CS447: Natural Language Processing

http://courses.engr.illinois.edu/cs447

Lecture 24: Discourse Coherence

Julia Hockenmaier

juliahmr@illinois.edu

3324 Siebel Center


What makes
discourse
coherent?

Discourse: going beyond single sentences

> On Monday, John went to Einstein's. He wanted to buy lunch. But the cafe was closed. That made him angry, so the next day he went to Green Street instead.

'Discourse':

Any linguistic unit that consists of **multiple sentences**

Speakers describe "some situation or state of the real or some hypothetical world" (Webber, 1983)

Speakers attempt to get the **listener** to construct a similar **model of the situation**.

Local vs. global coherence

Local coherence:

There is coherence between **adjacent sentences:**

- **topical** coherence
- **entity-based** coherence
- **rhetorical** coherence

Global coherence:

The **overall structure** of a discourse is coherent (in ways that depend on the genre of the discourse):

Compare the structure of stories, persuasive arguments, scientific papers.

Topical coherence

Before winter I **built** a **chimney**, and **shingled**
the **sides** of my **house**...
I have thus a tight **shingled** and **plastered house**...
with a **garret** and a **closet**, a large **window** on each **side**....

These sentences clearly talk about the same topic: both contain a lot of words having to do with the structures of houses and building (they belong to the same '**semantic field**').

When nearby sentences talk about the same topic, they often exhibit **lexical cohesion** (they use the same or semantically related words).

Rhetorical coherence

John took a train from Paris to Istanbul.
He likes spinach.

This discourse is incoherent because there is no apparent rhetorical relation between the two sentences.

(Did you try to construct some explanation, perhaps that Istanbul has exceptionally good spinach, making the very long train ride worthwhile?)

Jane took a train from Paris to Istanbul.
She had to attend a conference.

This discourse is coherent because there is clear rhetorical relation between the two sentences.

The second sentence provides a REASON or EXPLANATION for the first.

Entity-based coherence

John wanted to buy a piano for his living room.
Jenny also wanted to buy a piano.
He went to the piano store.
It was nearby.
The living room was on the second floor.
She didn't find anything she liked.
The piano he bought was hard to get up to that floor.

This is incoherent because the sentences switch back and forth between **entities** (John, Jenny, the piano, the store, the living room)

Entity-based coherence

Entity-based coherence

Discourse 1:

John went to his favorite music store to buy a piano.

It was a store John had frequented for many years.

He was excited that he could finally buy a piano.

It was closing just as John arrived.

Discourse 2:

John went to his favorite music store to buy a piano.

He had frequented the store for many years.

He was excited that he could finally buy a piano.

He arrived just as the store was closing for the day.

Entity-based coherence

Discourse 1:

John went to **his favorite music store** to buy **a piano**.

It was **a store John** had frequented for many years.

He was excited that **he** could finally buy **a piano**.

It was closing just as **John** arrived.

Discourse 2:

John went to **his favorite music store** to buy **a piano**.

He had frequented **the store** for many years.

He was excited that **he** could finally buy **a piano**.

He arrived just as **the store** was closing for the day.

How we refer to entities in adjacent sentences influences how coherent a discourse is (**Centering theory)**

Centering Theory

Grosz, Joshi, Weinstein (1986, 1995)

A linguistic theory of entity-based coherence and salience

It predicts which entities are salient at any point during a discourse.

It also predicts whether a discourse is entity-coherent, based on its referring expressions.

Centering is about **local (=within a discourse segment) coherence and salience**

Centering theory itself is **not a computational model** or an algorithm: many of its assumptions are not precise enough to be implemented directly. (Poesio et al. 2004)

But many algorithms have been developed based on specific instantiations of the assumptions that Centering theory makes. The textbook presents a centering-based pronoun-resolution algorithm

Centering Theory: Definitions

Utterance:

A sequence of words (typically a sentence or clause) at a particular point in a discourse.

The centers of an utterance:

Entities that are evoked (mentioned) in the utterance and that link the utterance to the previous and following utterances.

Centering Theory: Assumptions

In each utterance, some discourse entities are more **salient** than others.

Salient entities

Discourse is more coherent if each utterance

Centering Theory: Assumptions

In each utterance, some discourse entities are more **salient** than others.

Salience in the current utterance determines how easy it is to refer to the same entity in the next utterance

This affects…

… how pronouns are used

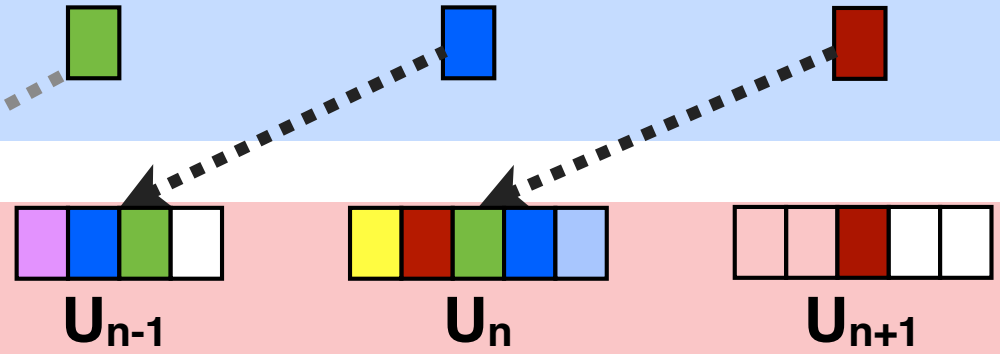
… whether a discourse is coherent or not

The two centers of an utterance

The **backward-looking center** of utterance $\mathbf{U_n}$ is the **highest ranked entity in $\mathbf{U_{n-1}}$** (the forward looking center of the previous utterance) **that is mentioned in $\mathbf{U_n}$**.

Backward-looking:
Mentioned in U_n and U_{n-1}

Forward-looking:
mentioned in U_n



The **forward-looking center** of utterance $\mathbf{U_n}$
is a **partially ordered list of the entities** mentioned in $\mathbf{U_n}$.

The ordering reflects **salience** within $\mathbf{U_n}$:

subject > direct object > object,….

Center realization and pronouns

Observation: Only the most salient entities of $\mathbf{U_{n-1}}$ can be referred to by pronouns in $\mathbf{U_n}$.

Constraint/Rule 1:

If ***any* element of $\mathbf{FW(U_{n-1})}$** is realized as **a pronoun in $\mathbf{U_n}$**, then the **$\mathbf{BW(U_n)}$** has to be realized **as a pronoun in $\mathbf{U_n}$** as well.

***Sue** told **Joe** to feed **her dog**.*
$\mathbf{BW(U_{n-1})}$=Sue, **$\mathbf{FW_{n-1}}$**={Sue, Joe, dog}

***He** asked **her** what to feed **it**.* **$\mathbf{BW(U_n)}$**=Sue, **$\mathbf{FW(U_n)}$**={Joe, Sue, dog} ✔ Constraint obeyed	***He** asked **Sue** what to feed **it**.* **$\mathbf{BW(U_n)}$**=Sue, **$\mathbf{FW(U_n)}$**={Joe, Sue, dog} ✘ Constraint violated: *Sue* should be a pronoun as well.

Transitions between sentences

Center continuation: BW stays the same, remains most salient

$BW(U_n) = BW(U_{n-1})$. $BW(U_n)$ is highest ranked element in $FW(U_n)$

U_{n-2} ***Sue*** *gave* ***Joe*** ***a dog***.
U_{n-1} ***She*** *told* ***him*** *to feed* ***it*** *well.* **BW**=Sue, **FW**={Sue, Joe, dog}
U_n ***She*** *asked* ***him*** *whether he liked* ***the gift.*** **BW**=Sue, **FW**={Sue, Joe, gift}

Center retaining: BW stays the same, but no longer most salient

$BW(U_n) = BW(U_{n-1})$. $BW(U_n) \neq$ highest ranked element in $FW(U_n)$

U_{n-2} ***Sue*** *gave* ***Joe*** ***a dog***.
U_{n-1} ***She*** *told* ***him*** *to feed* ***it*** *well.* **BW**=Sue, **FW**={Sue, Joe, dog}
U_n ***John*** *asked* ***her*** *what to feed* ***it***. **BW**=Sue, **FW**={Joe, Sue, dog}

Center shifting: BW changes

$BW(S_n) \neq BW(S_{n-1})$

U_{n-2} ***Sue*** *gave* ***Joe*** ***a dog***.
U_{n-1} ***She*** *told* ***him*** *to feed* ***it*** *well.* **BW**=Sue, **FW**={Sue, Joe, dog}
U_n ***The dog*** *was very cute.* **BW**=dog, **FW**={dog}

Local coherence: Preferred Transitions

Rule/Constraint 2:
Center continuation is preferred over center retaining.
Center retaining is preferred over center shifting.

Local coherence is achieved by **maximizing** the number of **center continuations.**

Example: Coherent discourse

John went to **his** **favorite music store** to buy **a piano**.

backward-looking center: ? (no previous discourse)

forward-looking center: {*John'*, *store'*, *piano'* }

He had frequented **the store** for many years.

backward-looking center: {*John'* }

forward-looking center: {*John'*, *store'* }

He was excited that **he** could finally buy **a piano**.

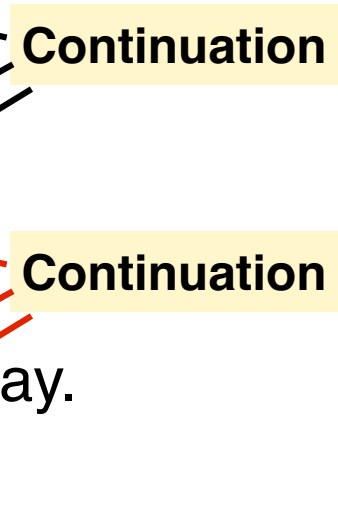
backward-looking center: {*John'* }

forward-looking center: {*John'*, *piano'* }

He arrived just as **the store** was closing for the day.

backward-looking center: {John' }

forward-looking center: {*John'*, *store'* }



Example: incoherent discourse

John went to **his favorite music store** to buy **a piano**.

backward-looking center: ? (no previous discourse)

forward-looking center: {*John'*, *store'*, *piano'* }

It was **a store John** had frequented for many years.

backward-looking center: {*John'* }

forward-looking center: {*store'*, *John'* }

He was excited that **he** could finally buy **a piano**.

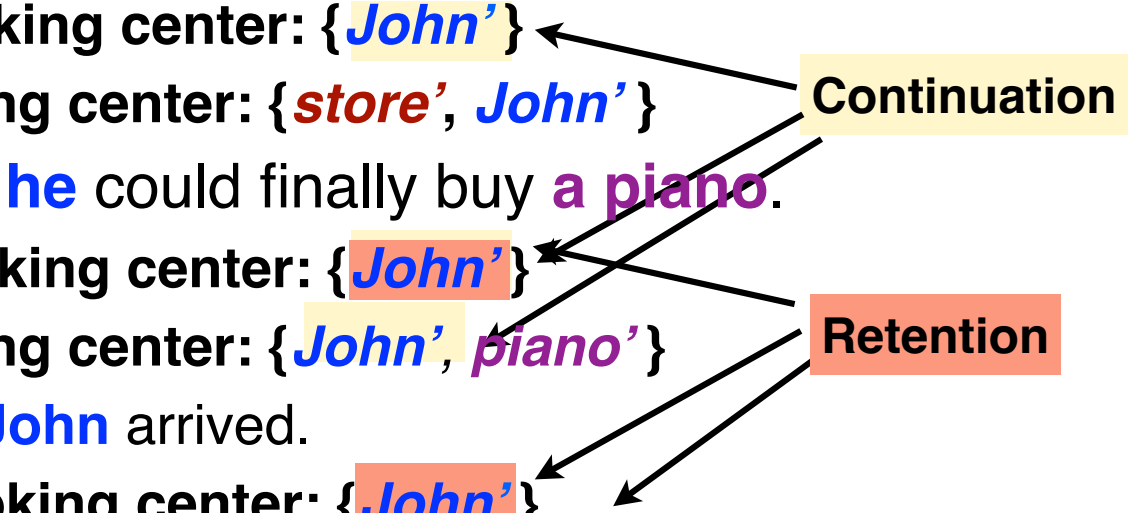
backward-looking center: {*John'* }

forward-looking center: {*John'*, *piano'* }

It was closing just as **John** arrived.

backward-looking center: {*John'* }

forward-looking center: {*store'*, *John'* }



Entity grid (Barzilay & Lapata '08)

1 [The Justice Department]$_S$ is conducting an [anti-trust trial]$_O$ against [Microsoft Corp.]$_X$ with [evidence]$_X$ that [the company]$_S$ is increasingly attempting to crush [competitors]$_O$.
2 [Microsoft]$_O$ is accused of trying to forcefully buy into [markets]$_X$ where [its own products]$_S$ are not competitive enough to unseat [established brands]$_O$.
3 [The case]$_S$ revolves around [evidence]$_O$ of [Microsoft]$_S$ aggressively pressuring [Netscape]$_O$ into merging [browser software]$_O$.
4 [Microsoft]$_S$ claims [its tactics]$_S$ are commonplace and good economically.
5 [The government]$_S$ may file [a civil suit]$_O$ ruling that [conspiracy]$_S$ to curb [competition]$_O$ through [collusion]$_X$ is [a violation of the Sherman Act]$_O$.
6 [Microsoft]$_S$ continues to show [increased earnings]$_O$ despite [the trial]$_X$.

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	S	O	S	X	O	–	–	–	–	–	–	–	–	–	–	1
2	–	–	O	–	–	X	S	O	–	–	–	–	–	–	–	2
3	–	–	S	O	–	–	–	–	S	O	O	–	–	–	–	3
4	–	–	S	–	–	–	–	–	–	–	–	S	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	S	O	–	5
6	–	X	S	–	–	–	–	–	–	–	–	–	–	–	O	6

Entity grid (Barzilay & Lapata '08)

Entity grid:

rows = sentences

columns = entities in a document

entries = Subject (S), Object (O), other (X), none (–)

Transitions:

How an entity's entry changes from utterance to utterance

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	S	O	S	X	O	–	–	–	–	–	–	–	–	–	–	1
2	–	–	O	–	–	X	S	O	–	–	–	–	–	–	–	2
3	–	–	S	O	–	–	–	–	S	O	O	–	–	–	–	3
4	–	–	S	–	–	–	–	–	–	–	–	S	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	S	O	–	5
6	–	X	S	–	–	–	–	–	–	–	–	–	–	–	O	6

Entity grid (Barzilay & Lapata '08)

Entity grid:

rows = sentences

columns = entities in a document

entries = Subject (S), Object (O), other (X), none (–)

Transitions:

How an entity's entry changes from utterance to utterance

Aim: Learn a classifier that predicts whether a text is coherent or not based on the transitions in its entity grid

(Negative examples = randomly shuffled sentences)


Rhetorical
(Discourse)
relations

Rhetorical relations

Discourse 1:
John hid Bill's car keys. He was drunk.

Discourse 2:
John hid Bill's car keys. He likes spinach.

Discourse 1 is more coherent than Discourse 2 because "He(=Bill) was drunk" provides an **explanation** for "John hid Bill's car keys"

What **kind of relations** between two consecutive utterances (=sentences, clauses, paragraphs,…) make a discourse coherent?

Rhetorical Structure Theory; also lots of recent work on **discourse parsing (Penn Discourse Treebank)**

Example: The *Result* relation

The reader can infer that the **state/event described in S0 causes** (or: could cause) **the state/event asserted in S1**:

S0: The Tin Woodman was caught in the rain.

S1: His joints rusted.

This can be rephrased as:
"S0. **As a result,** S1"

Example: The *Explanation* relation

The reader can infer that **the state/event in S1 provides an explanation** (reason) **for the state/event in S0**:

S0: John hid Bill's car keys.

S1: He was drunk.

This can be rephrased as:
"S0 **because** S1"

Rhetorical Structure Theory (RST)

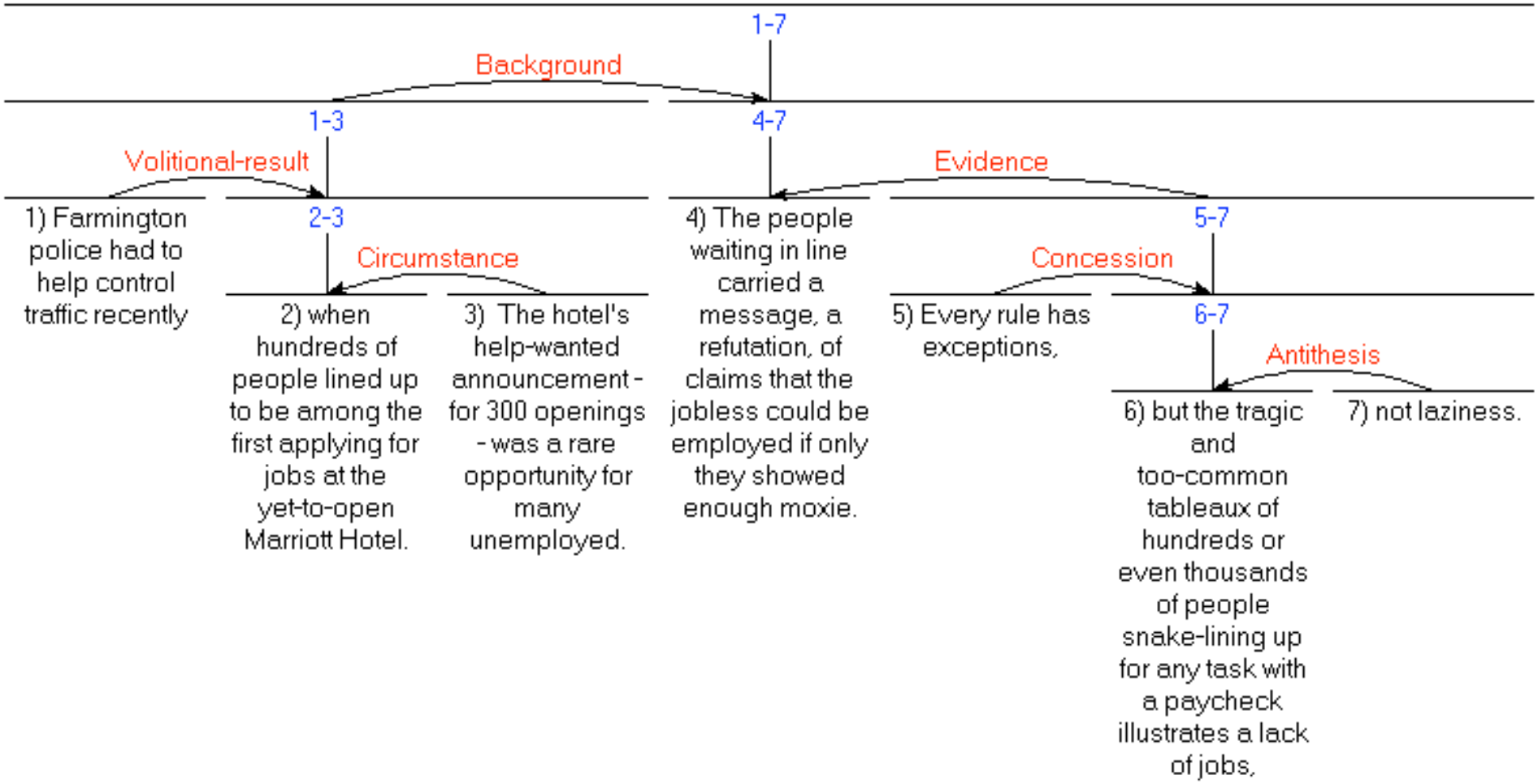
RST (Mann & Thompson, 1987) describes **rhetorical relations** between utterances:
Evidence, Elaboration, Attribution, Contrast, List,…

Different variants of RST assume different sets of relations.

Most relations hold between a **nucleus** (N) and a **satellite** (S).
Some relations (e.g. *List*) have **multiple nuclei** (and no satellite).

Every relation imposes certain **constraints** on its arguments (N,S), that describe the goals and beliefs of the **reader** R and **writer** W, and the effect of the utterance on the reader.

Discourse structure is hierarchical

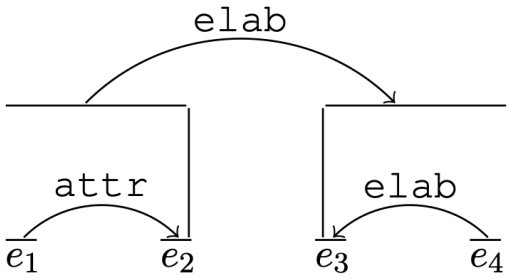


RST website: http://www.sfu.ca/rst/

RST parsing (Yu et al. 2018)

Step 1: Identify "**elementary discourse units**" (EDUs, text spans)

Step 2: Use a **shift-reduce parser** to predict RST relations between EDUs



e_1: American Telephone & Telegraph Co. said it
e_2: will lay off 75 to 85 technicians here , effective Nov. 1.
e_3: The workers install , maintain and repair its private branch exchanges,
e_4: which are large intracompany telephone networks.

Figure 27.5 Example RST discourse tree, showing four EDUs. Figure from Yu et al. (2018).

RST parsing (Yu et al. 2018)

Step 1: Identify "**elementary discourse units**" (EDUs, text spans)

Step 2: Use a **shift-reduce parser** to predict RST relations between EDUs

Step	Stack	Queue	Action	Relation
1	$\varnothing$	e_1, e_2, e_3, e_4	SH	$\varnothing$
2	e_1	e_2, e_3, e_4	SH	$\varnothing$
3	e_1, e_2	e_3, e_4	RD(attr, SN)	$\varnothing$
4	$e_{1:2}$	e_3, e_4	SH	$\widehat{e_1 \mathbf{e_2}}$
5	$e_{1:2}, e_3$	e_4	SH	$\widehat{e_1 \mathbf{e_2}}$
6	$e_{1:2}, e_3, e_4$	$\varnothing$	RD(elab, NS)	$\widehat{e_1 \mathbf{e_2}}$
7	$e_{1:2}, e_{3:4}$	$\varnothing$	RD(elab, SN)	$\widehat{e_1 \mathbf{e_2}}, \widehat{\mathbf{e_3} e_4}$
8	$e_{1:4}$	$\varnothing$	PR	$\widehat{e_1 \mathbf{e_2}}, \widehat{\mathbf{e_3} e_4}, \widehat{e_{1:2} \mathbf{e_{3:4}}}$

Figure 27.6 Parsing the example of Fig. 27.5 using a shift-reduce parser. Figure from Yu et al. (2018).

Penn Discourse Treebank (PDTB)

Miltsakaki et al. 2004, Prasad et al. 2008, 2014

The PDTB annotates explicit and implicit discourse connectives and their argument spans.

Explicit connective ("as a result")

[arg1 *Jewelry displays in department stores were often cluttered and uninspired. And the merchandise was, well, fake*].
As a result, [arg2 marketers of faux gems steadily lost space in department stores to more fashionable rivals—cosmetics makers]

Implicit connective (no lexical item)

[arg1 *In July, the Environmental Protection Agency imposed a gradual ban on virtually all uses of asbestos.*]
[arg2 By 1997, almost all remaining uses of cancer-causing asbestos will be outlawed]

PDTB semantic distinctions

Class	Type	Example
TEMPORAL	SYNCHRONOUS	The parishioners of St. Michael and All Angels stop to chat at the church door, as members here always have. (Implicit while) In the tower, five men and women pull rhythmically on ropes attached to the same five bells that first sounded here in 1614.
CONTINGENCY	REASON	Also unlike Mr. Ruder, Mr. Breeden appears to be in a position to get somewhere with his agenda. (implicit=because) **As a former White House aide who worked closely with Congress, he is savvy in the ways of Washington.**
COMPARISON	CONTRAST	The U.S. wants the removal of what it perceives as barriers to investment; Japan denies there are real barriers.
EXPANSION	CONJUNCTION	Not only do the actors stand outside their characters and make it clear they are at odds with them, but they often literally stand on their heads.

PDTB sense hierarchy

Temporal • Asynchronous • Synchronous (Precedence, Succession)	**Comparison** • Contrast (Juxtaposition, Opposition) •*Pragmatic Contrast (Juxtaposition, Opposition)* • Concession (Expectation, Contra-expectation) • *Pragmatic Concession*
Contingency • Cause (Reason, Result) • Pragmatic Cause (Justification) • *Condition (Hypothetical, General, Unreal Present/Past, Factual Present/Past)* • *Pragmatic Condition (Relevance, Implicit Assertion)*	**Expansion** • *Exception* • Instantiation • Restatement (Specification, Equivalence, Generalization) • Alternative (Conjunction, Disjunction, Chosen Alternative) • List

Figure 23.3 The PDTB sense hierarchy. There are four top-level classes, 16 types, and 23 subtypes (not all types have subtypes). 11 of the 16 types are commonly used for implicit argument classification; the 5 types in italics are too rare in implicit labeling to be used.


Global
coherence

Global coherence: Argumentation structure

In persuasive essays, **claims** (1) may be followed (or preceded) by **premises** (2,3) that support the claim, (some of which might be supported by their own premises (4) (Stab and Gurevych, 2014)

(1) Museums and art galleries provide a better understanding about arts than Internet. (2) In most museums and art galleries, detailed descriptions in terms of the background, history and author are provided. (3) Seeing an artwork online is not the same as watching it with our own eyes, as (4) the picture online does not show the texture or three-dimensional structure of the art, which is important to study."

Argumentation mining

Can we automatically detect claims
and the premises that are made to support them?

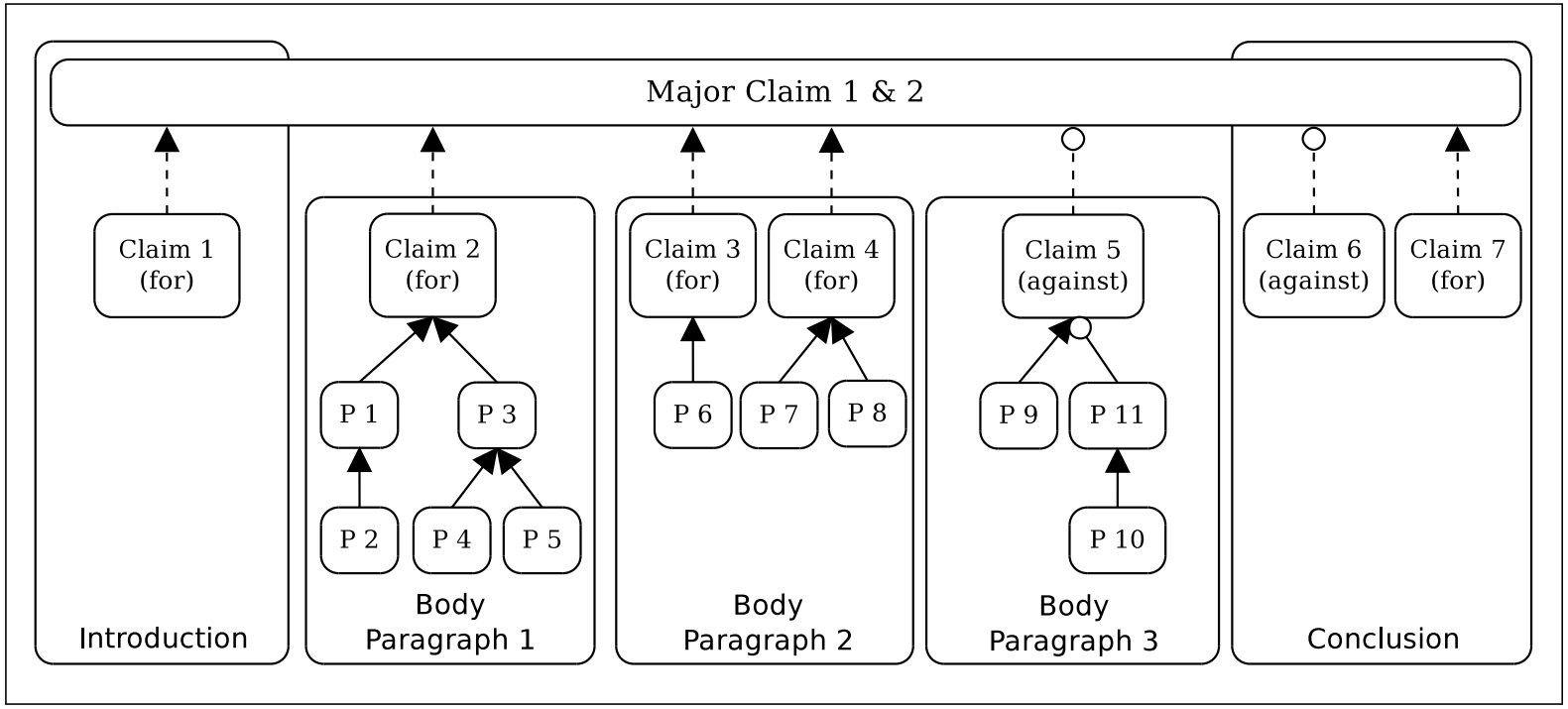


Figure 23.12 Argumentation structure of a persuasive essay. Arrows indicate argumentation relations, either of SUPPORT (with arrowheads) or ATTACK (with circleheads); P denotes premises. Figure from Stab and Gurevych (2017).

The structure of scientific discourse

Category	Description	Example
AIM	Statement of specific research goal, or hypothesis of current paper	"The aim of this process is to examine the role that training plays in the tagging process"
OWN_METHOD	New Knowledge claim, own work: methods	"In order for it to be useful for our purposes, the following extensions must be made:"
OWN_RESULTS	Measurable/objective outcome of own work	"All the curves have a generally upward trend but always lie far below backoff (51% error rate)"
USE	Other work is used in own work	"We use the framework for the allocation and transfer of control of Whittaker...."
GAP_WEAK	Lack of solution in field, problem with other solutions	"Here, we will produce experimental evidence suggesting that this simple model leads to serious overestimates"
SUPPORT	Other work supports current work or is supported by current work	"Work similar to that described here has been carried out by Merialdo (1994), with broadly similar conclusions."
ANTISUPPORT	Clash with other's results or theory; superiority of own work	"This result challenges the claims of..."

Figure 23.13 Examples for 7 of the 15 labels from the Argumentative Zoning labelset (Teufel et al., 2009).

We can also label spans in scientific papers with the role they play in the overall argumentation of the paper.