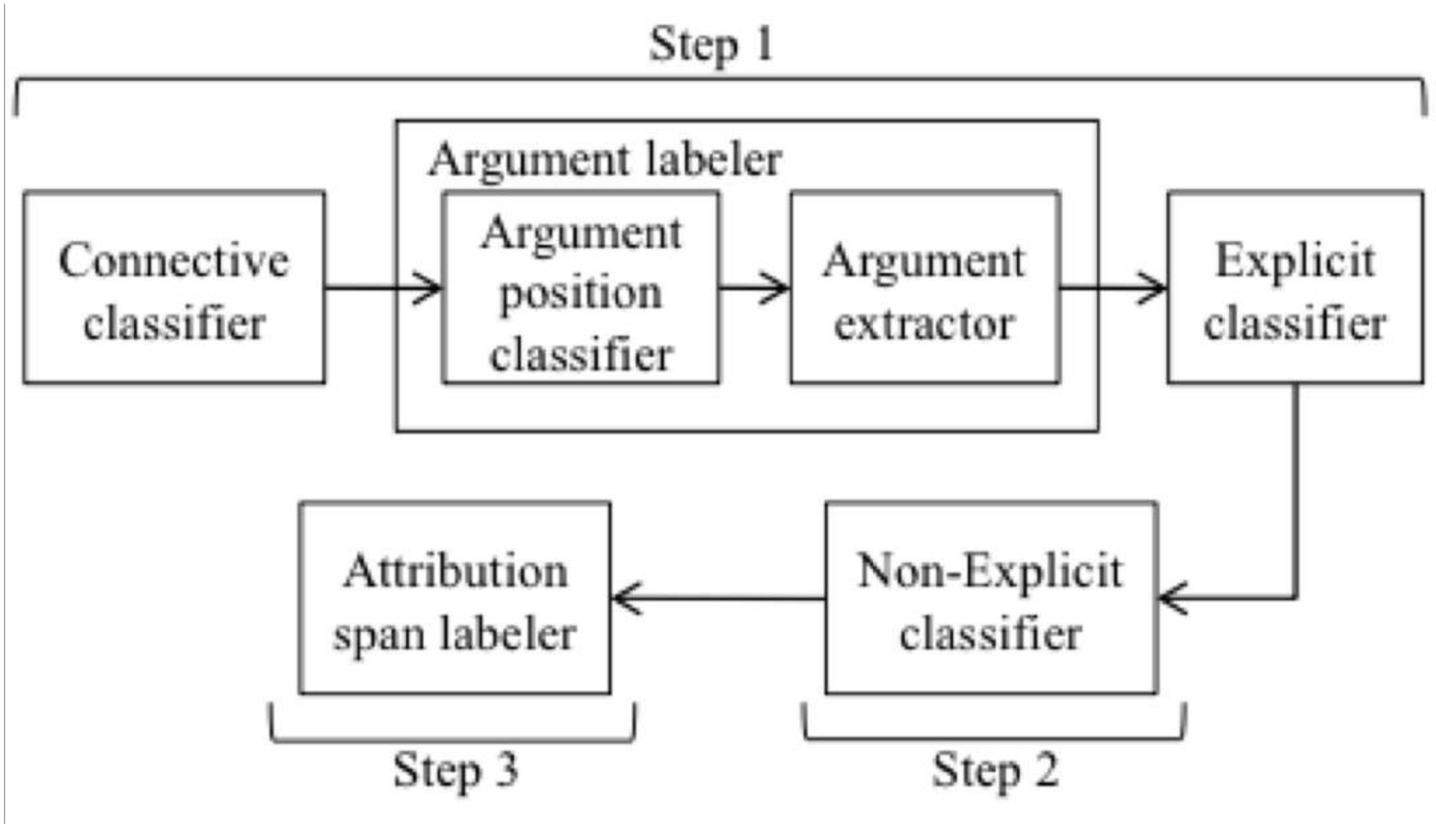
- https://www.cs.brandeis.edu/~clp/conll16st/
- common data sets (training, dev) and test platform
- well-defined problem
- standard input and output formats
- https://nbviewer.jupyter.org/github/attapol/conll16st/blob/master/tutorial/tutorial.ipynb

Task

- Identify the text span of an explicit discourse connective, if present, or the position between adjacent sentences as the proxy site of an implicit discourse relation
- Identify the two text spans that serve as arguments to the relation
- Label the argument spans as Arg1 or Arg2, as appropriate
- Predict the sense of the discourse relation (e.g., "Cause", "Condition", "Contrast")

PDTB-style Discourse Parser

Lin et al. (2012)



https://www.comp.nus.edu.sg/~kanmy/papers/nleLin2012.pdf

Evaluation

- Gold standard (GS) parses without error propagation

1. connective classifier

- assume a list of explicit connectives
- disambiguation problem
- MaxEnt classifier (logistic regression)
- syntactic and lexical features: connective, its POS-tag and immediate context, syntactic sisters and path to root
- F_1 = 95.76%

2. argument position

- relative position of Arg1 and Arg2
- same sentence, previous sentence
- features:
  - position of connective
  - contextual features
- Component $F_1 = 97.94\%$

3. argument extractor

Input: a discourse connective C and the text T
Output: Arg1 and Arg2 spans of C

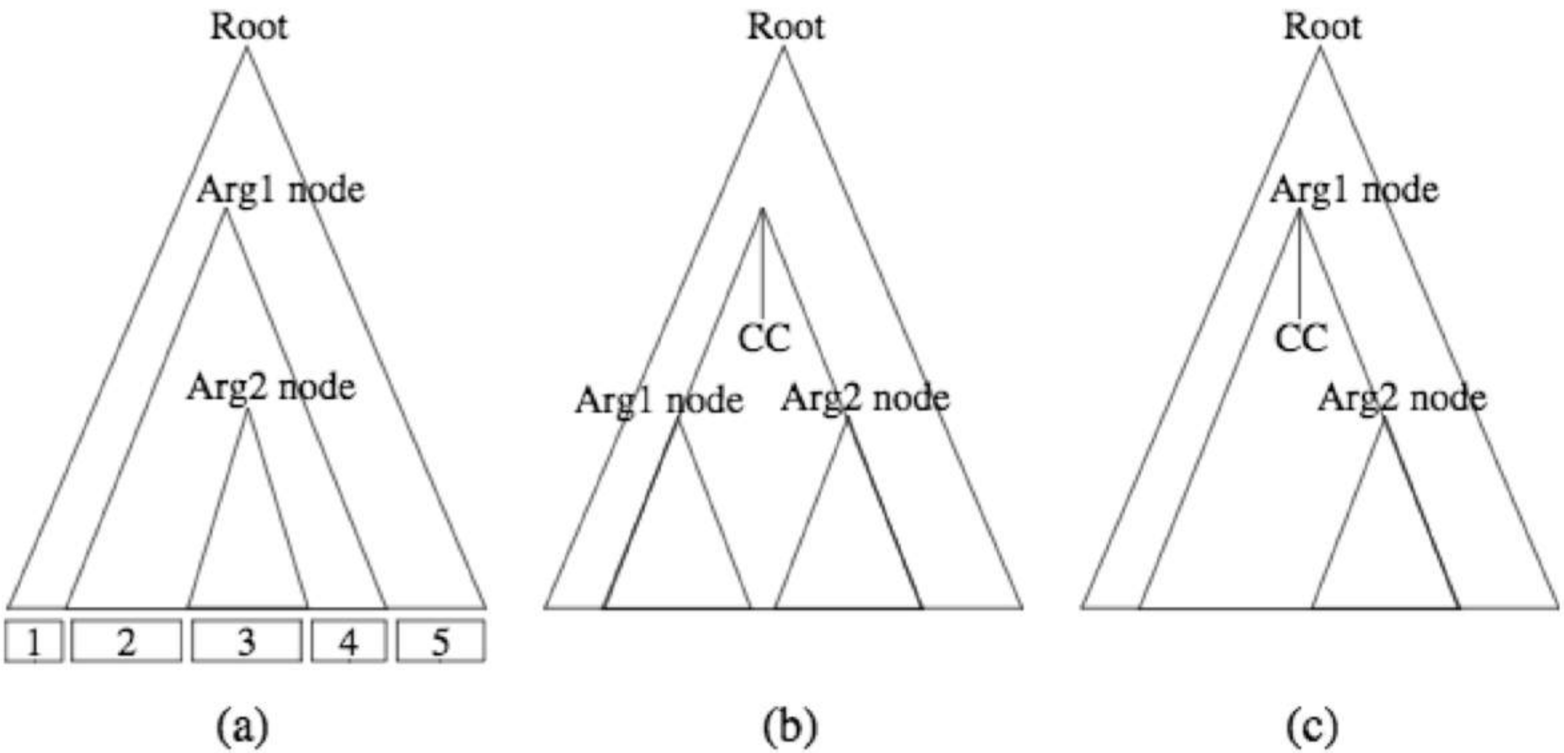
```
// Argument position classifier
Classify the relative position of Arg1 as SS or PS

// Argument extractor
if the relative position of Arg1 is SS then
    Identify the Arg1 and Arg2 subtree nodes within the sentence parse tree
    Apply tree subtraction to extract the Arg1 and Arg2 spans
else // the relative position of Arg1 is PS
    Label the sentence containing C as Arg2
    Identify and label the Arg1 sentence from all previous sentences of Arg2
end if
```

3. argument extractor (results)

- split sentence into clauses
- assign probabilities to each node
- subtract subtrees from potential arguments
- component F_1 = 86.24% for partial matches
- component F_1 = 53.85% for exact matches

relation of Arg1 and Arg2 (*SS*)



example: nesting arguments

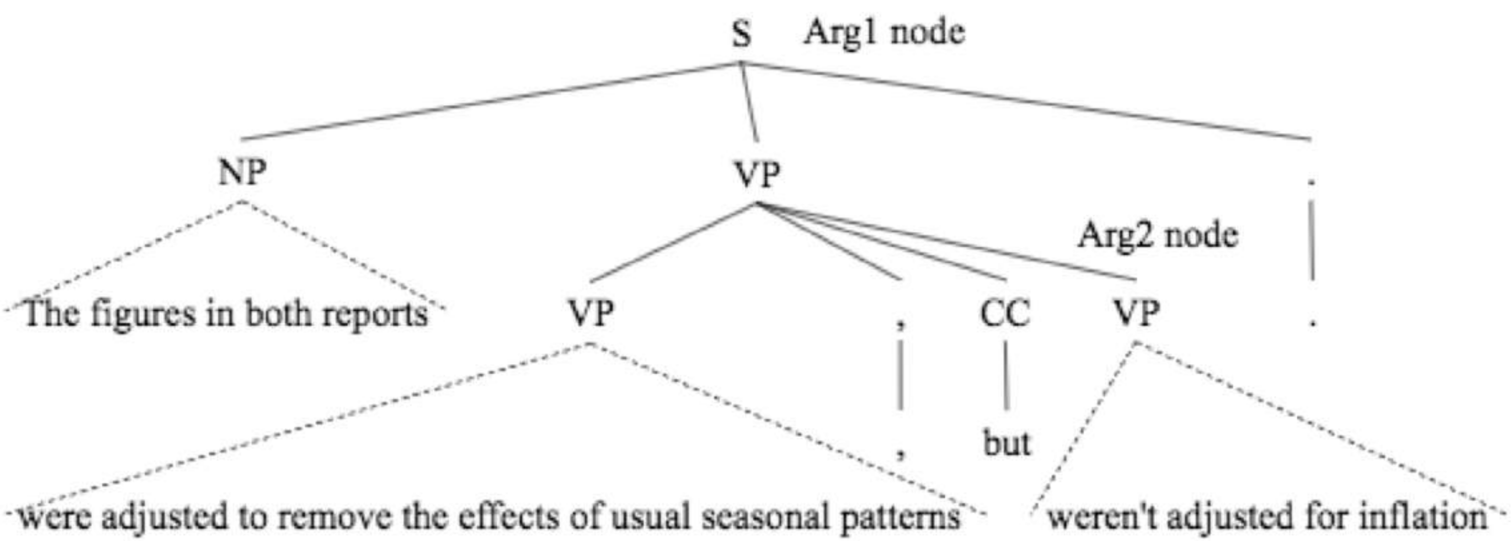


Fig. 7. The parse tree for Example 14 to illustrate Figure 6(c).

4. explicit (sense) classifier

- assign sense tag
- features: (only!)
  - connective C
  - its POS tag
  - C + previous word
- component F_1 = 86.77%

5. non-explicit classifier

- classify all remaining adjacent sentence pairs as Implicit/AltLex, EntRel, NoRel
- one classifier for 11 sense types + EntRel, NoRel
- separately determine AltLex using first three words of Arg2
- features:
  - contextual
  - dependency/constituent parses
  - word pairs
- component F_1 = 39.63%

6. attribution span labeler

- split text into clauses and determine which clauses are attribution spans
- mostly lexical features
- component F_1 = 79.68% for partial matches
- component F_1 = 65.95% for exact matches

results shallow discourse parsing

	Partial match F_1	Exact match F_1
GS + EP	46.80%	33.00%
Auto + EP	38.18%	20.64%

OPT pipeline system

OPT: Oslo–Potsdam–Teesside
Pipelining Rules, Rankers, and Classifier Ensembles for Shallow Discourse Parsing

Stephan Oepen[1], Jonathon Read[2], **Tatjana Scheffler**[3], Uladzimir Sidarenka[3,4], Manfred Stede[3], Erik Velldal[1], and Lilja Øvrelid[1]

[1]University of Oslo, Department of Informatics
[2]Teesside University, School of Computing
[3]University of Potsdam, FSP Cognitive Science
[4]Retresco GmbH

tatjana.scheffler@uni-potsdam.de

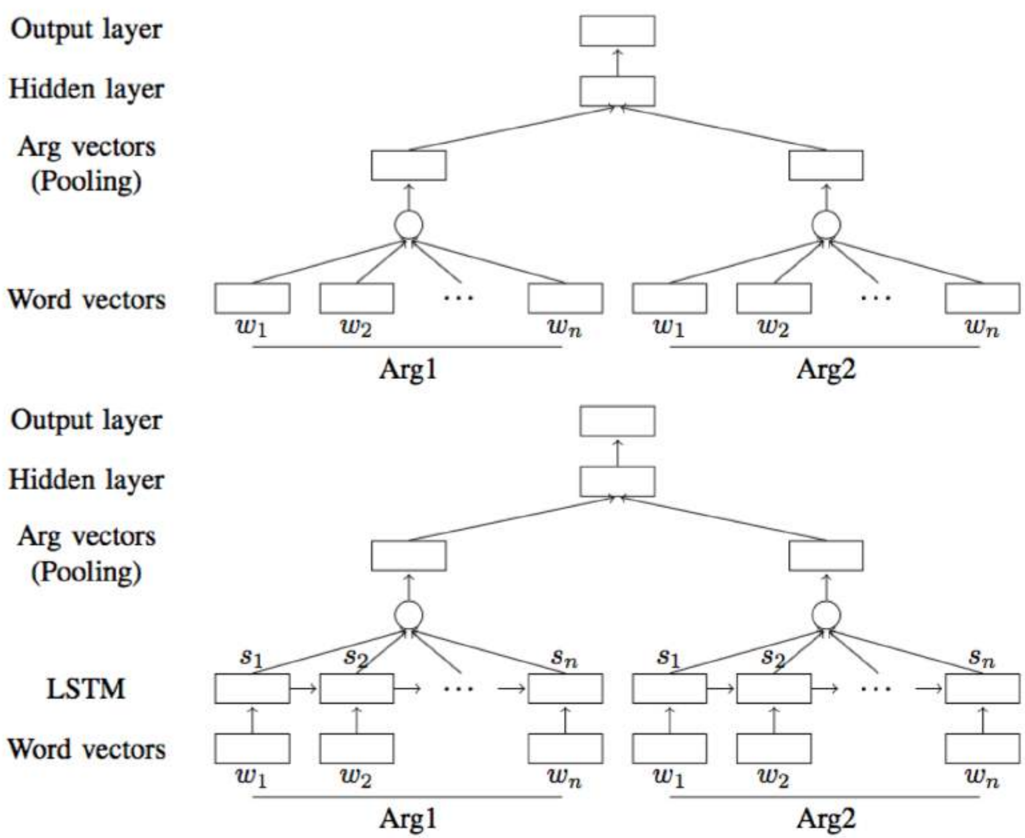
August 12, 2016

Neural models

- Dominant mainly for implicit sense classification
- End-to-end systems also available
- Performance gains only moderate/incremental

Rutherford/Demberg/Xue 2017

- Compared neural models
- Simple feed forward model best
- 1 hidden layer, sum pooling
- 300 pretrained (word2vec) features



Rutherford/Demberg/Xue 2017

Model	Acc.
CoNLL-ST 2015-2016 English (WSJ Test set)	
Most frequent tag baseline	21.36
Our best LSTM variant	31.76
Wang and Lan (2015) - winning team	34.45
Our best feedforward variant	**36.13**
CoNLL-ST 2016 Chinese (CTB Test set)	
Most frequent tag baseline	77.14
ME + Production rules	80.81
ME + Dependency rules	82.34
ME + Brown pairs (1000 clusters)	82.36
Out best LSTM variant	82.48
ME + Brown pairs (3200 clusters)	82.98
ME + Word pairs	83.13
ME + All feature sets	84.16
Our best feedforward variant	**85.45**

Parser adaptation

Lin parser implementation

- Rene Knaebel (Potsdam PhD student)
- https://github.com/rknaebel/discopy
- basic pipeline architecture
- classifiers for each module
- runs on CONLL-style JSON files
- needs text + parse input

Adapting the parser

- Features for explicit sense disambiguation:
  - Conn, ConnHead, ConnPOS, ConnPrev, ConnPosition
  - explicit.py
- Features for non-explicit sense disambiguation:
  - production rules, dependencies, word pairs
  - nonexplicit.py
- Add features?
  - negation
  - modals

Thank you!

Questions?

Contact me at:

tatjana.scheffler@uni-potsdam.de

@tschfflr