Coherence Relations in (Machine) Translation

Jet Hoek j.hoek@uu.nl

Joint work with Sandrine Zufferey (University of Bern), Jacqueline Evers-Vermeul (UU), and Ted Sanders (UU)

Time in Translation workshop
23 June 2017
Introduction

*Coherence relations* connect discourse segments

They can, but need not be marked explicitly with a connective or cue phrase

Some relations are easy to convey implicitly

**Teen kills younger brother because he thought he deleted his Pokémon**
- snopes.com

**It’s summer, so let the tabloid body shaming begin.**
- boingboing.net

while other relations become very hard to reconstruct without explicit marking

**5 ways to show you care even though you forgot about Valentine’s Day**
- 12news.com

**If you are a young couple, Toronto Island wants you**
- thestar.com

→ Supported by analyses on discourse-annotated corpora
   (PDTB – Asr & Demberg 2012, RST – Das & Taboada 2013)
The marking of coherence relations

**General principle:**
Relations can remain unmarked if they can be easily inferred without explicit marking; if not, they should be explicitly marked.

When can a relation be easily inferred?

- Default expectations
- Signaled by other means
Parallel corpus study

Europarl Direct (Koehn 2005; Cartoni, Zufferey, & Meyer 2013)

1916 English source text relations

Connectives:  

Also  If
Although  In addition
Because  So
But  Unless

→ Annotated for relation type using CCR
(Cognitive approach to Coherence Relations, Sanders et al. 1992 and later work)

Translations into:  

Dutch  French  
German  Spanish

→ Is the ‘meaning’ of the connective explicitly expressed in the target language, and, if so, how?
Translations of coherence relations

Explicit:
Because John won the race, he very is happy.

Paraphrase:
John’s victory made him very happy.

ExplicitUNDSP:
John won the race and was very happy.

Syntax:
John, who won the race, is happy.

Implicit:
John is happy. Ø He won the race.

Other:
ParaphraseCONSTR:
If we want to stop climate change, we have to...
→ To stop climate change, we have to...

→ The relation in the target text has to correspond to the relation in the source text
Translations of coherence relations

Explicit:
Because John won the race, he is very happy.

Paraphrase:
John’s victory made him very happy.

ExplicitUNDSP:
John won the race and was very happy.

Syntax:
John, who won the race, is happy.

Implicit:
John is happy. Ø He won the race.

Other:
ParaphraseCONSTR:
If we want to stop climate change, we have to...
→ To stop climate change, we have to...

→ The relation in the target text has to correspond to the relation in the source text

**Subtype**: Restrictive RCs
Currently testing whether people can and do indeed infer coherence relations between restrictive RCs and their matrix clauses

- **Continuation experiment**
- **Self-paced reading**
- **Eye-tracking**
## Translations of coherence relations

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit</td>
<td><em>Because</em> John won the race, he very is happy.</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>John’s victory <em>made</em> him very happy.</td>
</tr>
<tr>
<td>Explicit UNDSP</td>
<td>John won the race <em>and</em> was very happy.</td>
</tr>
<tr>
<td>Syntax</td>
<td>John, <em>who won the race</em>, is happy.</td>
</tr>
<tr>
<td>Implicit</td>
<td>John is happy. <em>Ø</em> He won the race.</td>
</tr>
</tbody>
</table>

### Other:

**Paraphrase CONSTR**:  
- If we want to stop climate change, we have to...  
  - To stop climate change, we have to...

→ The relation in the target text has to correspond to the relation in the source text
Default expectations

In translation, connectives are often removed, added or rephrased

This should be bound by a relation’s potential to be implicit

Hypothesis: Cognitively simple relations are more expected than relations that are cognitively more complex

Implicitness: simple > complex
Implicitation: simple > complex

Cognitive complexity based on relation type → CCR primitives

• Processing
• Acquisition
• Logic
• Mental space theory

Teen kills younger brother because he thought he deleted his Pokémon

Polarity Positive
Basic operation Causal
Source of Coh Objective
Order Non-basic
Default expectations

In translation, connectives are often removed, added or rephrased. This should be bound by a relation’s potential to be implicit.

**Hypothesis:** Cognitively simple relations are more expected than relations that are cognitively more complex.

**Implicitness:** simple > complex

**Implicitation:** simple > complex

Cognitive complexity based on *relation type* → CCR primitives

- **Teen kills younger brother because he thought he deleted his Pokémon**

**Cross-linguistically**

- Processing
- Acquisition
- Mental space theory

<table>
<thead>
<tr>
<th>Polarity</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic operation</td>
<td>Causal</td>
</tr>
<tr>
<td>Source of Coh</td>
<td>Objective</td>
</tr>
<tr>
<td>Order</td>
<td>Non-basic</td>
</tr>
</tbody>
</table>
Segment-internal features

How do segment-internal features and connectives interact?

Collocation
I cannot come to your party, because my mom doesn’t like you
- telegraph.co.uk

Agreement
Doordat deze cellen zo dicht op elkaar zitten kunnen er alleen kleine deeltjes zoals zuurstof en voedingsstoffen doorheen.
- laboratorium.nl

‘Division of labor’
Don’t eat your coconut oil, use it for this stuff (instead).
- lifehacker.com
Segment-internal features

How do segment-internal features and connectives interact?

Collocation
I cannot come to your party, **because** my mom doesn’t like you
- telegraph.co.uk

Agreement
**Doordat** **deze cellen zo dicht op elkaar zitten** kunnen er alleen kleine deeltjes zoals zuurstof en voedingsstoffen doorheen.
- laboratorium.nl

‘Division of labor’
Don’t eat your coconut oil, use it for this stuff (**instead**).
- lifehacker.com
Segment-internal features and connectives

→ Look at changes in marking in the TT compared to the ST

**Less specific**

**EN**  Unless [we take that way,] [the only alternative will be more misery, more destruction and more death.]

**DE**  Wenn [wir diesen Weg nicht einschlagen,] dann [wird die einzige Alternative mehr Elend, mehr Zerstörung und mehr Tod sein.]

**More specific**

**EN**  [Help us to be more precise with the road-map] because [we want to follow it.]

**NL**  [Help ons de routebeschrijving nauwkeuriger te maken,] zodat [we hem kunnen volgen.]
Discussion

In this project, we used parallel corpora:

• To create a discourse-annotated parallel corpus

• To learn something about translation
  o Created an inventory of the ways in which coherence relations are translated
  o Improved knowledge about which types of coherence relations should be explicitly translated

  → Input for MT research

• As a method to study monolingual phenomena → cross-linguistically
  o The explicit vs. implicit marking of coherence relations is influenced by the relation’s cognitive complexity
  o The systematicity in the explicit vs. implicit marking is similar across languages
  o Explored the interaction between connectives and segment-internal elements

  → Further our understanding of discourse and human cognition
Thank you!

Jet Hoek
j.hoek@uu.nl