Uniform Information Density at the Level of Discourse Relations

Fatemeh Torabi Asr and Vera Demberg
Collaborative Research Center SFB1102
Saarland University, Germany
John did not go to the concert. He was ill.
John did not go to the concert. He went to the cinema.
Discourse relations

Linguistic features in the arguments

John did not go to the concert. He went to the cinema.
Explicit discourse connective

John did not go to the concert. **Instead**, he went to the cinema.
Why aren’t the connectives always used?

Relations in Penn Discourse Treebank

<table>
<thead>
<tr>
<th># Implicit</th>
<th># Explicit</th>
</tr>
</thead>
<tbody>
<tr>
<td>47%</td>
<td>53%</td>
</tr>
</tbody>
</table>
When are discourse relations explicitly marked in natural text?

A communication & information perspective:
  – Uniform Information Density hypothesis
  – Applying UID to discourse connective utilization
  – Case study: Chosen alternative relations

  Connective omission in presence of negation cues
Hypothesis: connectives are used only if the information they deliver is essential for communication.
Hypothesis: connectives are used only if the information they deliver is essential for communication.

Early related notions:

• Principle of least effort (Zipf 1965)
• Maxim of quantity (Grice 1975)
Uniform Information Density (Levy & Jaeger 2007)

Among equivalent forms speakers naturally choose the one that delivers information more uniformly across the utterances.
Uniform Information Density (Levy & Jaeger 2007)

Among equivalent forms speakers naturally choose the one that delivers information more uniformly across the utterances.
Uniform Information Density (Levy & Jaeger 2007)

Among equivalent forms speakers naturally choose the one that delivers information more uniformly across the utterances.
Uniform Information Density (Levy & Jaeger 2007)

Among equivalent forms speakers naturally choose the one that delivers information more uniformly across the utterances.
Uniform Information Density (Levy & Jaeger 2007)

Among equivalent forms speakers naturally choose the one that delivers information more uniformly across the utterances.
Measuring the information delivered by a unit:

\[ \text{Surprisal}(\text{unit}) = - \log p(\text{unit} | \text{context}) \]

(Hale 2001, 2003; Levy 2008)
Uniform Information Density

Measuring the information delivered by a unit:

**Surprisal**\(\text{unit}) = - \log p(\text{unit} | \text{context})

(Hale 2001, 2003; Levy 2008)

Term comes from comprehension studies

\[-\log P(w_i | w_1 \ldots w_n)\]
Jaeger 2010:

“My boss confirmed ...

our appointment with...

we were absolutely...

(1) start of a complement clause

(2) subject being “we”
Jaeger 2010:

“My boss confirmed ...

our appointment with...

we were absolutely...

(1) start of a complement clause
(2) subject being “we”

“Confirm”: not predictive of the continuation type
“think”: highly predictive of complement clause continuation

=> “that” is more needed after “confirm” to deliver info (1)
Uniform Information Density

Jaeger 2010

(a)
Jaeger 2010

Uniform Information Density

(b)

Information density in bits/word

seconds

My
boss
thinks
am
that
absolutely
crazy

Asr & Demberg, Saarland Uni.

IWCS 2015
Applied to discourse connectives:
Connective omission ~ Relation’s predictability


- Continuity hypothesis (Segal et al., 1991; Murray, 1997)
- Causality-by-default hypothesis (Sanders, 2005; Kuperberg et al., 2011)
Discourse-level UID: Local cues

Connective omission ~ Relation’s predictability

Current study: Local cues

Sentence 1  Connective  Sentence 2
Discourse-level UID

Find relations in natural language

Find present cues

Make sure cues make sense

Measure predictability

Correlated with connective omission?
• Available annotated resource: Penn Discourse Treebank
  – Gold standard annotation of connectives & relations
  – Other features need to be automatically extracted

• Motivation for studying specific relations: Chosen alternative
  – Downward entailing in Arg1 (Webber 2013)
  – Example:
    “No price for the new shares has been set. Instead, the companies will leave it up to the marketplace to decide.”
We expect:

1. **Negation** is a statically plausible cue for Chosen alternative (CA) relations.
2. **Connective** is less likely to occur in presence of negation (for CA).
State-of-the-art features of the discourse relations:

- connectives,
- n-grams,
- production rules,
- polarity markers,
- verb classes, etc.
State-of-the-art features of the discourse relations:

- connectives,
- n-grams,
- production rules,
- polarity markers,
- verb classes, etc.

Highly informative features of the *Chosen Alternative* relation (auto-extraction)
Discourse-level UID: Case study

Highly informative features of the *Chosen Alternative* relation (auto-extraction)

State-of-the-art features of the discourse relations:

- connectives,
- n-grams,
- production rules,
- polarity markers,
- verb classes, etc.
• Cleaning the data:
  – Relations with sentence-initial connectives removed (~ 6%)
  – A binary feature indicates presence/absence of:
    » not, n’t, no, without, never, neither, none, non, nor, nobody, nothing

14% of all relations have some negation in Arg1
Relation senses obtaining positive npmi with negation cues (all but last significant at p<0.001)

- Expansion.Alternative.Chosen alternative (171,117)
- Comparison.Concession.Expectation (31,179)
- Comparison.Contrast (1236,2226)
- Comparison (157,378)
- Contingency.Cause.Reason (2467,1316)
- Contingency.Cause.Result (1704,748)
- Comparison.Concession.Contra-exception (186,791)
We expect:

✓ **Negation** is a statically plausible cue for Chosen alternative (CA) relations.

2. **Connective** is less likely to occur in presence of negation (for CA).
Relation senses obtaining positive npmi with negation cues (all but last significant at p<0.001)
Correlation between presence of negation & absence of connective (significant ones are stared)

*Expansion.Alternative.Chosen alternative (171,117)*

*Comparison.Concession.Expectation (31,179)*

*Comparison.Contrast (1236,2226)*

*Comparison (157,378)*

*Contingency.Cause.Reason (2467,1316)*

*Contingency.Cause.Result (1704,748)*

*Comparison.Concession.Contra-expectation (186,791)*

---

**Connective ~ Negation**

- npmi(Implicit $R$, Arg1Neg)
- npmi(Explicit $R$, Arg1Neg)
- npmi($R$, Arg1Neg)

Asr & Demberg, Saarland Uni.

IWCS 2015
We expect:

✓ **Negation** is a statically plausible cue for Chosen alternative (CA) relations.

✓ **Connective** is less likely to occur in presence of negation (for CA).
Conclusion

• Discourse connectives are dropped when the relation is expected,
  – given general cognitive biases (our previous work)
  – given local cues in the first argument of the relation (here).
• This is a support for the mechanism that UID proposes.
Conclusion

• Discourse connectives are dropped when the relation is expected,
  – given general cognitive biases (our previous work)
  – given local cues in the first argument of the relation (here).
• This is a support for the mechanism that UID proposes.

• Future directions:
  – Larger scale evaluation
  – NLG application
Thank you!

Also, thanks to Florian Jaeger and Bonnie Webber for great discussions and feedback on the experiments.
Thanks to Johannes Pietsch and Anne-Marie Friedrich for sharing code.

References (please see the paper for a full list):

- Grice (1975) *Logic and conversation*.
- Levy and Jaeger (2007) *Speakers optimize information density through syntactic reduction*.
- Webber (2013) *What excludes an alternative in coherence relations?*
- Asr and Demberg (2012) *Implicitness of the discourse relations*.

Some pictures in this presentation have been borrowed from public domain (internet) but might be subject to copy right.