School of CIT –
Social Computing
Research Group

Technical
University
of Munich

Natural Language Processing
IN2361

Prof. Dr. Georg Groh

Chapter 27
Discourse Coherence

- content is based on [1]
- certain elements (e.g. equations or tables) were taken over or taken over in a modified form from [1]
- citations of [1] or from [1] are omitted for legibility
- errors are fully in the responsibility of Georg Groh
- BIG thanks to Dan and James for a great book!

Discourse Coherence

- local coherence:
  - John took a train from Paris to Istanbul. He likes spinach.

    incoherent

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

    coherent. coherence relation: *REASON*
  - coherence: focus on salient entities → 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.

    ←→ Centering Theory, Entity Grid theory
  - topical coherence → lexical cohesion

    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....

Discourse Coherence

- global coherence: ←→ particular conventional discourse structures:
  scientific paper, persuasive essay, story, news article etc.

- coherence detection / quantitative measures for coherence: e.g.
  - for text quality assessment
  - for summarization
  - for medical applications (e.g. mental disease detection)

Coherence Relations: Rhetorical Structure Theory

relations between phrases / spans: nucleus and satellite

- Reason

  [$_{NUC}$ Jane took a train from Paris to Istanbul.] [$_{SAT}$ She had to attend a conference.]

- Elaboration

  [$_{NUC}$ Dorothy was from Kansas.] [$_{SAT}$ She lived in the midst of the great Kansas prairies.]

- Evidence

  [$_{NUC}$ Kevin must be here.] [$_{SAT}$ His car is parked outside.]

- Attribution

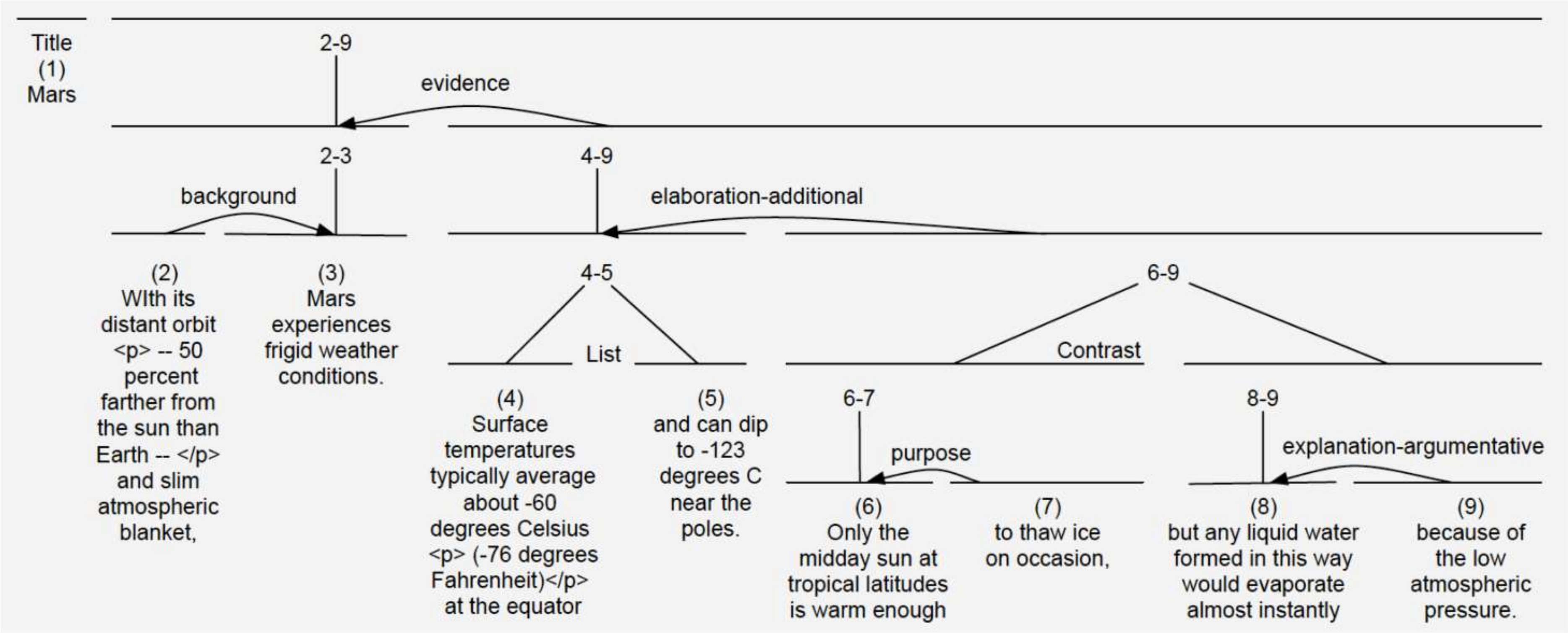
  [$_{SAT}$ Analysts estimated] [$_{NUC}$ that sales at U.S. stores declined in the quarter, too]

- List

  [$_{NUC}$ Billy Bones was the mate;] [$_{NUC}$ Long John, he was quartermaster]

Coherence Relations: Rhetorical Structure Theory (RST)

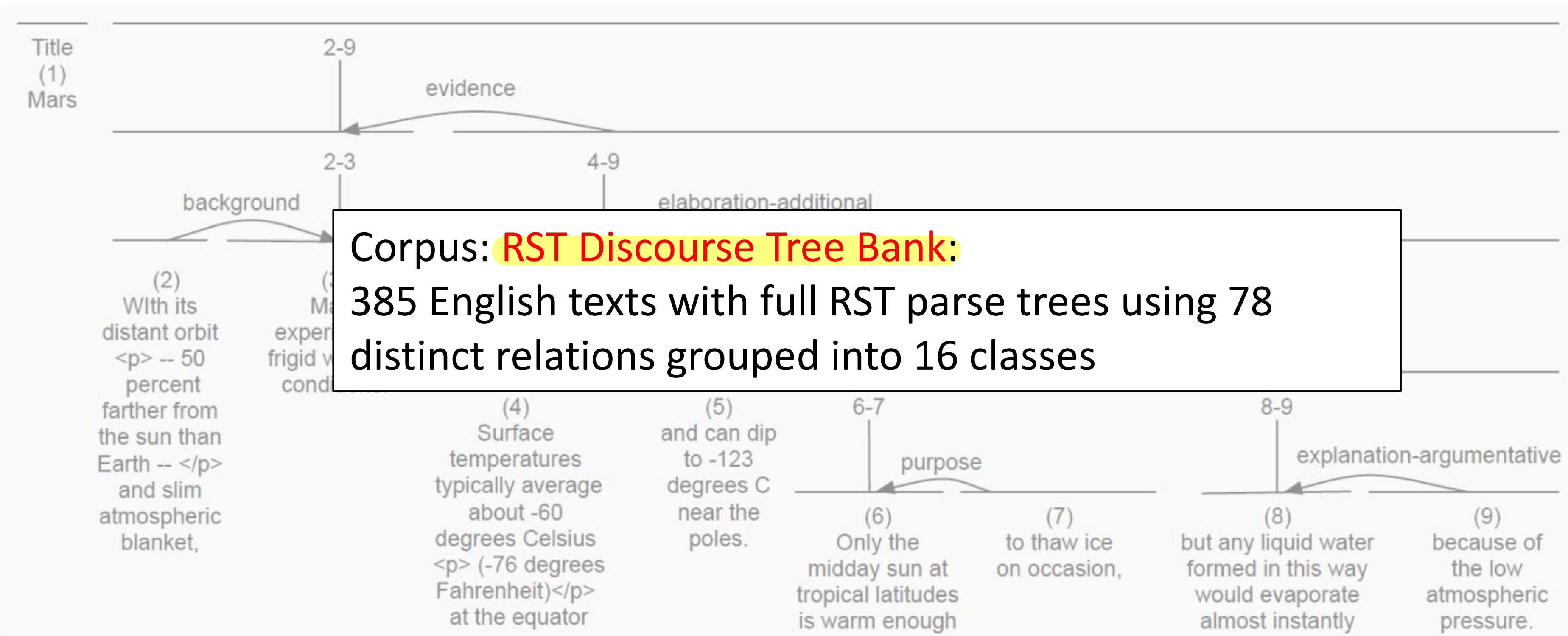
With its distant orbit–50 percent farther from the sun than Earth–and slim atmospheric blanket, Mars experiences frigid weather conditions. Surface temperatures typically average about -60 degrees Celsius (-76 degrees Fahrenheit) at the equator and can dip to -123 degrees C near the poles. Only the midday 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.



RST parse tree: nodes: elementary discourse units / discourse segments

Coherence Relations: Rhetorical Structure Theory (RST)

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.



Corpus: RST Discourse Tree Bank:
385 English texts with full RST parse trees using 78 distinct relations grouped into 16 classes

RST parse tree: nodes: elementary discourse units / discourse segments

Coherence Relations: Alternative: Penn Discourse TreeBank

- lexical grounded annotation: based on lexical discourse connectives (e.g. *because, although, when, since,* or *as a result*)

Arg1

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

Arg2

[**Conn** 为] [**Arg2** 推动图们江地区开发]，[Arg1 韩国捐款一百万美元设立了图们江发展基金]
"[In order to] [**Arg2** promote the development of the Tumen River region], [**Arg1** South Korea donated one million dollars to establish the Tumen River Development Fund]."

- Also possible: relations between sentences without discourse connectives:

In July, the Environmental Protection Agency imposed a gradual ban on virtually all uses of asbestos. (implicit=as a result) **By 1997, almost all remaining uses of cancer-causing asbestos will be outlawed.**

Coherence Relations: Penn Discourse TreeBank

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.

Figure 22.2 The four high-level semantic distinctions in the 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

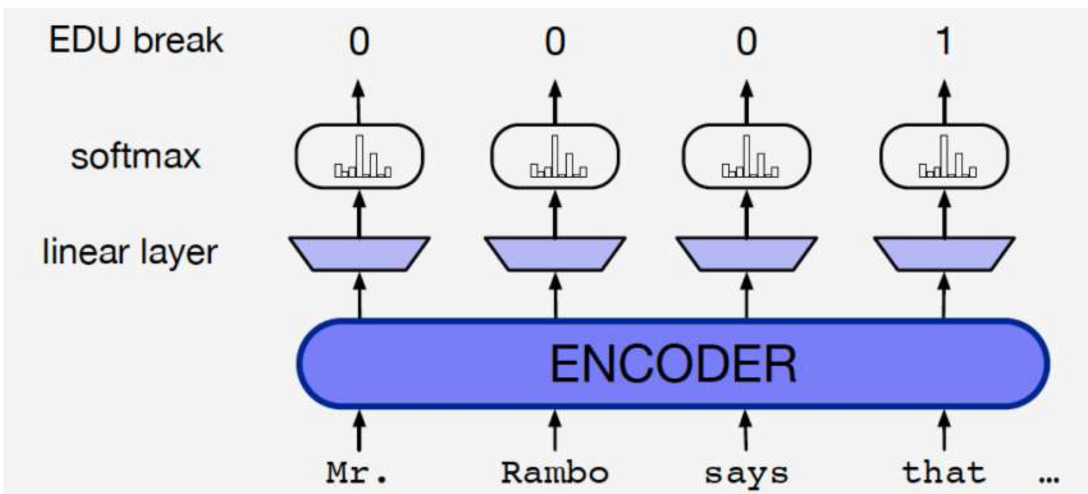
Figure 22.3 The PDTB sense hierarchy. There are four top-level classes, 16 types, and 23 subtypes (not all types have subtypes). 11 of the 16 types are commonly used for implicit argument classification; the 5 types in italics are too rare in implicit labeling to be used.

Discourse Structure Parsing: RST

- determine discourse relations btw. sentences: discourse parsing

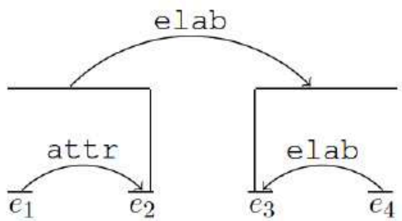
- Step 1: identifying (segmenting) elementary discourse units (RST-EDUs) / discourse segments

[Mr. Rambo says]$_{e1}$ [that a 3.2-acre property]$_{e2}$ [overlooking the San Fernando Valley]$_{e3}$ [is priced at \$4 million]$_{e4}$ [because the late actor Erroll Flynn once lived there.]$_{e5}$



Discourse Structure Parsing: RST

- Step 2: RST parsing (identify rhetorical structure theory (RST) relations): NN shift-reduce style architecture:



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

- shift: pushes first EDU in the queue onto stack creating a single-node subtree.
- reduce(l,d): merges the top two subtrees on the stack, where l is the coherence relation label, and d is the nuclearity direction, d ∈ {NN,NS,SN}.
- pop root: remove final tree from stack

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

Figure 22.6 Parsing the example of Fig. 22.5 using a shift-reduce parser. Figure from Yu et al. (2018).

nuc nuc, nuc sat, sat nuc

Discourse Structure Parsing: RST

- Yu et al 2018 [2] encoder-decoder architecture:

$w_1, w_2, \ldots, w_m$ input text word sequence

$\mathbf{x}_1^w, \mathbf{x}_2^w, \ldots, \mathbf{x}_m^w$ (static word-, static character-, or contextual embeddings

$$\mathbf{h}_1^w, \mathbf{h}_2^w, \ldots, \mathbf{h}_m^w = \text{biLSTM}(\mathbf{x}_1^w, \mathbf{x}_2^w, \ldots, \mathbf{x}_m^w)$$

EDU of span $w_s, w_{s+1}, \ldots, w_t$
has biLSTM output representation $\mathbf{h}_s^w, \mathbf{h}_{s+1}^w, \ldots, \mathbf{h}_t^w$

elementary discourse unit

$$\mathbf{x}^e = \frac{1}{t-s+1} \sum_{k=s}^{t} \mathbf{h}_k^w$$

EDU initial representation via average-pooling

$$\mathbf{h}_1^e, \mathbf{h}_2^e, \ldots, \mathbf{h}_n^e = \text{biLSTM}(\mathbf{x}_1^e, \mathbf{x}_2^e, \ldots, \mathbf{x}_n^e)$$

Sequence of n EDUs final representations

Discourse Structure Parsing: RST

- Yu et al 2018 [2] encoder-decoder architecture:

$$\mathbf{o} = \mathbf{W}(\mathbf{h}^{\mathrm{t}}_{\mathrm{s}0}, \mathbf{h}^{\mathrm{t}}_{\mathrm{s}1}, \mathbf{h}^{\mathrm{t}}_{\mathrm{s}2}, \mathbf{h}^{\mathrm{e}}_{\mathrm{q}0})$$

action o predicted by $\mathrm{FFN}_{\mathrm{W}}$

where

$$\left\{ \begin{array}{l} (s_o, s_1, s_2) \\ (q_0) \\ \mathbf{h}^{\mathrm{t}}_{\mathrm{S}} = \frac{1}{j-i+1} \sum_{k=i}^{j} \mathbf{h}^{e}_{k} \end{array} \right.$$

top 3 subtrees on stack

first EDU in queue

subtree representation: avg.-pooling over all EDUs in subtree

$$p_a = \frac{\exp(\mathbf{o}_a)}{\sum_{a' \in A} \exp(\mathbf{o}_{a'})}$$

softmax to get probabilities for actions

$$L_{CE}() = -\log(p_a) + \frac{\lambda}{2} ||\Theta||^2$$

regularized cross entropy loss using gold action for each step derived from gold parse

Centering and Entity-Based Coherence

- Centering:
  - discourse is "about" some entities
  - discourse salience: at any point in discourse one of the entities is salient (is being centered on)
  - discourse coherence: higher if adjacent sentences are about the same entity (Continue) vs. shifting back and forth btw. Entities (Shift)

a. John went to his favorite music store to buy a piano.
b. He had frequented the store for many years.
c. He was excited that he could finally buy a piano.
d. He arrived just as the store was closing for the day.

about John (Continue)

a. John went to his favorite music store to buy a piano.
b. It was a store John had frequented for many years.
c. He was excited that he could finally buy a piano.
d. It was closing just as John arrived.

about store and John (Shift)

Centering and Entity-Based Coherence

- For chain of utterances $\{U_n\}_{n=1}^{N}$: John, he, he, he
  - backward looking center $C_b(U_n)$: current salient entity
  - forward looking center $C_f(U_n)$: potential future salient entities (from which $C_b(U_{n+1})$ is "picked") ; $C_p(U_n)$: most probable ("preferred")

Rule 1: If any element of $C_f(U_n)$ is realized by a pronoun in utterance U_{n+1}, then $C_b(U_{n+1})$ must be realized as a pronoun also.
Rule 2: Transition states are ordered. Continue is preferred to Retain is preferred to Smooth-Shift is preferred to Rough-Shift.

	$C_b(U_{n+1}) = C_b(U_n)$ or undefined $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	Continue	Smooth-Shift
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	Retain	Rough-Shift

preference

Figure 22.7 Centering Transitions for Rule 2 from Brennan et al. (1987).

John went to his favorite music store to buy a piano. (U_1)
He was excited that he could finally buy a piano. (U_2)
He arrived just as the store was closing for the day. (U_3)
It was closing just as John arrived (U_4)

$C_f(U_1)$: {John, music store, piano}
$C_p(U_1)$: John
$C_b(U_1)$: undefined

$C_f(U_2)$: {John, piano}
$C_p(U_2)$: John
$C_b(U_2)$: John
Result: Continue ($C_p(U_2)$=$C_b(U_2)$; $C_b(U_1)$ undefined)

Entity Grid Model

- entity grid: represents the distribution of entity mentions across sentences
  - grammatical roles: subject (S), object (O), neither (X), absent (–)

	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

Figure 22.8 Part of the entity grid for the text in Fig. 22.9. Entities are listed by their head noun; each cell represents whether an entity appears as subject (S), object (O), neither (X), or is absent (–). Figure from Barzilay and Lapata (2008).

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]$_X$.

Figure 22.9 A discourse with the entities marked and annotated with grammatical functions. Figure from Barzilay and Lapata (2008).

Entity Grid Model

- entity grid: represents the distribution of entity mentions across sentences
  - grammatical roles: subject (S), object (O), neither (X), absent (–)

	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

Figure 22.8 Part of the entity grid for the text in Fig. 22.9. Entities are listed by their head noun; each cell represents whether an entity appears as subject (S), object (O), neither (X), or is absent (–). Figure from Barzilay and Lapata (2008).

	S S	S O	S X	S –	O S	O O	O X	O –	X S	X O	X X	X –	– S	– O	– X	– –
d_1	.01	.01	0	.08	.01	0	0	.09	0	0	0	.03	.05	.07	.03	.59
d_2	.02	.01	.01	.02	0	.07	0	.02	.14	.14	.06	.04	.03	.07	0.1	.36
d_3	.02	0	0	.03	.09	0	.09	.06	0	0	0	.05	.03	.07	.17	.39

Figure 22.10 A feature vector for representing documents using all transitions of length 2. Document d_1 is the text in Fig. 22.9. Figure from Barzilay and Lapata (2008).

Representation Learning for Local Coherence

- Third kind of local coherence: topical coherence
  - e.g. sentence embedding based:

$$\mathrm{sim}(s,t) = \cos(\mathbf{s},\mathbf{t})$$
$$\underset{\text{(simplest case)}}{=} \cos\Big(\sum_{w\in s}\mathbf{w}, \sum_{w\in t}\mathbf{w}\Big)$$
$$\mathrm{coherence}(T) = \frac{1}{n-1}\sum_{i=1}^{n-1}\cos(s_i, s_{i+1})$$

  - Use self supervision: distinguish natural (coherent) texts T, from unnatural (incoherent) versions by permutating sentences. Example: LCD model:

$$L_\theta = \sum_{d\in C}\sum_{s_i\in d}\mathbb{E}_{p(s'|s_i)}\Big[L\big(\underbrace{f_\theta(s_i,s_{i+1})}_{\text{maximized}}, \underbrace{f_\theta(s_i,s')}_{\text{minimized}}\big)\Big]$$

Documents d in collection C

sentence s_i in document d

Implemented via sampling negative sentence s' from each s_i

any sentence embedding allowed (e.g. BERT)

Global Coherence – Discourse Structure

- Argument Mining:

> “(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.”

Argument relations: SUPPORT(2,1), SUPPORT(3,1), SUPPORT(4,3)

Global Coherence – Discourse Structure

- Argument Mining:

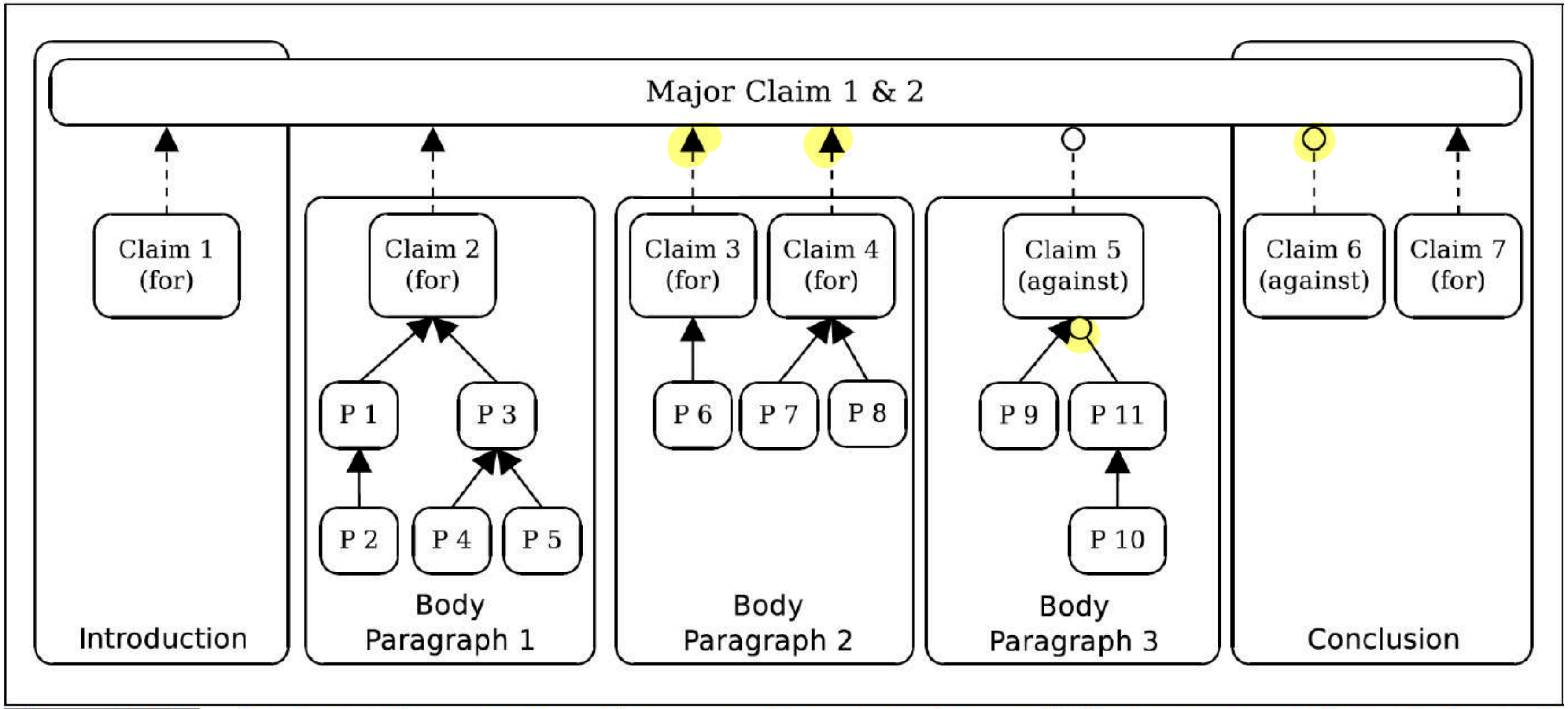


Figure 22.12 Argumentation structure of a persuasive essay. Arrows indicate argumentation relations, either of SUPPORT (with arrowheads) or ATTACK (with circleheads); P denotes premises. Figure from Stab and Gurevych (2017).

Global Coherence – Discourse Structure

- Argument Mining algorithms:
  - span-based classifiers for argument types (claim, premise, non-argument etc.)
  - span-based argument relation detection
  - or higher order argumentation scheme classifiers (e.g. argument from examples)

- Higher order approaches: predict persuasive success of an argumentation. feature-motives:
  - reciprocity (people return favors),
  - social proof (people follow others' choices),
  - authority (people are influenced by those with power),
  - scarcity (people value things that are scarce),

Global Coherence – Discourse Structure

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

Bibliography

(1) Dan Jurafsky and James Martin: Speech and Language Processing (3rd ed. draft) (Jan 2023); Online: https://web.stanford.edu/~jurafsky/slp3/ (URL, Oct 2023)

(2) Yu, N., M. Zhang, and G. Fu. 2018: Transition-based neural RST parsing with implicit syntax features. COLING 2018.

Recommendations for Studying

- minimal approach:

  work with the slides and understand their contents! Think beyond instead of merely memorizing the contents

- standard approach:

  minimal approach + read the corresponding pages in Jurafsky [1]

- interested students

  == standard approach