

Uniform Information Density at the Level of Discourse Relations

Fatemeh Torabi Asr and Vera Demberg
Collaborative Research Center SFB1102
Saarland University, Germany

Discourse relations

Reason

John did not go to the concert. He was ill.



Discourse relations

Chosen alternative

John did not go to the concert. He went to the cinema.



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

# Implicit	# Explicit
47%	53%

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)

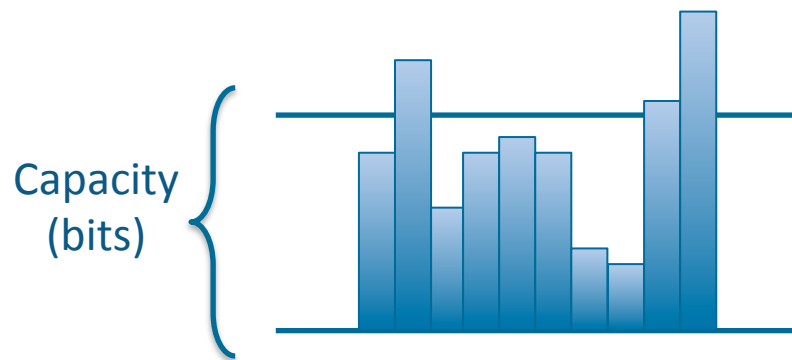
Uniform Information Density (Levy & Jaeger 2007)

Among equivalent forms speakers naturally choose the one that delivers information more uniformly across the utterances.



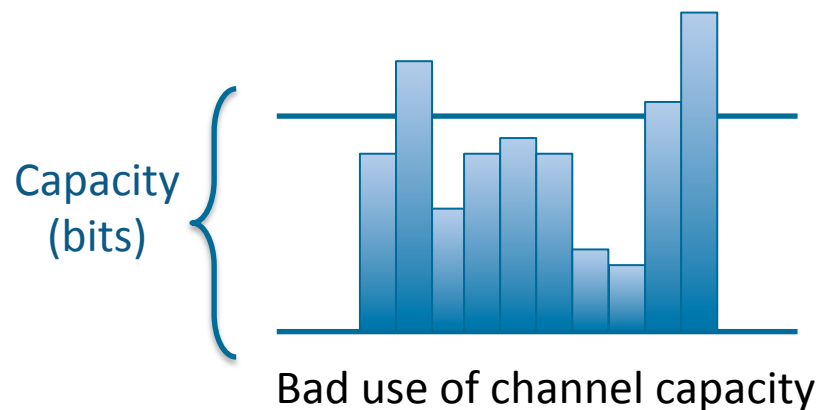
Uniform Information Density (Levy & Jaeger 2007)

Among equivalent forms speakers naturally choose the one that delivers information more uniformly across the utterances.



Uniform Information Density (Levy & Jaeger 2007)

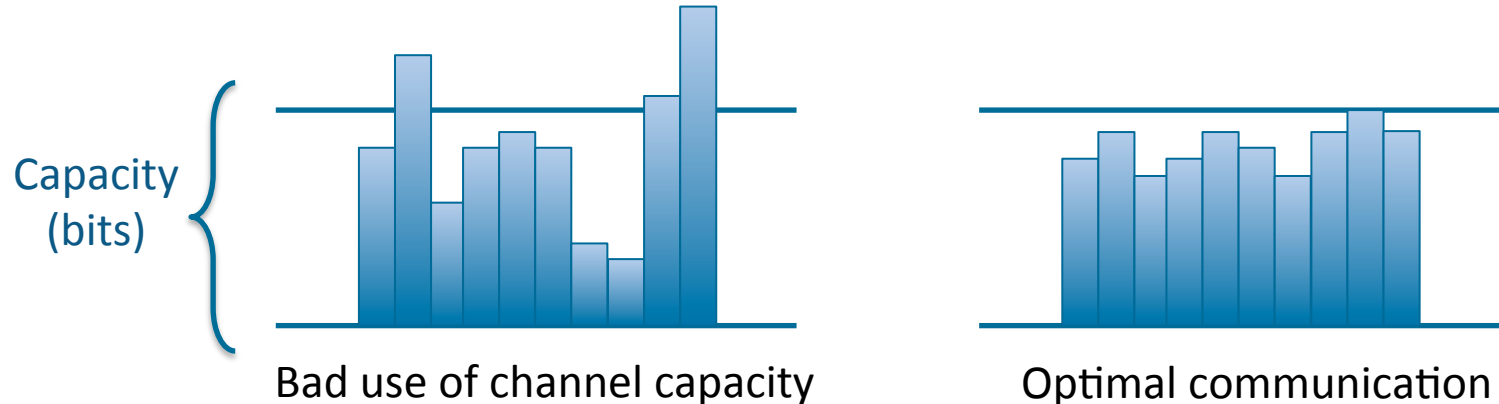
Among equivalent forms speakers naturally choose the one that delivers information more uniformly across the utterances.



Uniform Information Density

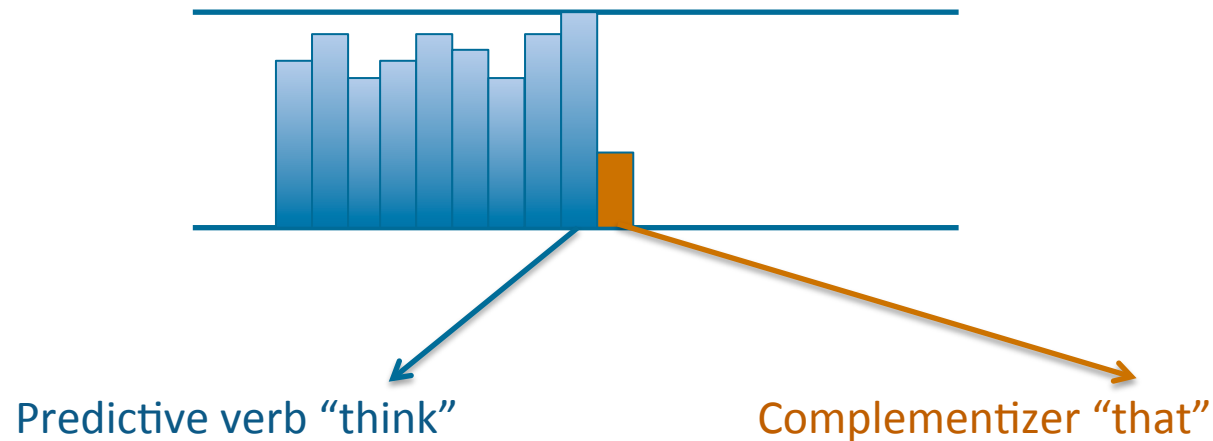
Uniform Information Density (Levy & Jaeger 2007)

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.



Jaeger 2010

Measuring the information delivered by a unit:

$$\text{Surprisal}(\text{unit}) = -\log p(\text{unit} \mid \text{context})$$



Term comes from
comprehension
studies

(Hale 2001, 2003; Levy 2008)

Uniform Information Density

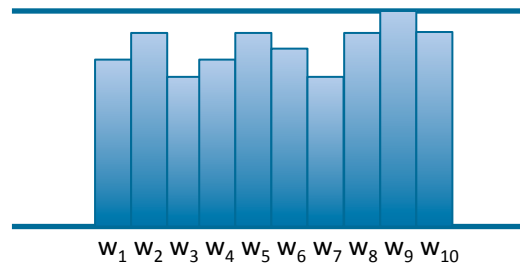
Measuring the information delivered by a unit:

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

Term comes from
comprehension
studies

(Hale 2001, 2003; Levy 2008)

$$-\log P(w_i | w_1 \dots w_n)$$



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”

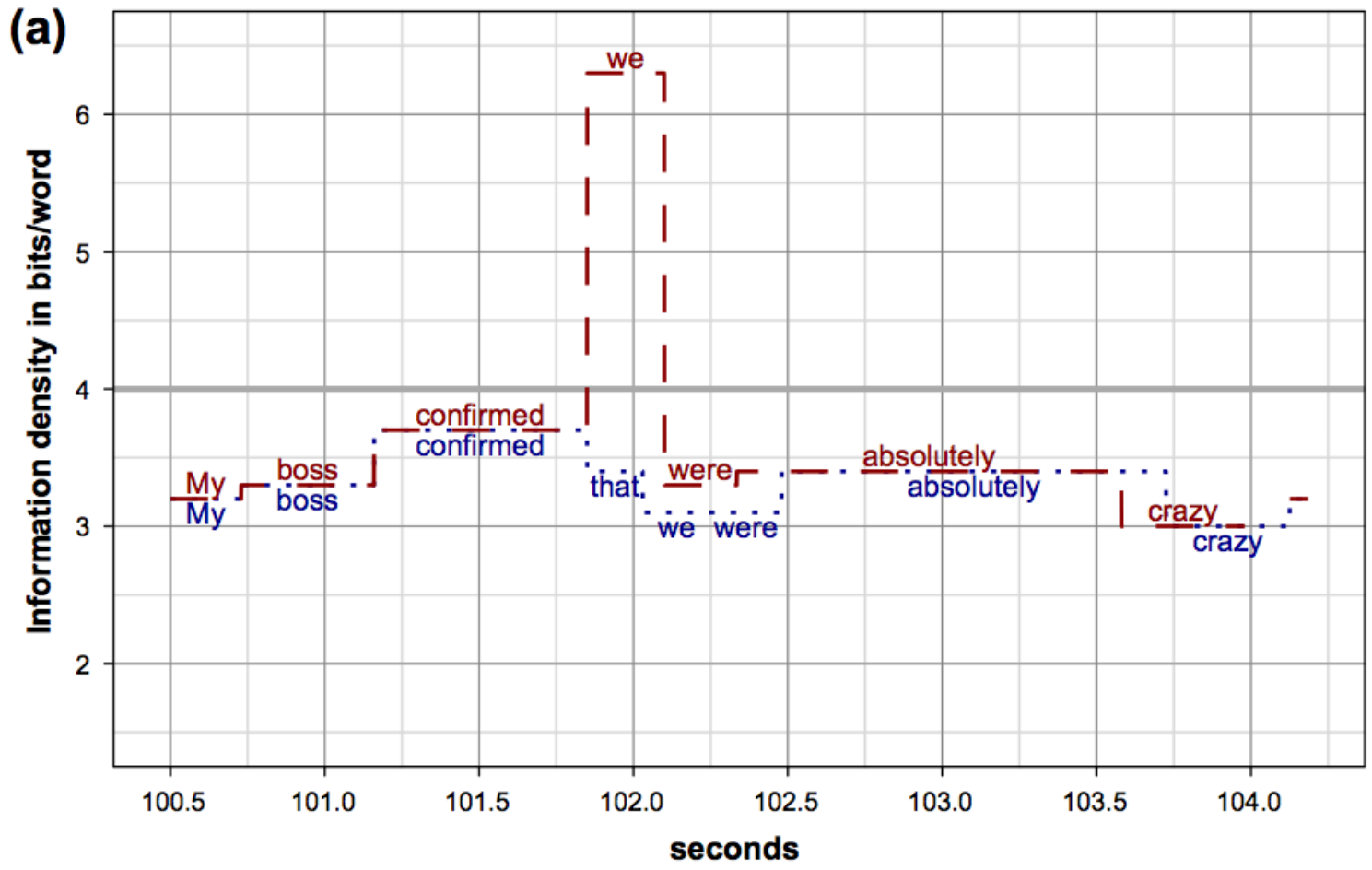
“Confirm”: not predictive of the continuation type

“think”: highly predictive of complement clause continuation

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

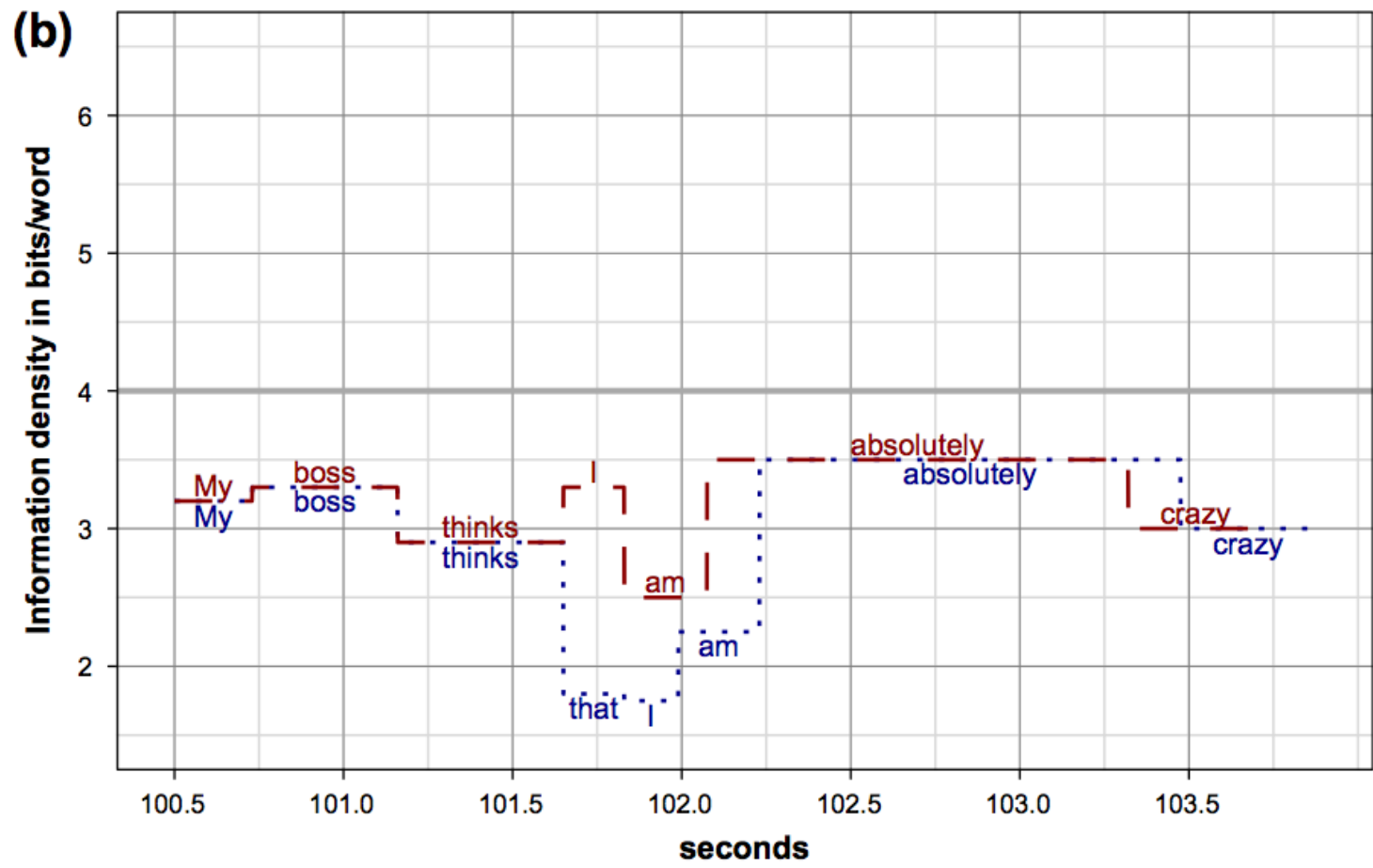
Uniform Information Density

Jaeger 2010



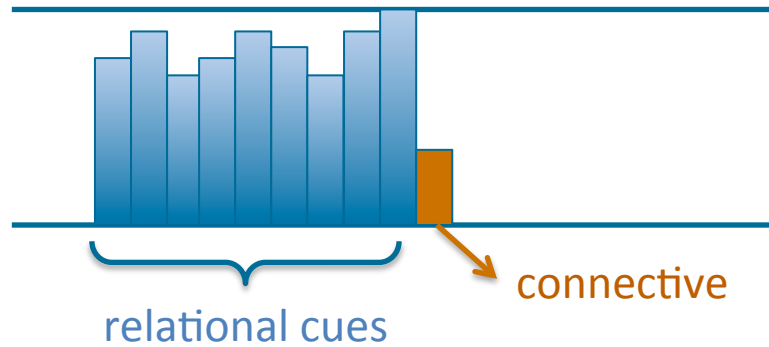
Uniform Information Density

Jaeger 2010



Uniform Information Density

Applied to discourse connectives:



Connective omission ~ Relation's predictability

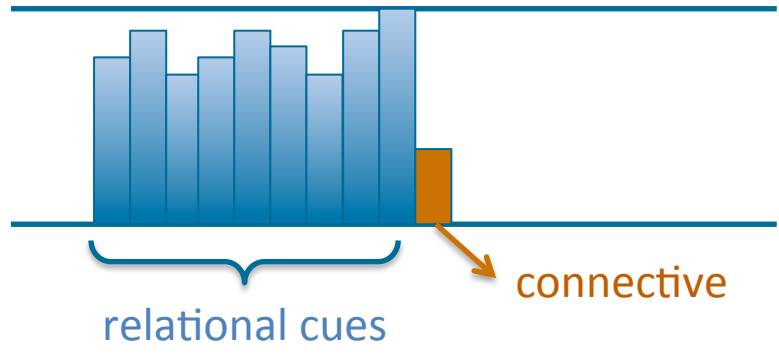
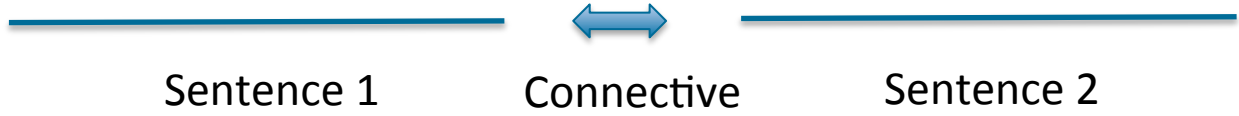
Asr & Demberg (2012, 2013): Cognitive biases

- Continuity hypothesis (Segal et al., 1991; Murray, 1997)
- Causality-by-default hypothesis (Sanders, 2005; Kuperberg et al., 2011)

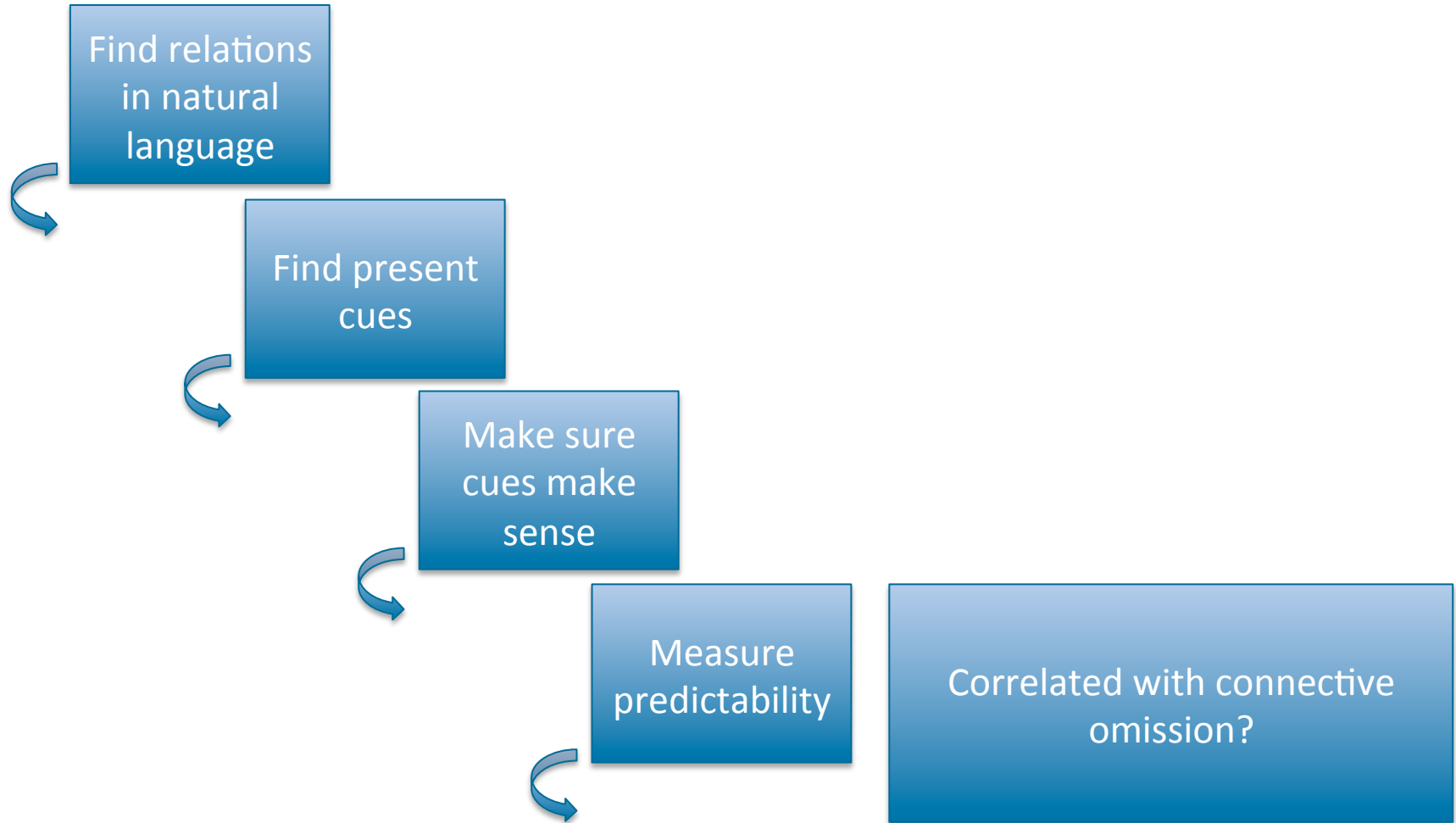


Connective omission ~ Relation's predictability

Current study: Local cues



Discourse-level UID



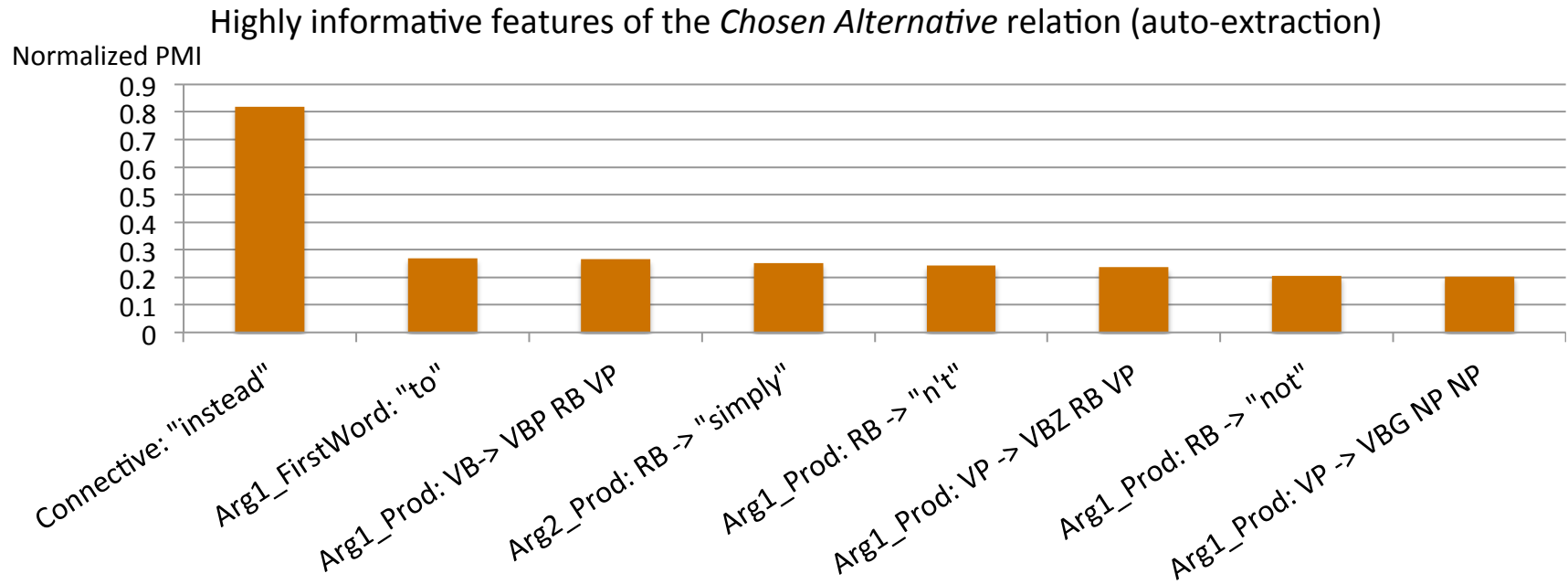
- Available annotated resource: Penn Discourse Treebank
 - Gold standard annotation of **connectives & relations**
 - **Other features** need to be automatically extracted

- 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).

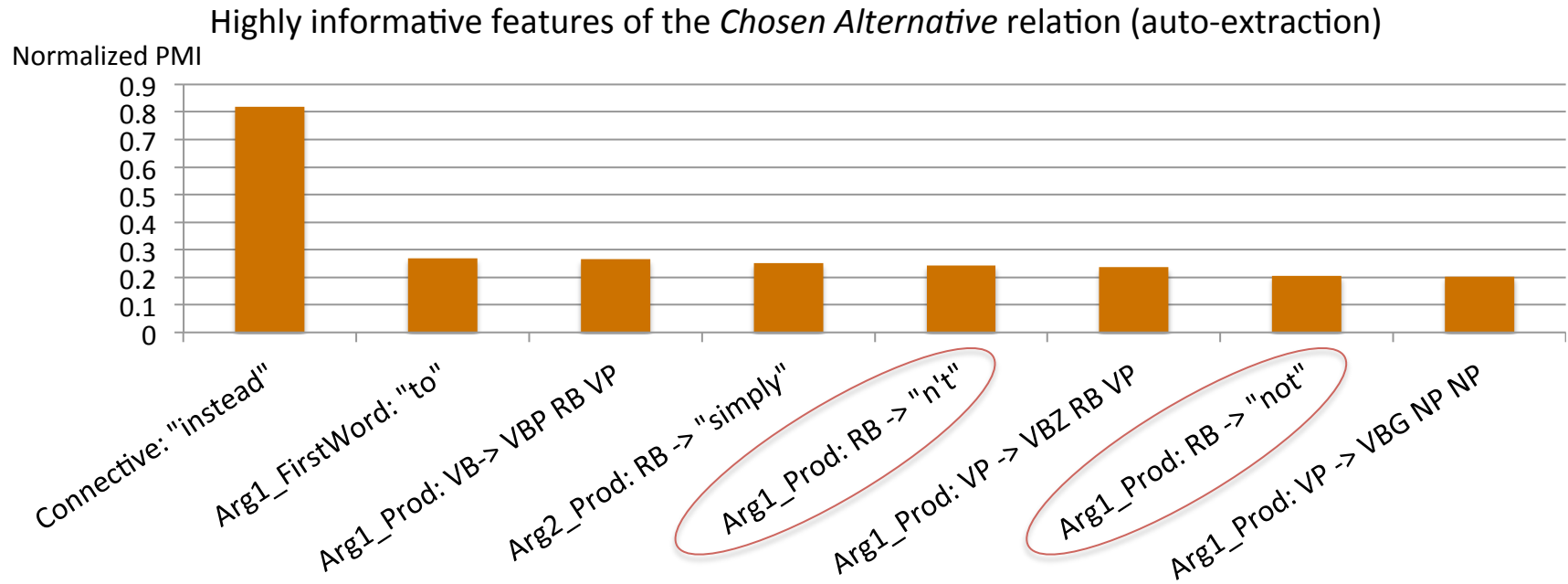
Discourse-level UID: Case study



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

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

Discourse-level UID: Case study

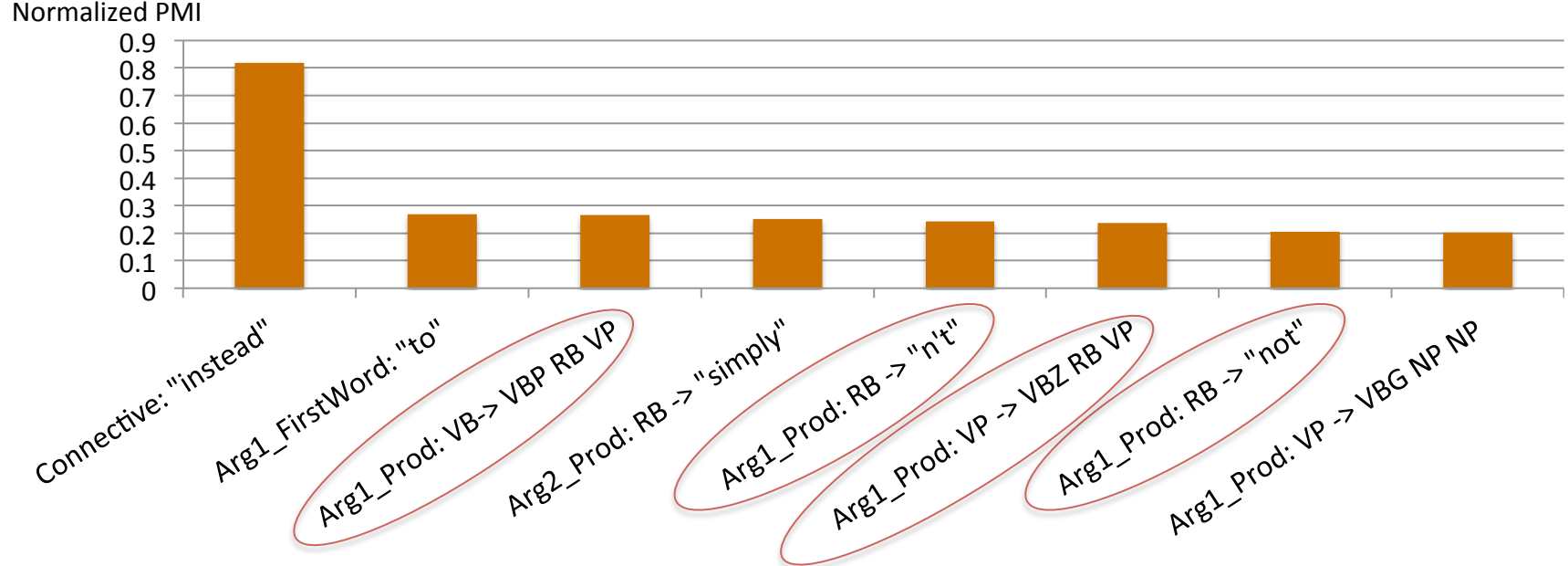


State-of-the-art features of the discourse relations:

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

Discourse-level UID: Case study

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



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

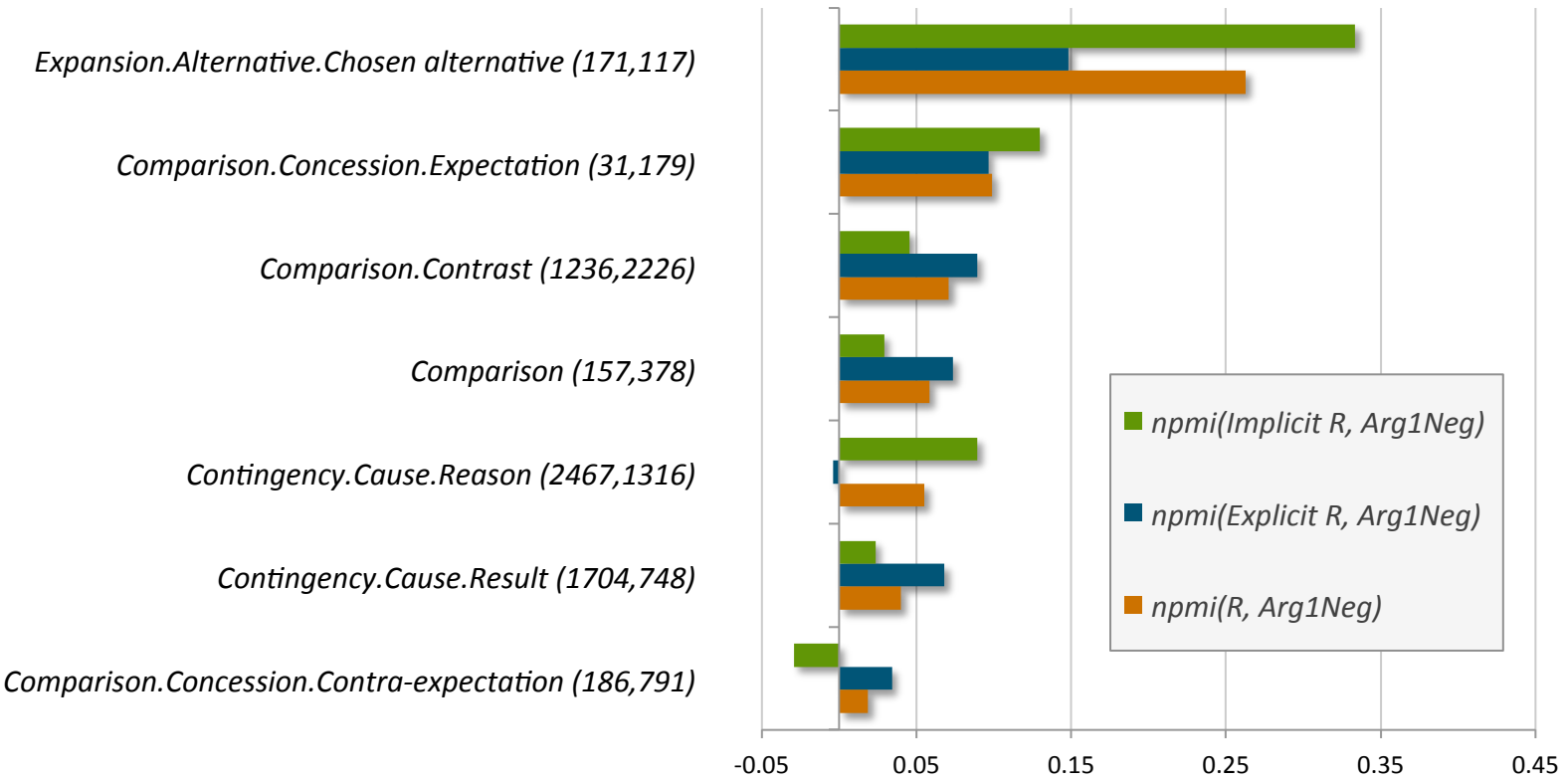
- connectives,
- n-grams,
- production rules,
- polarity markers,
- verb classes, etc.

- 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

Discourse-level UID: Case study

Relation senses obtaining positive npmi with negation cues (all but last significant at $p < 0.001$)

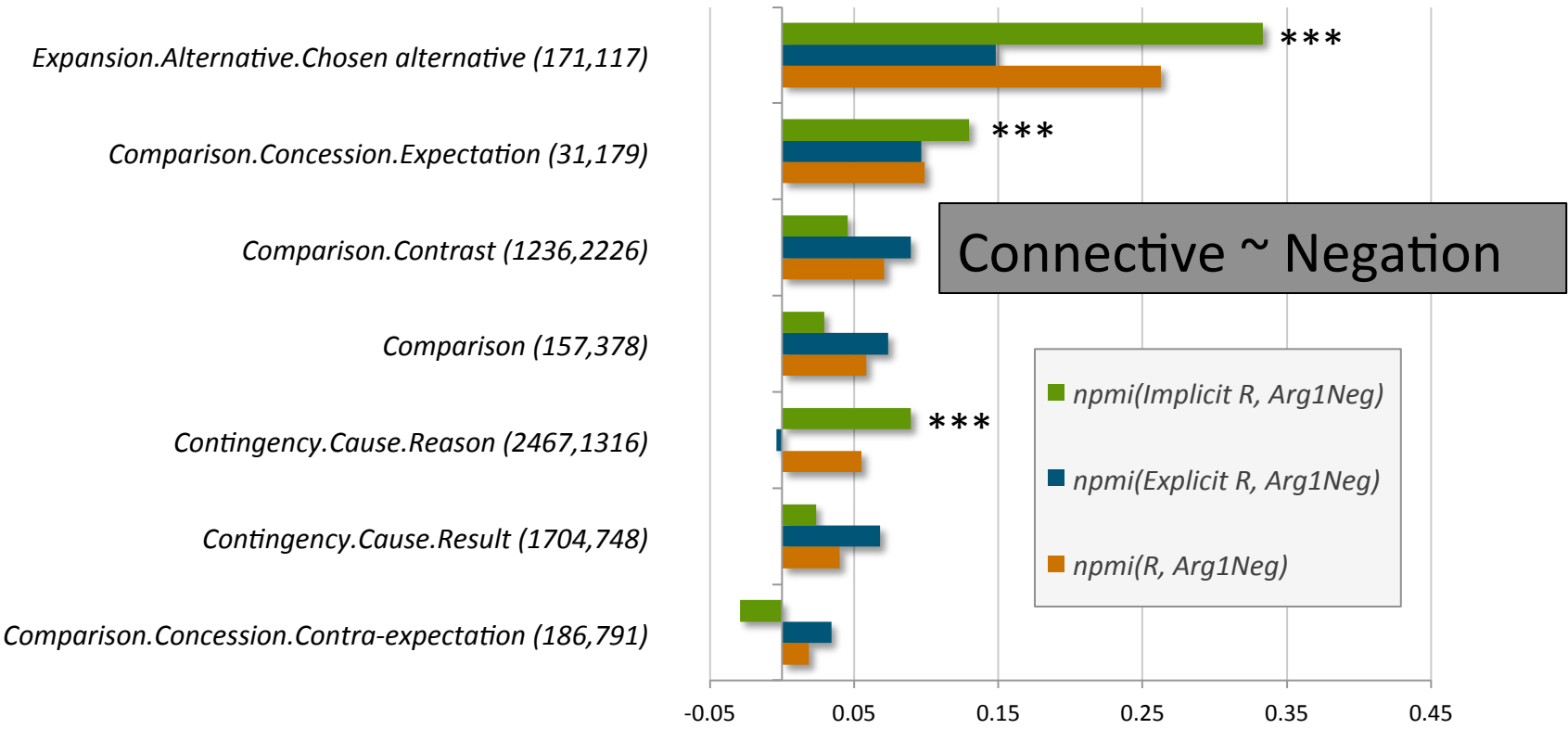


We expect:

- ✓ **Negation** is a statically plausible cue for Chosen alternative (CA) relations.
- 2. **Connective** is less likely to occur in presence of negation (for CA).

Discourse-level UID: Case study

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 starred)



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?*
- Prasad, Dinesh, Lee, Miltsakaki, Robaldo, Joshi, and Webber (2008) *The Penn Discourse Treebank 2.0*.
- 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.