An Integrated Architecture for Generating Parenthetical Constructions

Eva Banik

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):

\[
\begin{array}{c}
\text{nucleus} \quad \text{CONCESSION} \quad \text{satellite} \\
S_1: \text{is(surfing, fun)} \quad \quad \quad \quad S_2: \text{is(surfing, dangerous)}
\end{array}
\]

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

15-18
elaboration-additional-e

(15) Her husband and older son

15-19
Same-Unit

(19) run a software company with expected sales this year of $10 million.

16-18
elaboration-additional-e

(16) -- a computer prodigy

17-18
elaboration-additional-e

(17) profiled in The Wall Street Journal in 1981,

(18) when he was 13 --
10 most frequent relations within SAME UNIT

<table>
<thead>
<tr>
<th>Relation</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>elaboration-additional</td>
<td>331</td>
<td>42.93%</td>
</tr>
<tr>
<td>attribution</td>
<td>128</td>
<td>16.60%</td>
</tr>
<tr>
<td>circumstance</td>
<td>58</td>
<td>7.52%</td>
</tr>
<tr>
<td>purpose</td>
<td>35</td>
<td>4.54%</td>
</tr>
<tr>
<td>restatement</td>
<td>22</td>
<td>2.85%</td>
</tr>
<tr>
<td>condition</td>
<td>20</td>
<td>2.59%</td>
</tr>
<tr>
<td>example</td>
<td>19</td>
<td>2.46%</td>
</tr>
<tr>
<td>antithesis</td>
<td>18</td>
<td>2.33%</td>
</tr>
<tr>
<td>elaboration-set-member</td>
<td>14</td>
<td>1.82%</td>
</tr>
<tr>
<td>concession</td>
<td>13</td>
<td>1.69%</td>
</tr>
<tr>
<td>elaboration-general-specific</td>
<td>11</td>
<td>1.43%</td>
</tr>
<tr>
<td>Other</td>
<td>102</td>
<td>13.23%</td>
</tr>
</tbody>
</table>

771
## Correlation between Rhetorical Relations and Syntax

<table>
<thead>
<tr>
<th></th>
<th>Elab-add</th>
<th>Example</th>
<th>Elab-gen-spec</th>
<th>Restatement</th>
<th>Elab-set-mem</th>
<th>Attribution</th>
<th>Condition</th>
<th>Antithesis</th>
<th>Concession</th>
<th>Circumstance</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NP-modifiers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relative clause</td>
<td>143</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>11</td>
<td>4</td>
<td>147</td>
<td>117</td>
<td>64</td>
</tr>
<tr>
<td>participial clause</td>
<td>96</td>
<td>4</td>
<td>8</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>NP</td>
<td>34</td>
<td>13</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>including + NP</td>
<td>9</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>VP/S-modifiers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to-infinitive</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>NP + V</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>34</td>
</tr>
<tr>
<td>cue + S</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>106</td>
</tr>
<tr>
<td>PP</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>310</td>
<td>19</td>
<td>11</td>
<td>22</td>
<td>14</td>
<td>125</td>
<td>20</td>
<td>18</td>
<td>12</td>
<td>54</td>
<td>35</td>
</tr>
</tbody>
</table>
## Results: Penn Discourse Treebank

<table>
<thead>
<tr>
<th>Type of Connective</th>
<th>Connective in Host</th>
<th>Connective in Parenthetical</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subordinating Conjunction</td>
<td>0</td>
<td>205</td>
<td>205</td>
</tr>
<tr>
<td>Discourse Adverbial</td>
<td>12</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>12</td>
<td>207</td>
<td>219</td>
</tr>
</tbody>
</table>
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
Related work


The “integrated” representation

- **p**: concession(nucleus, satellite)
- **TS**: abstract document structure
- **TC**: rhetorical structure
- **S ↓ arg:n**: lexical item
- **S ↓ arg:s**: syntax, semantic arguments

Although
An example: trees for \text{CIRCUMSTANCE} (1)

Subordinate clause with discourse connective:

$$\text{CIRCUMSTANCE}(N, S)$$

\[ S \]

\[ S^*: n \quad T_E \]

\[ \text{PP} \]

\[ P \quad S\downarrow: s \]

before

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

S

T_E

S*: n

PP

P

NP↓: x

with

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
  S:p₂
    NP↓:x
      VP
        V
        did well
        NP
        overseas
The generation process — Step 1: Tree selection

\[ p_3: \text{elaboration\_additional}(x, p_1) \]
\[ p_1: \text{is}(x, \text{a\_gastrointestinal\_drug}) \]
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!