

An Integrated Architecture for Generating Parenthetical Constructions

Eva Banik

The Open University



The Open University

- Parenthetical constructions
- Corpus study on two discourse treebanks
- Results of corpus study formulated with a TAG
- An integrated generation architecture to generate parentheticals

What are parenthetical constructions?

- express less important information in the clause
- embedded: not part of the main predicate-argument structure

Some examples:

- APPOSITIVES AND OTHER NPs

The new goal of the Voting Rights Act [– more minorities in political office –] is laudable. (wsj1137)

What are parenthetical constructions?

- NON-RESTRICTIVE RELATIVE CLAUSES
GE, [which vehemently denies the government's allegations,] denounced Mr. Greenfield's suit. (wsj0617)
- TO-INFINITIVES
PandG's new powdered detergent [– to be called Cheer with Color Guard –] will be on shelves in that market by early November. (wsj2320)
- PARTICIPIAL CLAUSES
But most businesses in the Bay area, [including Silicon Valley,] weren't greatly affected. (wsj1930)

What are parenthetical constructions?

- SUBORDINATE CLAUSES WITH DISCOURSE CONNECTIVES
The show, [despite a promising start,] has slipped badly in the weekly ratings as compiled by A.C. Nielsen Co.[...] (wsj2395)
- FULL SENTENCES
The big questions [– Do you really need this much money to put up these investments? Have you told investors what is happening in your sector? What about your track record? –] aren't asked of companies coming to market. (wsj0629)

Why generate parentheticals?

- make texts easier to read
- allow reader to distinguish between more and less important information

Eprex is used by dialysis patients who are anemic. Prepulsid is a gastro-intestinal drug. Eprex and Prepulsid did well overseas.

Eprex, [used by dialysis patients who are anemic,] and Prepulsid, [a gastro-intestinal drug,] did well overseas. (wsj1156)

Why haven't parentheticals been generated before?

Commonly used input to an NLG system is Rhetorical Structure Tree (Mann & Thompson 87):



RST tree input to syntactic realizer; text spans concatenated:

[Surfing is fun.] [But surfing is dangerous.]
[Surfing is fun], [although it is dangerous].

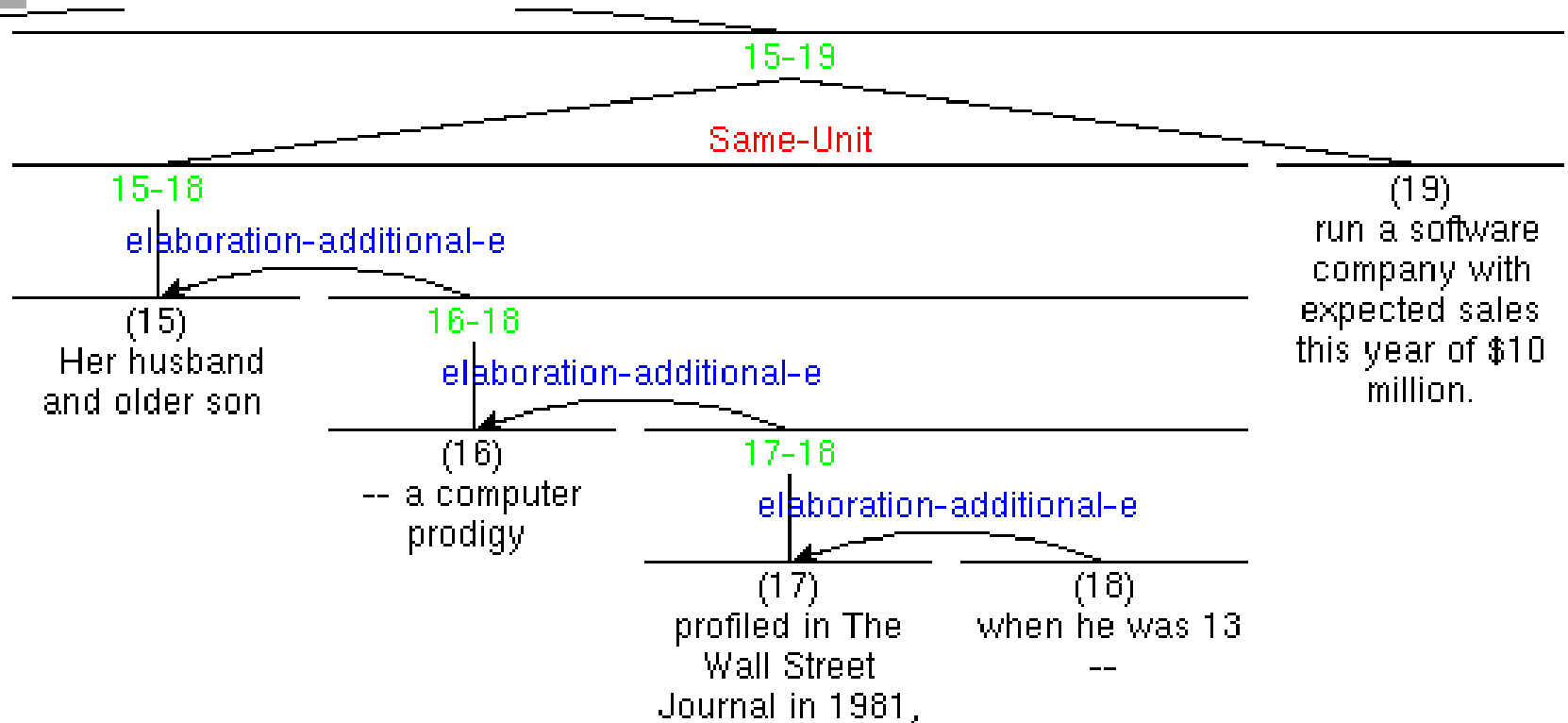
But parentheticals need one argument inside another:
Surfing, [despite being dangerous], is a lot of fun.

What rhetorical relations can be expressed by parentheticals?

Corpus study on two different discourse treebanks
(both annotate the same WSJ text)

- RST treebank (Carlson et al., 2001)
 - annotates rhetorical relations
 - distinguishes embedded relations
- Penn Discourse Treebank (PDTB-Group, 2008)
 - annotates discourse connectives and their arguments

RST Treebank: An Example



Results: RST Treebank

10 most frequent relations within SAME UNIT

331	42.93%	elaboration-additional
128	16.60%	attribution
58	7.52%	circumstance
35	4.54%	purpose
22	2.85%	restatement
20	2.59%	condition
19	2.46%	example
18	2.33%	antithesis
14	1.82%	elaboration-set-member
13	1.69%	concession
11	1.43%	elaboration-general-specific
102	13.23%	Other

771

Correlation between Rhetorical Relations and Syntax

		Elab-add	Example	Elab-gen-spec	Restatement	Elab-set-mem	Attribution	Condition	Antithesis	Concession	Circumstance	Purpose	
NP-modifiers	relative clause	143		2		2							147
	participial clause	96	4			1	1				11	4	117
	NP	34		8	22								64
	including + NP		13			5							18
	other	9	1			6			2	3	2		23
VP/S-modifiers	to-infinitive	4										30	34
	NP + V						106						106
	cue + S	5						20	14	9	29		77
	PP	11									9	1	21
	S	7	1	1									9
	other	1					18		2		3		24
		310	19	11	22	14	125	20	18	12	54	35	640

Results: Penn Discourse Treebank

Type of Connective	Connective in Host	Connective in Parenthetical	Total
Subordinating Conjunction	0	205	205
Discourse Adverbial	12	2	14
TOTAL	12	207	219

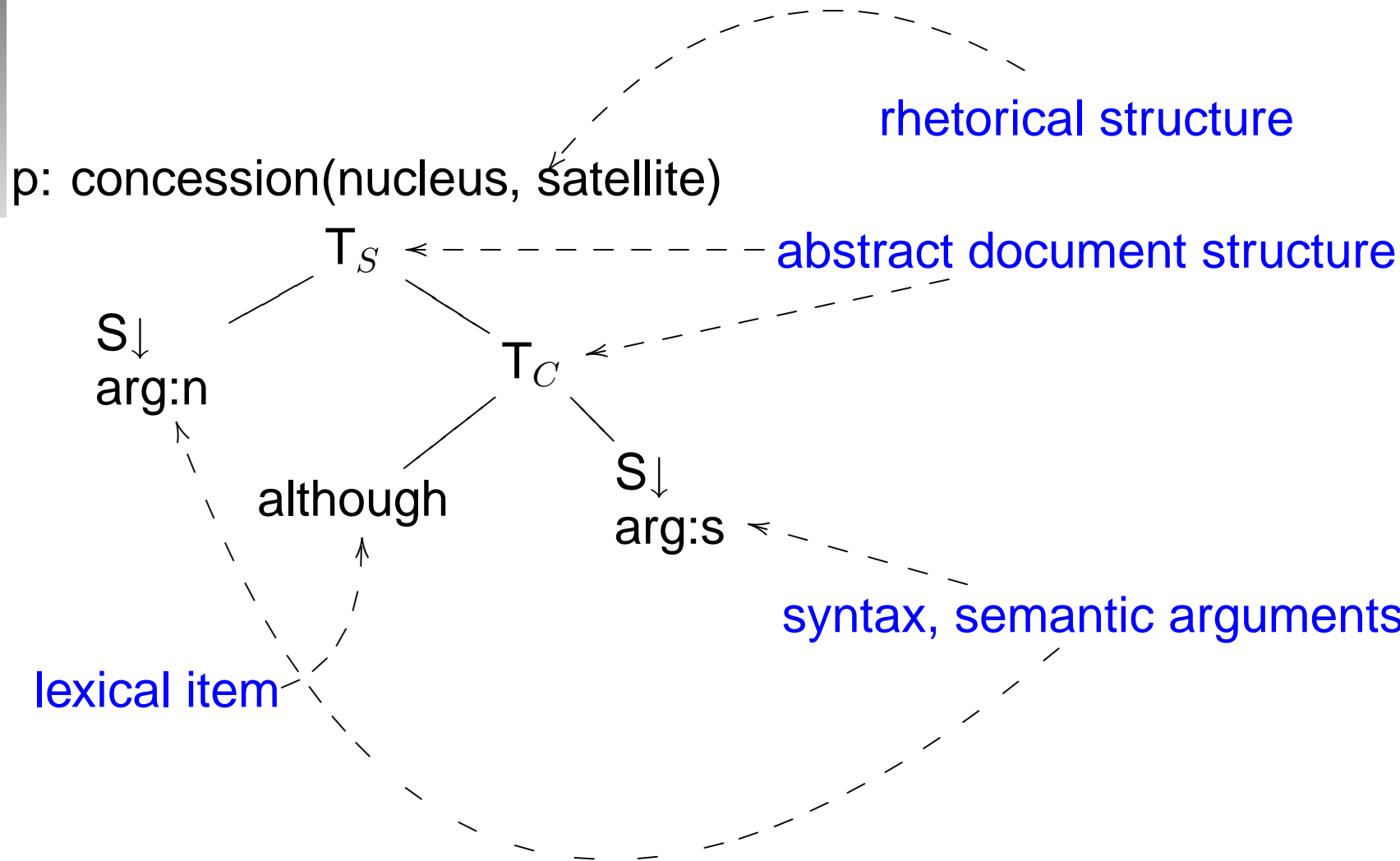
Incorporating the results of the study into an NLG system

Starting Points:

1. Rhetorical structure is a “semantic” concept
 - doesn't require arguments to be syntactically adjacent
 - interacts with syntax and abstract document structure
2. Integrated architecture
 - linguistic information stored in central knowledge base, using a Tree Adjoining Grammar

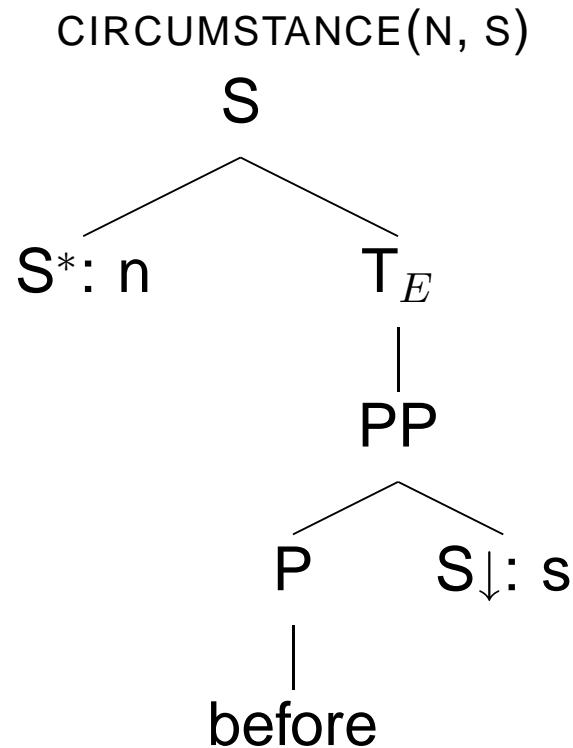
- an integrated representation using Tree Adjoining Grammar: Stone & Doran (1997), Koller & Striegnitz (2002)
- TAG-based realization and polarity filtering: Gardent and Kow (2007), Gardent and Kow (2006)
- abstract document structure and constraint-based NLG: Power Etal. (2003)

The “integrated” representation



An example: trees for CIRCUMSTANCE (1)

Subordinate clause with discourse connective:



In fiscal 1984, [before Mr. Gandhi came to power,] only \$810 million was raised. (wsj0629)

An example: trees for CIRCUMSTANCE (2)

Participial clause:

CIRCUMSTANCE(N, S)

VP

T_E

VP*: n

S↓:s

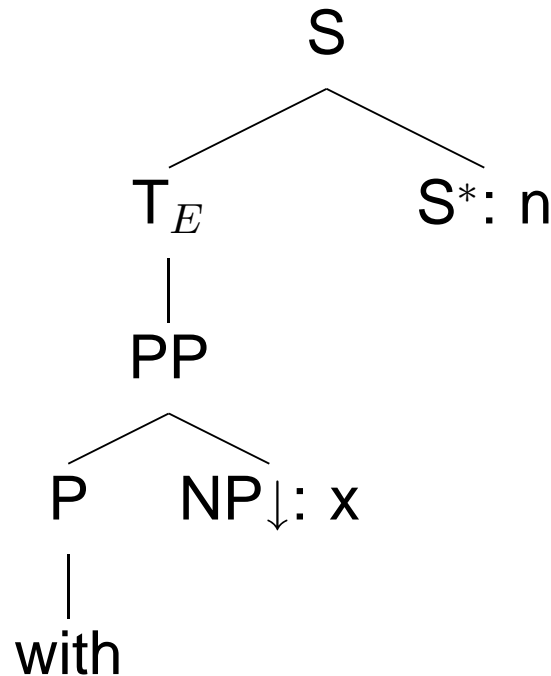
mode: ppart

The company, [currently using about 80% of its North American vehicle capacity,] has vowed it will run at 100% of capacity by 1992. (wsj2338)

An example: trees for CIRCUMSTANCE (3)

Prepositional Phrase (e.g. headed by 'with')

CIRCUMSTANCE(N, S), S: WITH(X)



But now, [with large amounts being raised from investors,] the government's dawdling on regulation has a more dangerous aspect. (wsj0629)

The generation process – Input

x: Prepulsid

p₁: is(x, a_gastrointestinal_drug)

p₂: do_well(x, overseas)

p₃: elaboration_additional(x, p₁)

Step 1. Tree selection

x: Prepulsid

NP: x

Prepulsid

p₂: do_well(x, overseas)

S:p₂

NP↓:x

VP

V

NP

did well

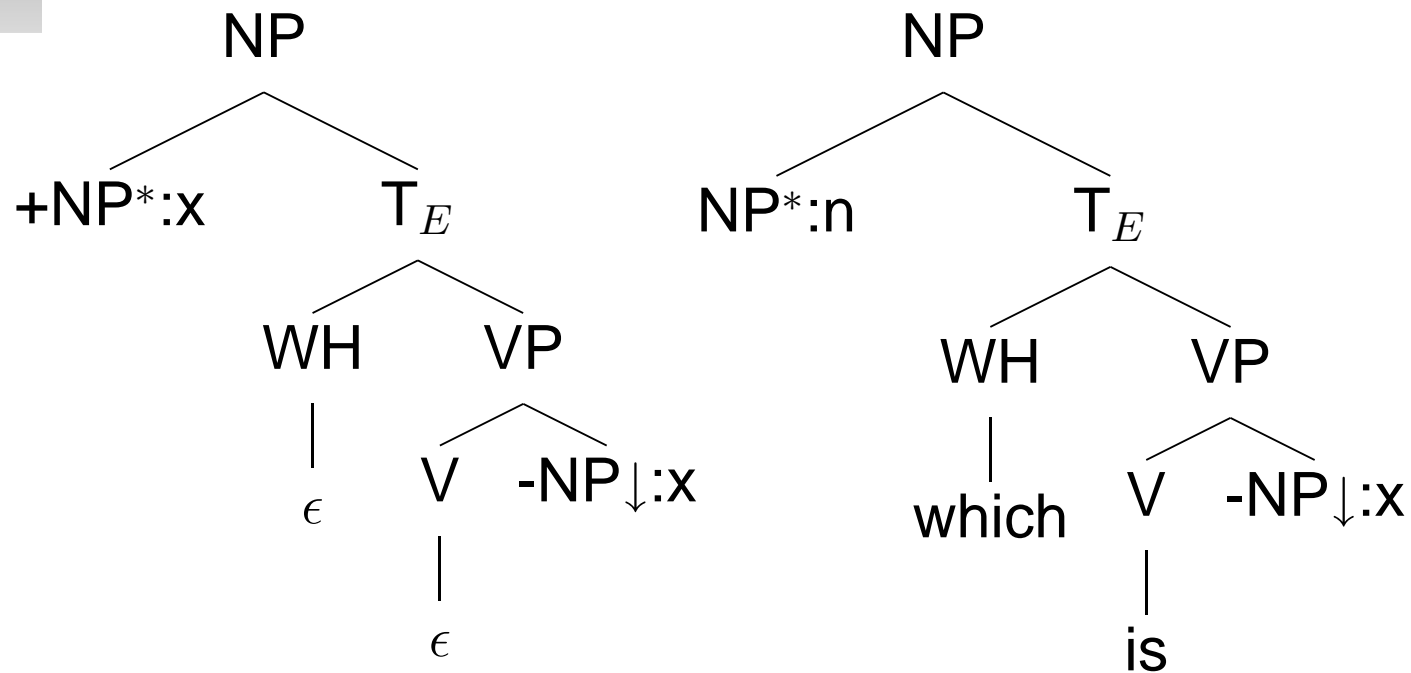
overseas

The generation process — Step 1:

Tree selection

p_3 : elaboration_additional(x, p_1)

p_1 : is(x, a_gastrointestinal_drug)



The generation process — Step 2:

Polarity Filtering

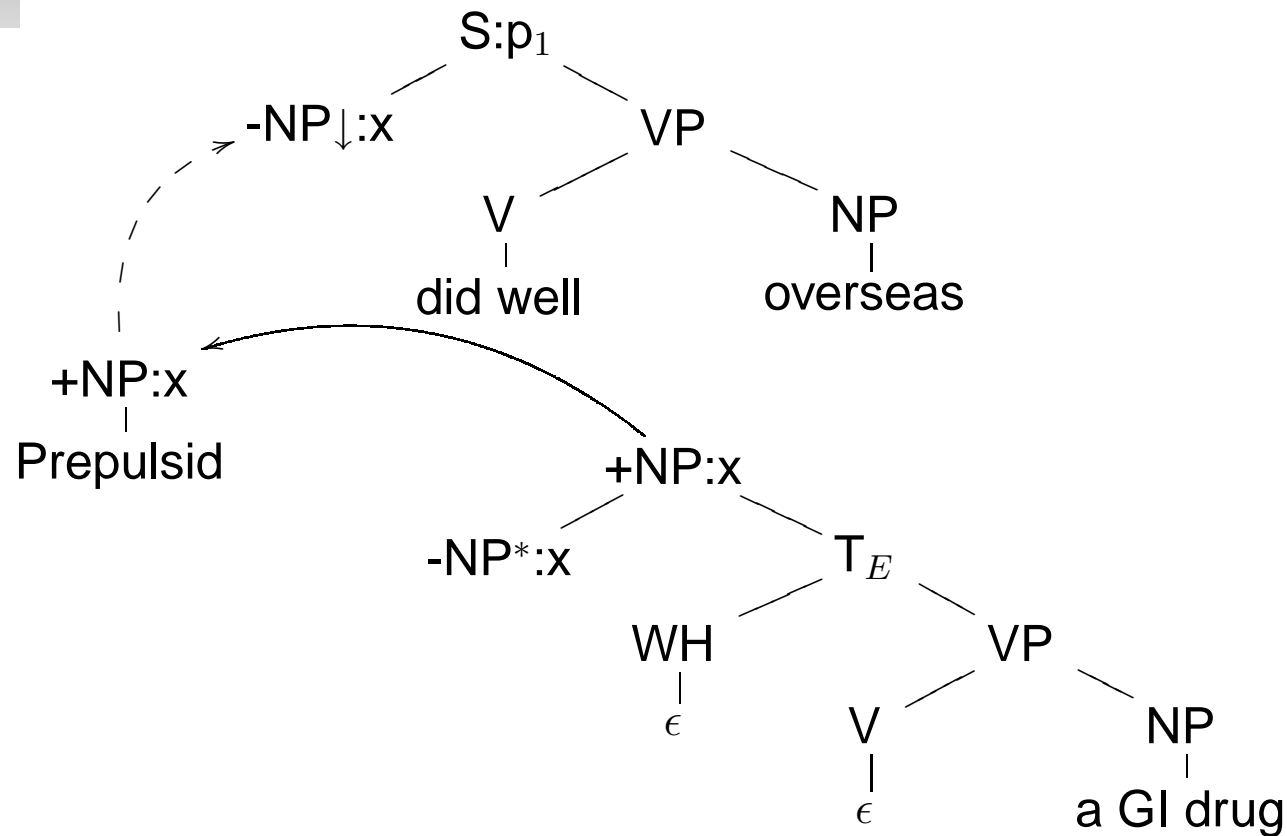
Polarity filtering (Gardent and Kow 2006) extended with semantic variables

- For substitution:
+NP:x, -NP:x,
- For adjunction:
+NP:x, -NP:x

The generation process — Step 3:

Combining the trees:

substitution and adjunction operations of Tree Adjoining Grammar (Joshi 1987)



The generation process — Step 4: linearization, punctuation

- punctuation marks inserted around the yield of T_E nodes

Prepulsid, [T_E a gastro-intestinal drug], did well overseas.

- Implementation currently under way.
- all possible solutions will be generated

- we have described an integrated generation architecture that is capable of realizing parenthetical constructions
- performed a corpus study to inform the generator:
 - studied rhetorical contexts that allow parentheticals
 - established correlation between syntactic types and rhetorical relations

Topics for further research:

- controlling the generator, e.g., by:
 - enriching the input with restrictions on trees that can be selected as in Gardent and Kow (2007)
 - adding ranking constraints to rank generated text
- further reducing computational complexity (e.g. by adding more parameters to polarity filtering)

Topics for further research:

How far can this definition of parentheticals be generalized?

"Express less important information and are not part of the main predicate-argument structure"

- The doctor examined Mary. He established that she had a sore throat.
- The doctor examined Mary. The doctor established that she had a sore throat. The doctor was male.

[+male] : a parenthetical?

Topics for further research:

- Last week, a jogger was hit by a car in Philadelphia.
- Last week, someone was jogging and was hit by a car in Philadelphia.

What is considered a parenthetical depends on the granularity of the semantic representation.
How far can we and should we decompose the semantics of words?

Thanks to Jerry Hobbs for pointing out these examples

Thank You!



The Open University