Explicit Discourse Connectives

Implicit Discourse Relations

THE UNIVERSITY of EDINBURGH

Bonnie Webber
Hannah Rohde
Anna Dickinson
Annie Louis

Nathan Schneider

GEORGETOWN UNIVERSITY

Aravind Joshi

1929–2017

Discourse coherence

<u>Recipe for whipped cream frosting</u>:

Put cream cheese and whipping cream into a bowl.

(then)

Add sugar and vanilla.

(then)

Beat the mixture until the cream can hold a stiff peak.

(then)

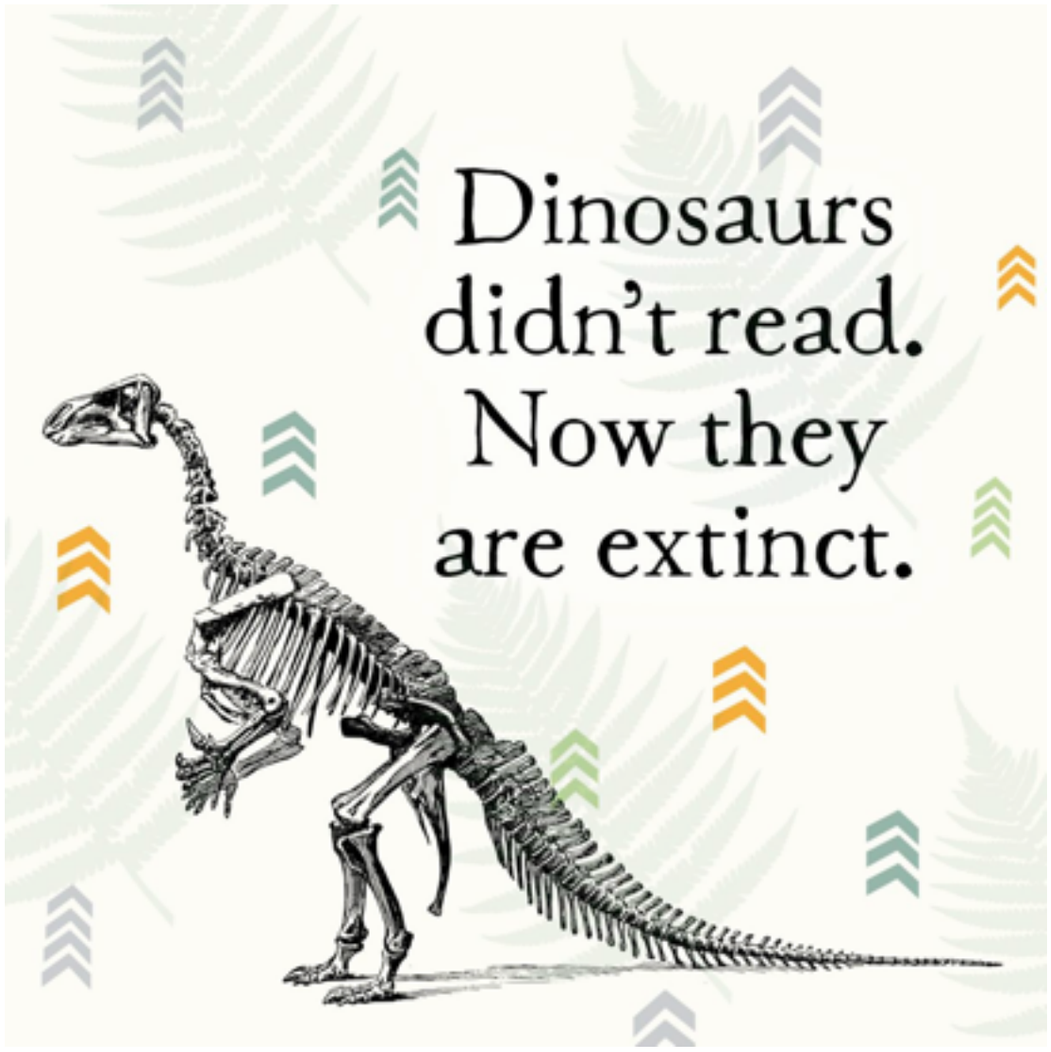
Cover cakes with this frosting that won't melt at room temperature.

Otherwise you'll be left with soggy cupcakes.

→ Some relations can be left implicit; others can't.

(Asher & Lascarides, 2003; Hobbs, 1979; Kehler, 2002; Mann & Thompson, 1988; Prasad et al, 2014; Roberts, 1996; Sanders et al., 1992)

Implicit discourse relations



Discourse connectives

Conjunctions

and
because
but
or
so

Adverbials

actually
after all
afterwards
first of all
for example
for instance
hence
however
in fact
in general
in other words
indeed
instead
meanwhile
nevertheless
nonetheless
on the one hand
on the other hand
otherwise
previously
specifically
then
therefore
thus

Both so therefore, or otherwise, ...

This talk

1. **Do inferable discourse relations hold when a discourse adverbial is already present?**
   → Yes, adverbials license co-occurring conjunctions
2. **How to characterise discourse adverbials with respect to inferred relations?**
   → Not predictable from adverbial or semantic class
   → More than one valid connection in some cases
3. **How to account for unexpected combinations?**
   → Multiple simultaneous sources of coherence

Unfortunately, nearly 75,000 acres of tropical forest are converted or deforested every day ____ in other words an area the size of Central Park disappears every 16 minutes.

○ or
○ so

→ are OR and SO substitutable in this context?

Implicit/explicit

- Deduction of implicit information from juxtaposed sentences

It's too far to walk. Let's take the bus.

Infer alternatives: walk/bus as means of transport
Infer causal relation: too far, therefore bus

It's too far to walk so let's take the bus.

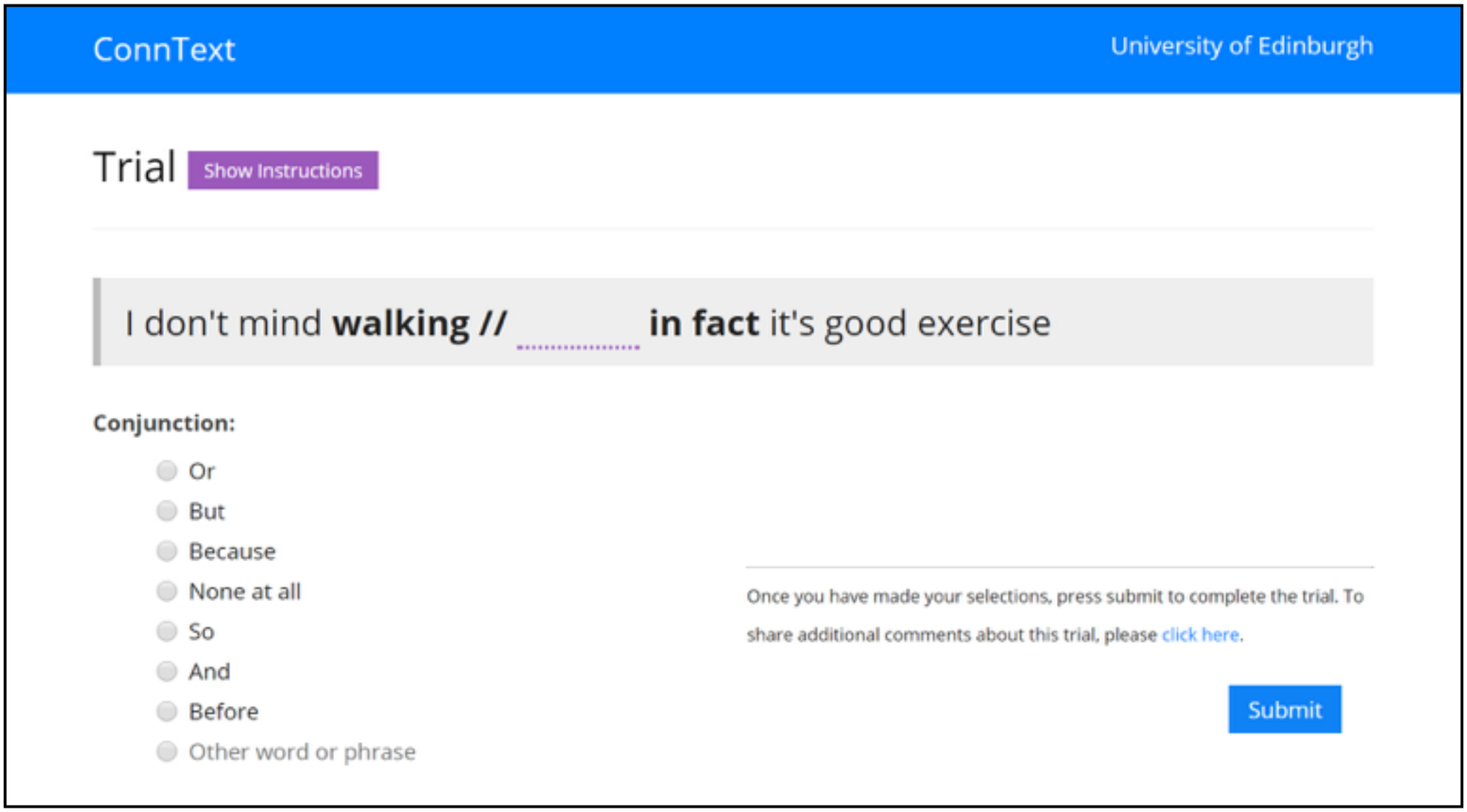
- **Assumption:** A passage marks its coherence relation either explicitly or implicitly — i.e., if explicit connective is present, no need for pragmatic inference about additional relations.

It's too far to walk. so? Instead **let's take the bus.**

Overarching question

- Given a **discourse adverbial**, which **conjunction(s)** is/are compatible and why?
  - Passage-dependent?
  - Reader-dependent/multiple interpretations?
    - If no conjunction, is there an implicit coherence relation?
    - With conjunction + adverbial, do they signal **different** coherence relations, or **the same** relation?
- Implications for corpus annotation and NLP (understanding/generation)

Fill-in-the-blank study



→ Dataset of judgments for 50 adverbials, each in 50+ passages, each passage judged by 28 people… 70,000+ data points

(Rohde et al., 2015, 2016, 2017)

Details for study 1

- Materials: for each adverbial, 50+ passages (mostly) from NYTimes Annotated Corpus (Sandhaus, 2008)
- Half originally explicit

"Nervous? No, my leg's not shaking," said Griffey, who caused everyone to laugh // _____ indeed his right foot was shaking.

Author=BECAUSE

- Half originally implicit

Sellers are usually happy, too // _____ after all they are the ones leaving with money.

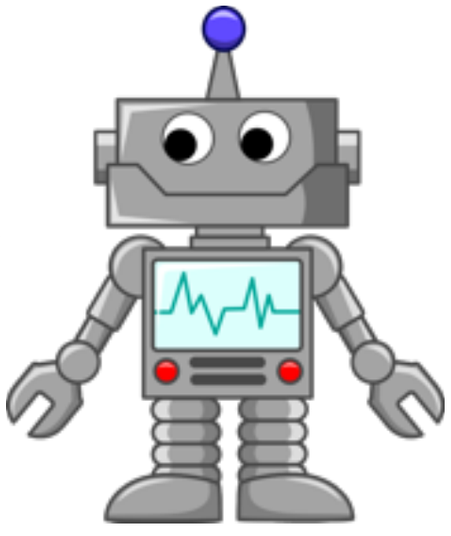
Author=NONE

Adverbials include: ACTUALLY, AFTER ALL, FIRST OF ALL, FOR EXAMPLE, FOR INSTANCE, IN FACT, IN OTHER WORDS, INDEED, INSTEAD, NEVERTHELESS, NONETHELESS, ON THE ONE HAND, ON THE OTHER HAND, OTHERWISE, SPECIFICALLY, THEN, THEREFORE, THUS, …

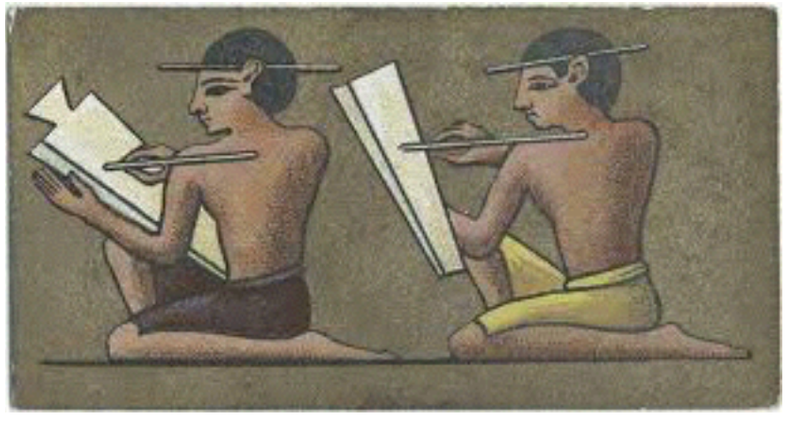
Hypotheses

- **Variability across adverbials:** Do adverbials pattern uniformly or vary across adverbials (by semantic type)?
- **Variability within adverbials:** Does the adverbial predict the same conjunction for all passages?

- If deterministic →



- If not →

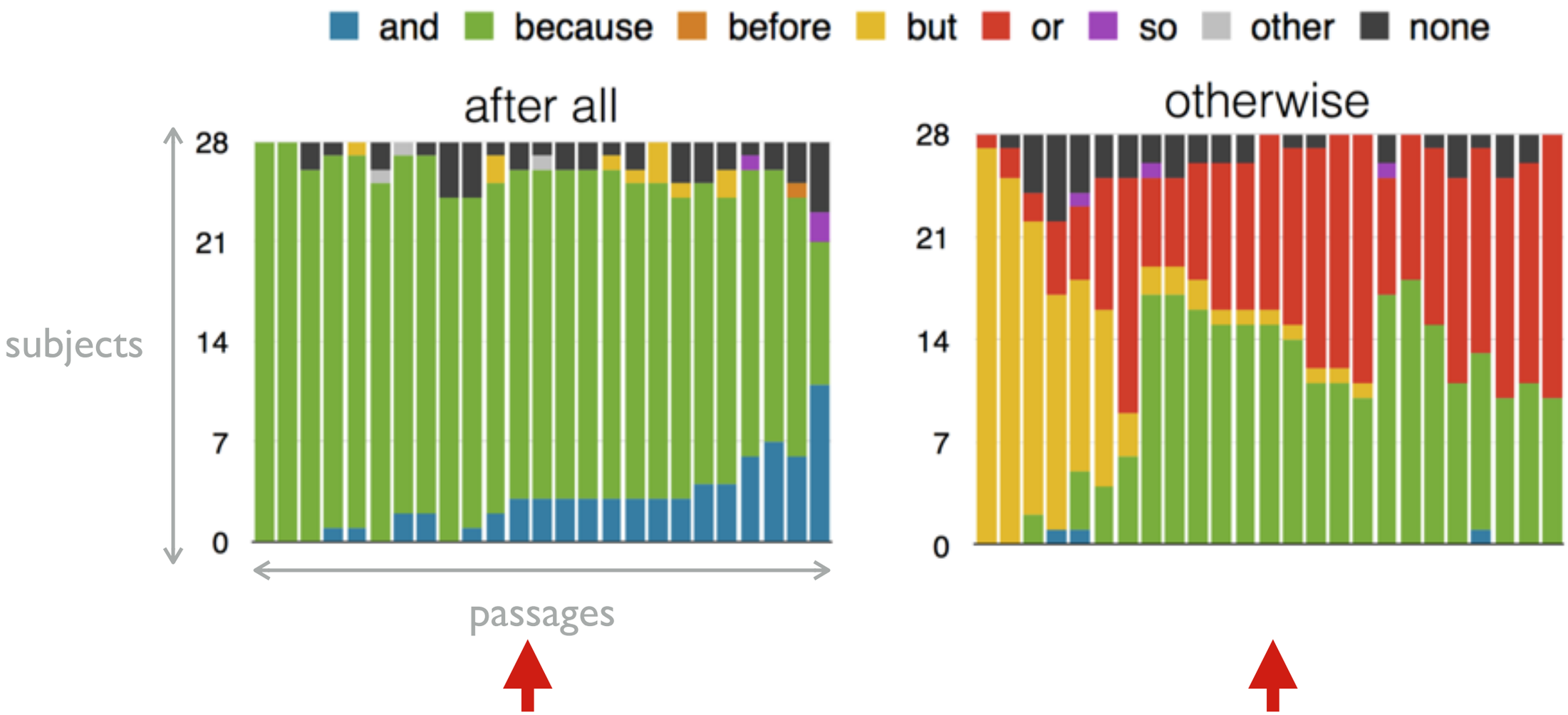


Results: Explicit passages

- Recover same conjunction author used: 57%
- If SO/BUT considered compatible with AND (Knott 1996), calculated match with author: 70%

Results: Implicit passages

- Dataset: 13,916 data points
- For each adverbial, visualize completions for all passages

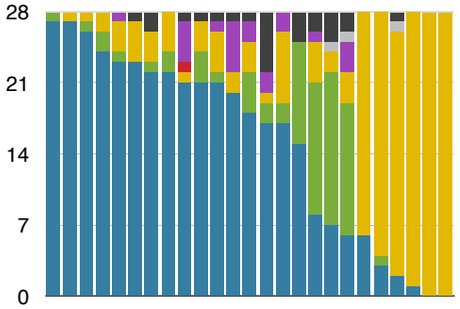


all passages favor 'because'

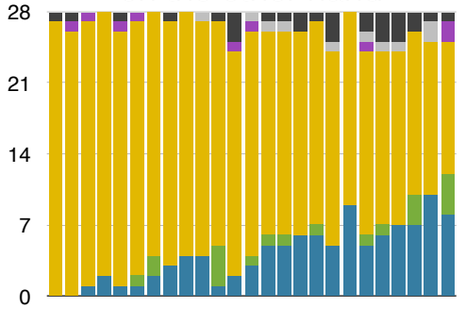
importance of passage context

and because before but or so other none

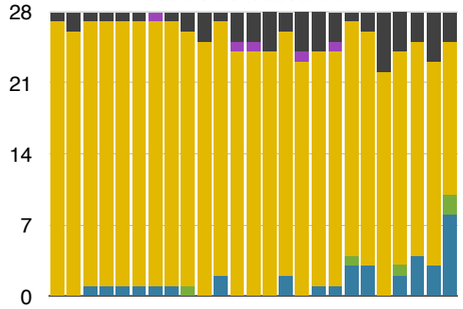
in fact



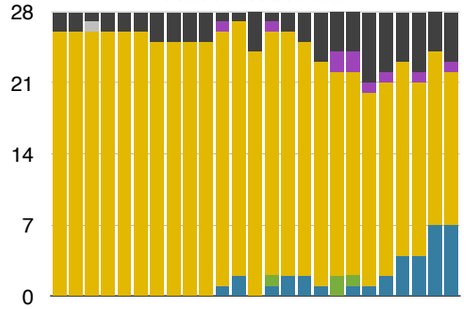
on the other hand



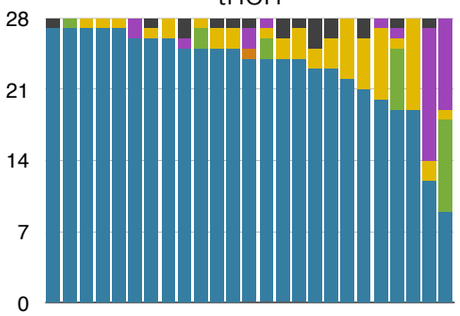
nevertheless



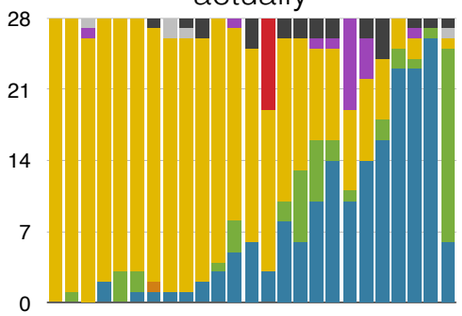
nonetheless



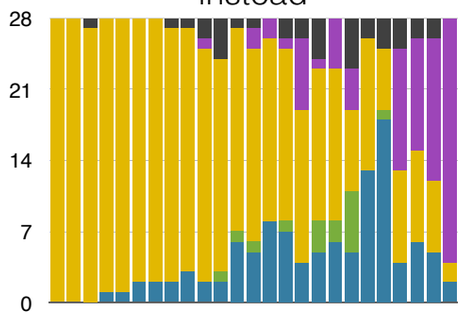
then



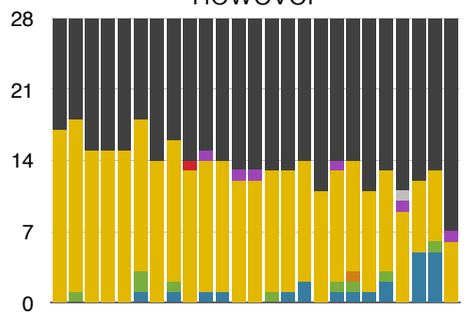
actually



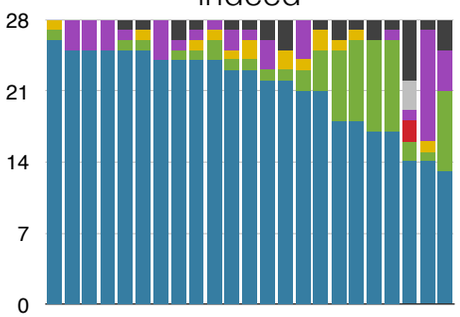
instead



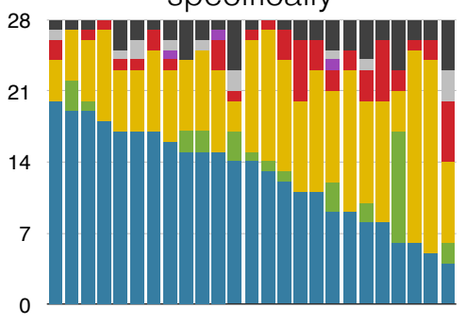
however



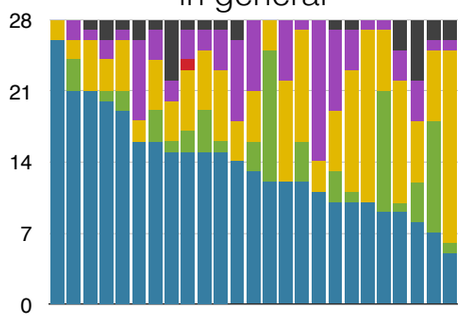
indeed



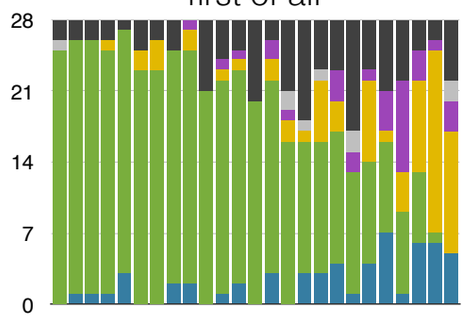
specifically



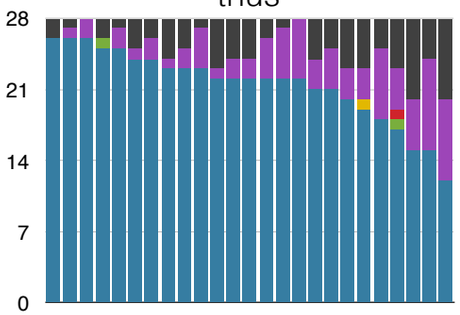
in general



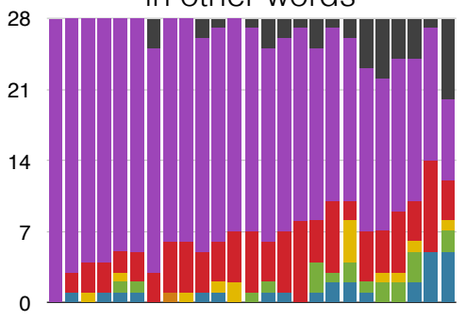
first of all



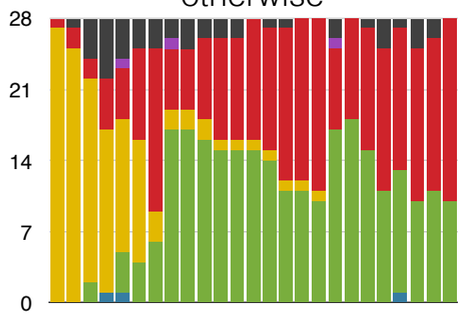
thus



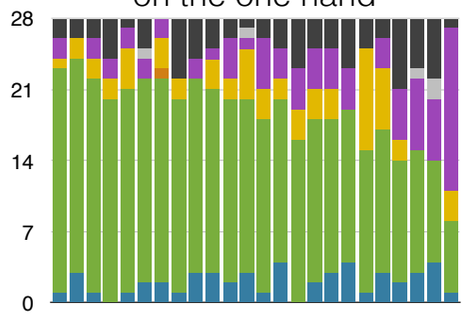
in other words



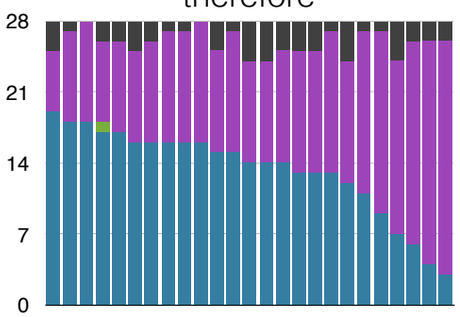
otherwise



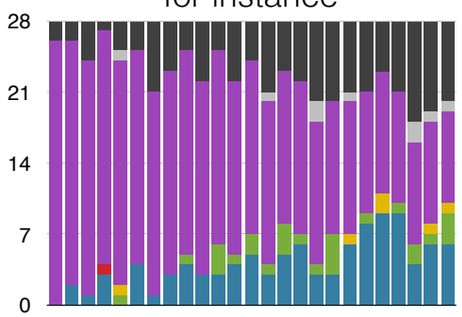
on the one hand



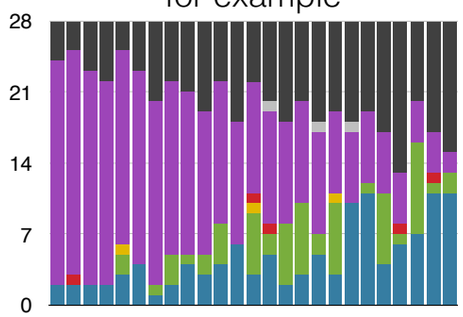
therefore



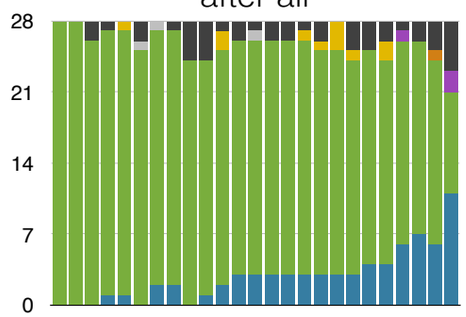
for instance



for example

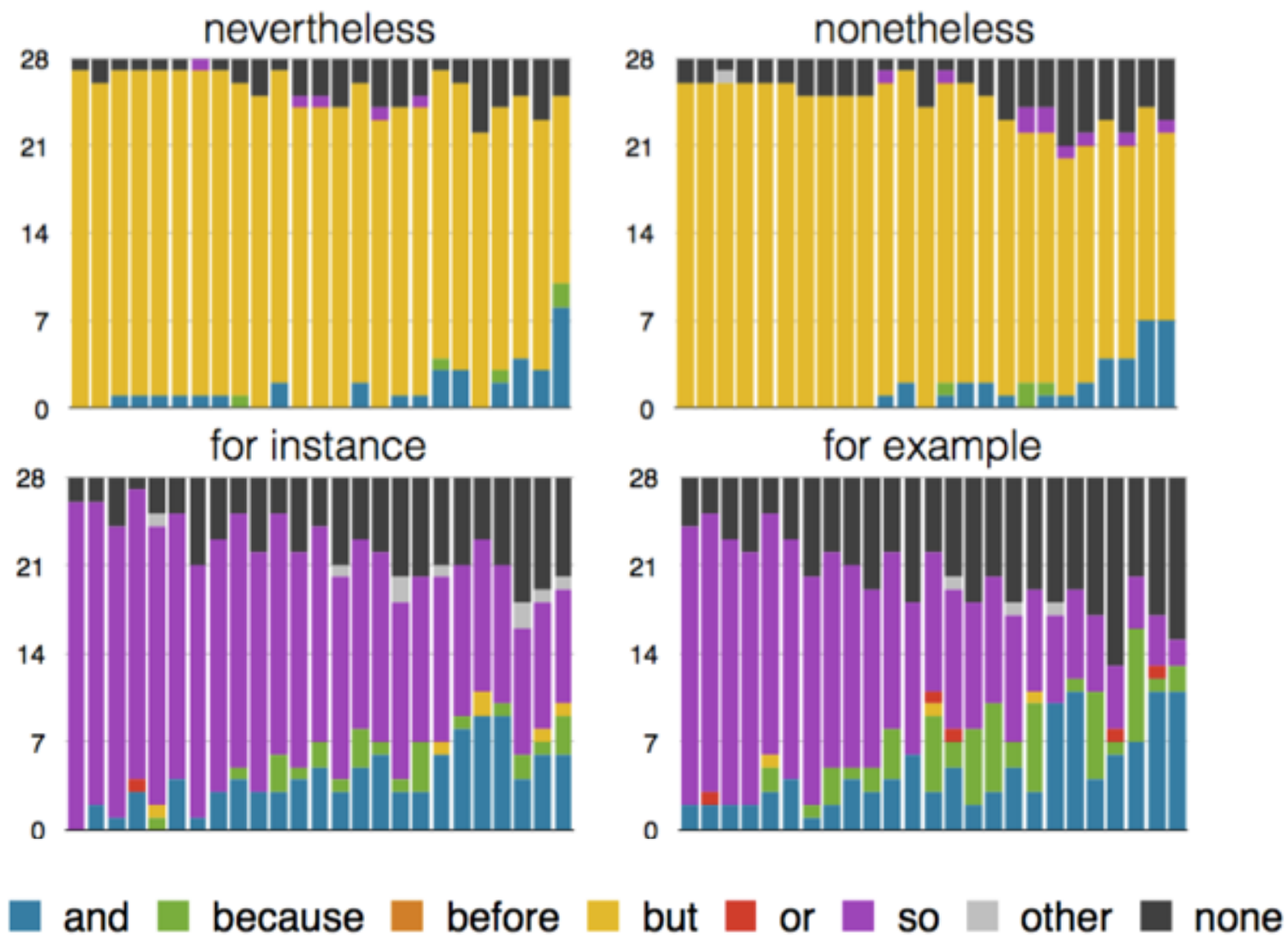


after all



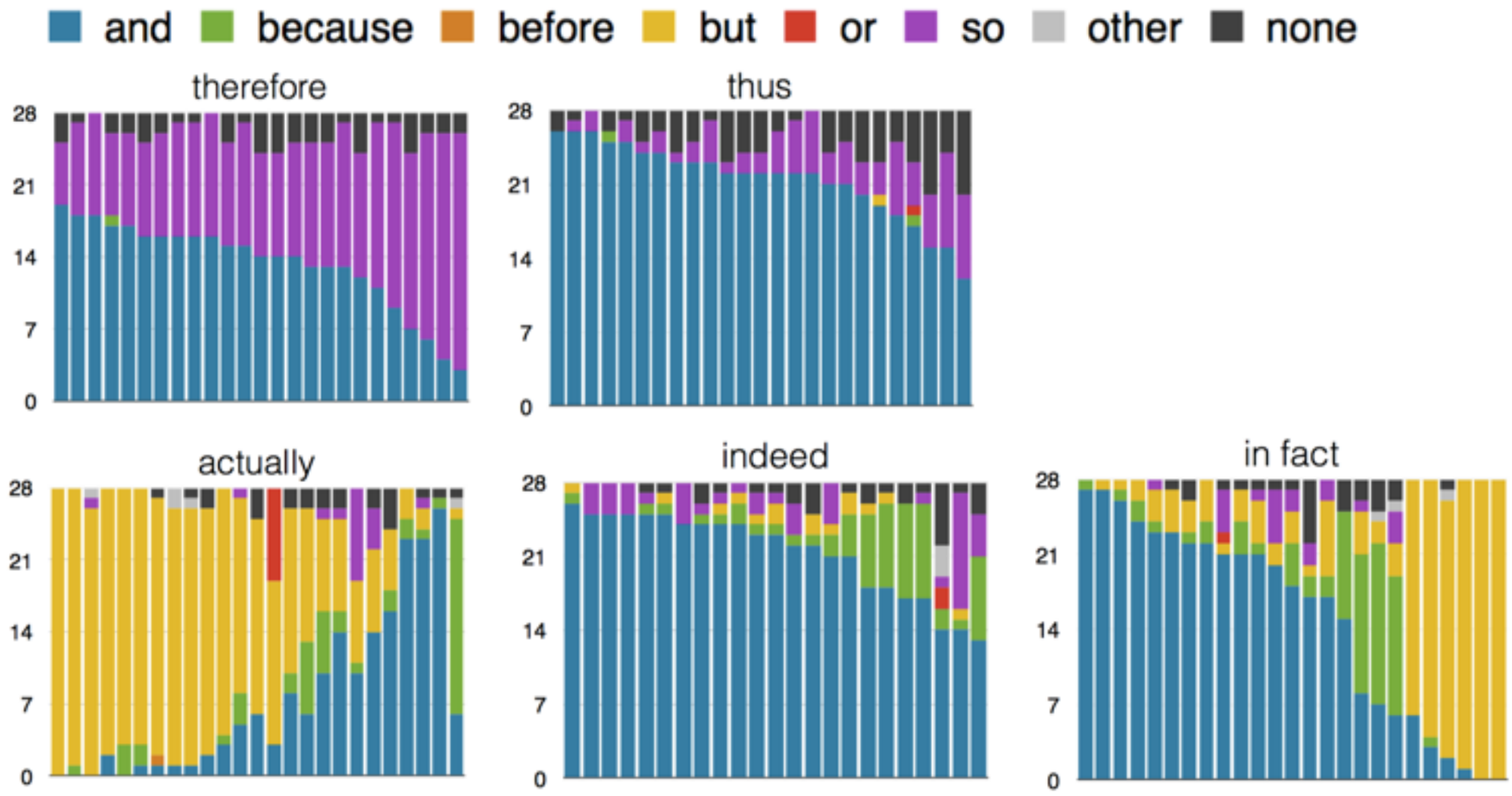
Implicit passages

- On one hand, we see some consistency in semantically related adverbial pairs.



Implicit passages

- But also divergence for near synonyms or for adverbials of a similar type (e.g., modal stance)



- Adverbial itself matters, as does passage content.

Informative disagreement

- Conjunction can disambiguate the attachment point

"Nervous? No, my leg's not shaking," said Griffey, who caused everyone to laugh // _____ indeed his right foot was shaking.

BECAUSE

BUT

Author=BECAUSE
13 Participants=BECAUSE
11 Participants=BUT

Implications for annotation efforts

- Disagreements are not errors, contra prior work on:
  - Corrections for biased/inattentive participants (Hovy et al. 2013, Passonneau & Carpenter 2014)
  - Importance of many annotators for reducing bias (Artstein & Poesio, 2005, 2008)
  - Use of naive annotators to infer discourse relations (Scholman et al., 2016)
- All with same assumption of a single correct answer

Summary so far

- **Multiple connectives:** Establish necessity of entertaining implicit relations when adverbial is present
- **Context sensitivity:** Adverbial alone does not completely predict discourse relation

- **Informative disagreement:** Demonstrate possibility of divergent valid annotations

Unexpected divergence

- Improbable combinations, but perfectly fine

Author=OR
17 Participants=OR
11 Participants=BECAUSE

"The Ravitch camp has had about 25 fund-raisers and has scheduled 20 more. Thirty others are in various stages of planning," Ms. Marcus said. "It has to be highly organized // _____ otherwise it's total chaos," she added.

Author=NONE
6 Participants=OR
19 Participants=SO

Unfortunately, nearly 75,000 acres of tropical forest are converted or deforested every day ____ in other words an area the size of Central Park disappears every 16 minutes.

- Which conjunctions permit substitution and in what contexts?

Categorizing connectives (Knott 1996)

- Division of sense relations into 10 categories:

SEQUENCE	CAUSE
RESULT	RESTATEMENT
TEMPORAL	HYPOTHETICAL
SIMILARITY	DIGRESSION
ADDITIONAL INFORMATION	NEGATIVE POLARITY

- Connectives belong to either a single category (e.g., *because*) or multiple categories (e.g., *since*).

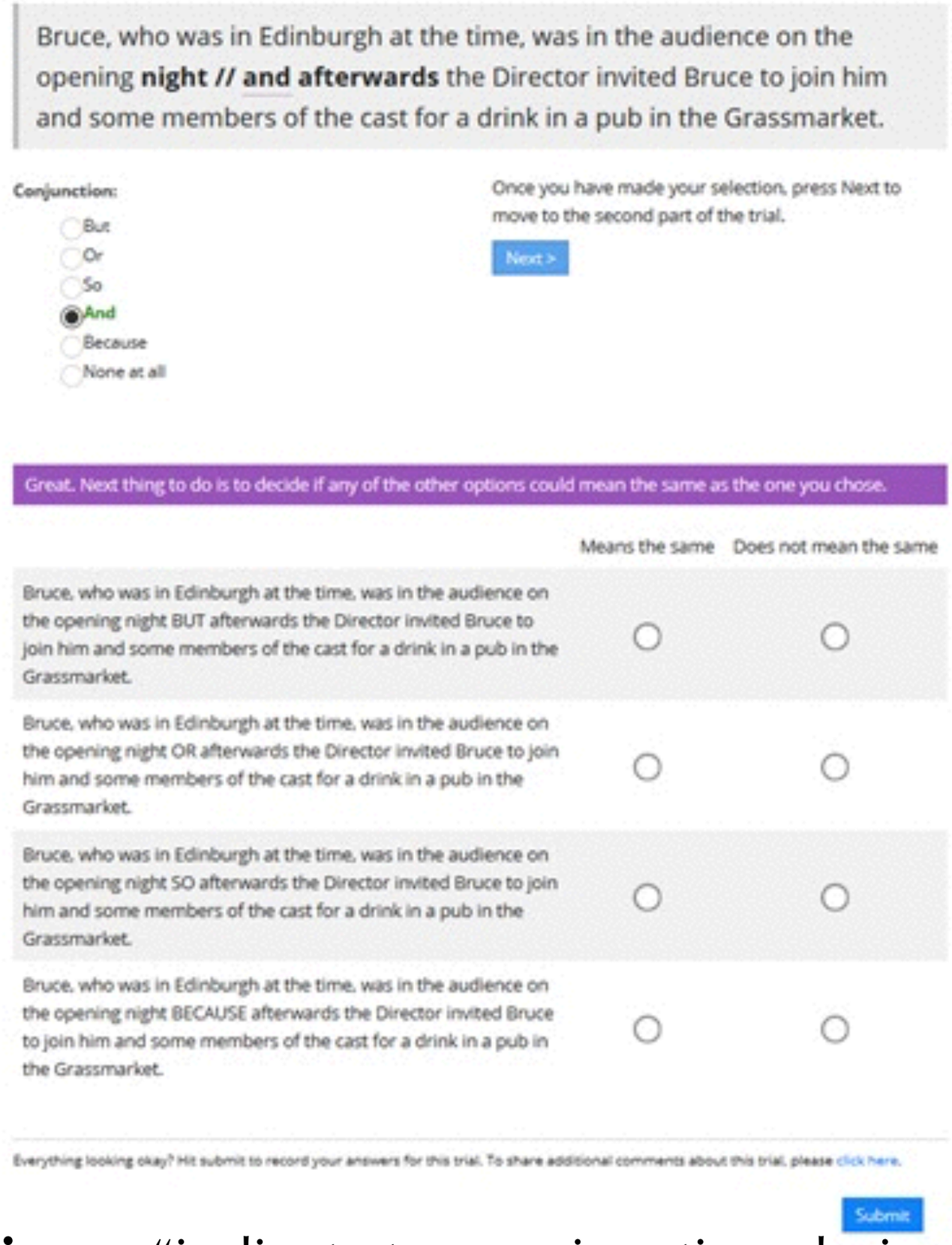
- Substitutability requires that two connectives belong to the same category to ensure that passage retains same meaning.

  → 2 connectives that don't share any sense categories are assumed to be EXCLUSIVE.

- Limits of Knott's approach: constructed examples, introspection

Why would participants differ?

- **Knott:** Substitutability arises if conjunctions belong to same category or if one/both are underspecified for certain features
- **What about connectives that substitute across categories?**
  - Hypothesis #1 ("mutually exclusive meanings"): different interpretations of same passage
  - Hypothesis #2 ("free-for-all"): with discourse adverbials, sense categories don't dictate substitutability, contra Knott's feature-based account
  - Hypothesis #3 ("systematic co-presence"): different conjunctions reflect different simultaneous sources of coherence
- **Method:** Fill-in-the-blank task to elicit **one or more** conjunctions



- **Instructions:** "indicate top conjunction choice and then select any other options that MEAN THE SAME AS THE ONE YOU CHOSE"

Results: exclusivity violations for cross-category conjunctions

- **BECAUSE** (category: CAUSE) ~ **BUT** (category: NEGATIVE POLARITY)
  - **Exclusive meanings or substitutability?** Did previous split between participants signal different meanings or can same interpretation be realized with both conjunctions?

    Yes, I suppose there's a certain element of danger in it, that you can't get around _____ after all, there's a certain amount of danger in living, whatever you do.

  - **Results:** 8+ participants out of 16 endorsed both BECAUSE and BUT

Results: exclusivity violations for cross-category conjunctions

- **BECAUSE** (category: CAUSE) ~ **BUT** (category: NEGATIVE POLARITY)
- **BECAUSE** (CAUSE) ~ **SO** (RESULT)

With a $50 credit in an on-line account, Jordan eagerly logged on. But as he tried to decide which video games to buy, he realized he had a new problem: shipping costs put him over budget. It took him a few weeks to figure out a solution: when he finally made his first purchase in July, he opted for less expensive items - videotapes - _____ then he could afford to pay the shipping costs.

- **Results:** 11+ out of 16 endorsed both BECAUSE and SO

Results: exclusivity violations for cross-category conjunctions

- **BECAUSE** (category: CAUSE) ~ **BUT** (category: NEGATIVE POLARITY)
- **BECAUSE** (CAUSE) ~ **SO** (RESULT)
- **BUT** (NEGATIVE POLARITY) ~ **OR** (SEQUENCE, RESTATEMENT, NEG POL)

Windows is a way of life to some degree _____ more specifically it's Microsoft's way of life, and you'd better like to live the way they tell you to live, or else.

"The Wild Hawaiian" is a Hawaiian rock album _____ more specifically it's an album of songs in the Hawaiian language, against a whiplash of percussion and distorted guitars.

- **Results**: 10+ out of 16 endorsed both BUT and OR

Results: exclusivity violations for cross-category conjunctions

- **BECAUSE** (category: CAUSE) ~ **BUT** (category: NEGATIVE POLARITY)
- **BECAUSE** (CAUSE) ~ **SO** (RESULT)
- **BUT** (NEGATIVE POLARITY) ~ **OR** (SEQUENCE, RESTATEMENT, NEG POL)
- **OR** (multiple, none causal) ~ **BECAUSE** (CAUSE)

Gouges are deep scratches that must be filled as well as colored _____ otherwise they will collect dirt and become permanently discolored.

- **Results:** 12+ out of 16 endorsed both OR and BECAUSE

Results: exclusivity violations for cross-category conjunctions

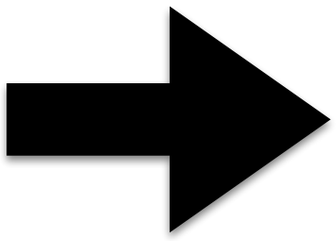
- **BECAUSE** (category: CAUSE) ~ **BUT** (category: NEGATIVE POLARITY)
- **BECAUSE** (CAUSE) ~ **SO** (RESULT)
- **BUT** (NEGATIVE POLARITY) ~ **OR** (SEQUENCE, RESTATEMENT, NEG POL)
- **OR** (multiple, none causal) ~ **BECAUSE** (CAUSE)
- **OR** (multiple, none causal) ~ **SO** (CAUSE)

- None of the above predicted by Knott
- Maybe substitutability isn't the only reason conjunctions alternate while the passage maintains the same meaning

Why would conjunctions substitute?

Cross-category substitution

- Hypothesis #1 (“mutually exclusive meanings”): different interpretations of same passage
- Hypothesis #2 (“free-for-all”): with discourse adverbials, sense categories don’t dictate substitutability, contra Knott’s feature-based account



- Hypothesis #3 (“systematic co-presence”): different conjunctions reflect different simultaneous sources of coherence

Copresent coherence relations

- Multiple coherence relations can be present in a passage.
- Mixture of **explicit** connectives and additional **implicit** relations
  - sometimes derived through pragmatic inference.

e.g., Adverbials that encode 'alternative'

- Adverbial meaning: *otherwise* and *in other words* license OR
- Additional pragmatic inference: Passage content licenses BECAUSE in some cases, SO in others

Gouges are deep scratches that must be filled as well as colored _____ otherwise they will collect dirt and become permanently discolored.

→ *otherwise* encodes 'otherness' (OR)
→ passage requires causal reasoning (BECAUSE)

Unfortunately, nearly 75,000 acres of tropical forest are converted or deforested every day _____ in other words an area the size of Central Park disappears every 16 minutes.

→ *in other words* encodes 'otherness' (OR)
→ reformulation conveys consequence (SO)

Exceptions

- Adverbials that encode alternatives sometimes fail to license ‘or’.
- What licenses which splits?

Different adjacent material

original

Unfortunately, nearly 75,000 acres of tropical forest are converted or deforested every day ____ in other words an area the size of Central Park disappears every 16 minutes.

→ OR/SO

intervening material

Unfortunately, nearly 75,000 acres of tropical forest are converted or deforested every day. *I don't remember where I heard that* _____ in other words an area the size of Central Park disappears every 16 minutes.

→ Prediction: BUT

Different underlying pragmatic logic

argumentation

Proper placement of the testing device is an important issue _____ otherwise the test results will be inaccurate.

→ Prediction: OR/BECAUSE #BUT

"a reason to place the test properly is to avoid inaccuracy"

enumeration

A baked potato, plonked on a side plate with sour cream flecked with chives, is the perfect accompaniment ____ otherwise you could serve a green salad and some good country bread.

→ Prediction: OR/BUT #BECAUSE

"there's more than one option for a side: potato or salad"

exception

Mr. Lurie and Mr. Jarmusch actually catch a shark, a thrashing 10-footer _____ otherwise the action is light.

→ Prediction: BUT #OR/BECAUSE

"shark catching is a special case; generally action is light"

Overall Conclusions

- Discourse **conjunctions** and discourse **adverbials** can both signal coherence relations
- **Crowdsourcing** with many lay subjects reveals mix of systematicity and variation in conjunction completions
- Conjunction, adverbial may signal the **same** relation or **different** relations
- Alternate choice of conjunction for a passage is sometimes predictable, and in some cases may highlight a different aspect of coherence (such as pragmatics)
- **Implicit** vs. **explicit**: not necessarily either/or!
  - Limitation of current approaches

Overall Conclusions

- Linguistics and NLP informing one another!
  - corpora
  - elicitation
  - theory refinement
  - goal of annotation & NLP

Further Details

- Rohde et al.: LAW 2016, IWCS 2017

Thanks

- Christopher N. L. Clark