Exploring Substitutability through Discourse Adverbials & Multiple Judgments

Hannah Rohde, Anna Dickinson, Nathan Schneider,
Annie Louis, & Bonnie Webber

THE UNIVERSITY
of EDINBURGH

GEORGETOWN UNIVERSITY

Inference of connections

▸ 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 any additional pragmatic inference.

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

▸ **Umbrella research question:** Inference of (additional) coherence relations in presence of an explicit connective

This talk

How are different (combinations of) connectives used
to realize particular types of coherence relations?

▶ Builds on Knott's (1996) analysis of substitutability conditions

▶ Stems from unexpected divergence in judgments in our own
(2015, 2016, 2017) large-scale connective elicitations

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.

O or
O so

→ are OR and SO substitutable in this context?

▶ **Goals:** Improve annotation and modelling, shed light on
translation divergence, inform psycholinguistic experimentation

▶ **New study:** Test substitutability by eliciting best-fitting
connective + additional connectives to express same meaning

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 two connectives from same category to ensure that passage retains same meaning

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

Categorizing connectives (Knott 1996)

- Substitutability

  - SYNONYM: *to begin with = to start with*

  - EXCLUSIVE: *first ≠ for one thing*

  - CONTINGENTLY SUBSTITUTABLE: *and, but*

  Bill's a liar. He said he can run a mile in three minutes, [and, but] that's impossible.

  I'm very tired, [and, #but] I don't want to be disturbed.

  Don't be too harsh on Bob. He arrived late, [#and, but] he's usually very punctual.

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

Prior work: Fill-in-the-blank with connectives

→ 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)

Substitutability puzzle

▶ Part of the story is simple

    ▶ Many adverbials prefer one conjunction (e.g., *after all* favors *because*)

    ▶ Many passages yield consistency (28/28 agreement)

    ▶ Semantically similar adverbials behave similarly (e.g., *nonetheless ~ nevertheless*)

▶ Part is not

    ▶ For some passages, competition among conjunctions

> **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

    ▶ Not noise; rather, persistent splits between conjunctions

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

Substitutability study

▶ **Participants:** 16 recruited from earlier task

▶ **Materials**

    ▶ Predicted by Knott to be contingently substitutable
  - AND:BUT (N=24 passages): *in fact, in general, meanwhile (more) specifically*
  - AND:SO (N=22): *for example, therefore, afterwards, then*

Neocons pushed for this war _____ therefore they deserve the blame for its failure or the credit for its success.

    ▶ Predicted to be exclusive
  - BECAUSE:BUT (N=6): *after all, previously, indeed*
  - BECAUSE:OR (N=3): *otherwise, hence*
  - SO:OR (N=2): *in other words*

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.

Substitutabilty study

▶ **Participants:** 16 recruited from earlier task

▶ **Materials to test Knott's claims**

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

Bruce, who was in Edinburgh at the time, was in the audience on the opening **night // and afterwards** the Director invited Bruce to join him and some members of the cast for a drink in a pub in the Grassmarket.

Conjunction:

- ○ But
- ○ Or
- ○ So
- ● And
- ○ Because
- ○ None at all

Once you have made your selection, press Next to move to the second part of the trial.

[Next >]

Great. Next thing to do is to decide if any of the other options could mean the same as the one you chose.

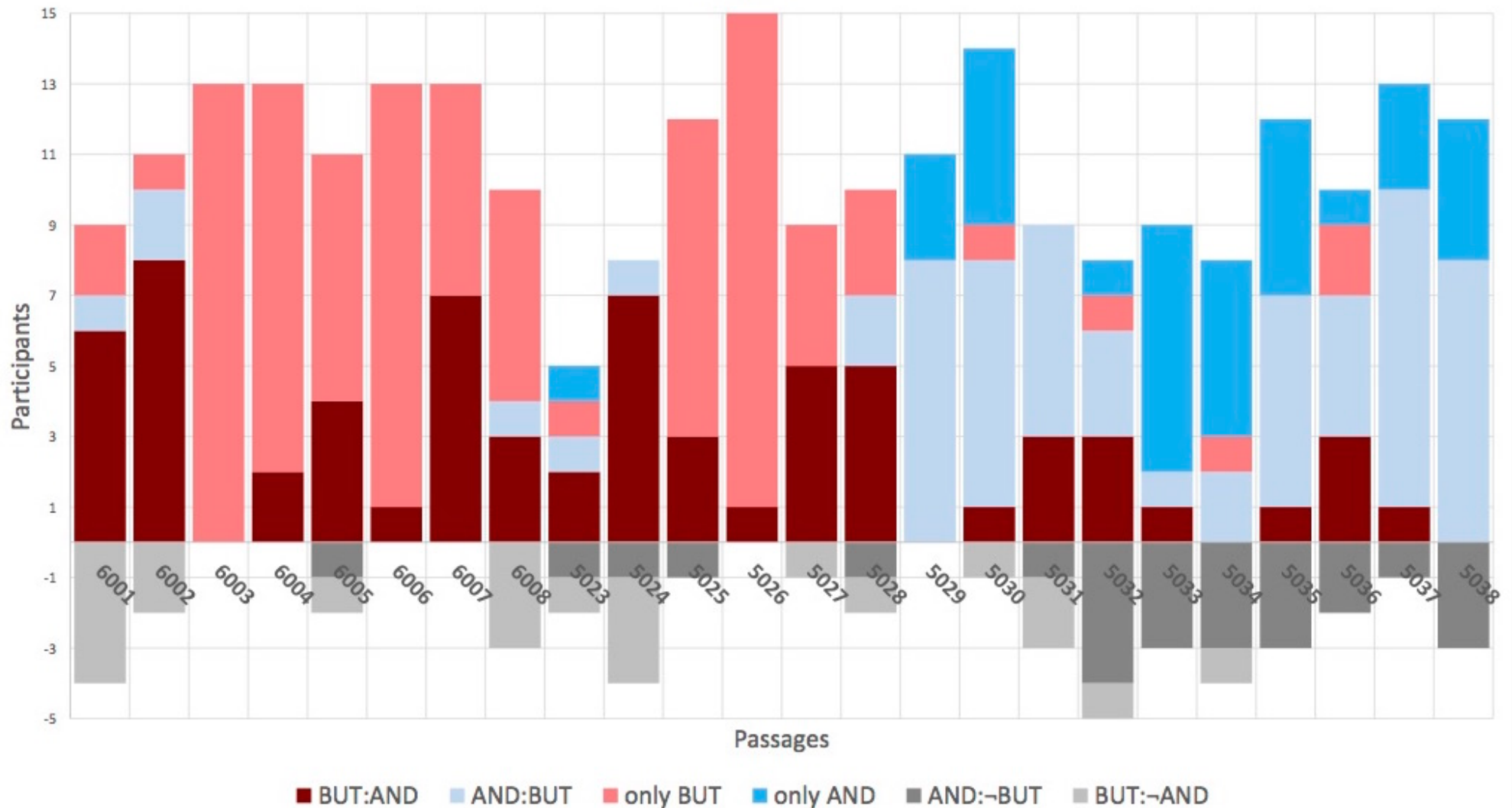|	Means the same	Does not mean the same
Bruce, who was in Edinburgh at the time, was in the audience on the opening night BUT afterwards the Director invited Bruce to join him and some members of the cast for a drink in a pub in the Grassmarket.	○	○
Bruce, who was in Edinburgh at the time, was in the audience on the opening night OR afterwards the Director invited Bruce to join him and some members of the cast for a drink in a pub in the Grassmarket.	○	○
Bruce, who was in Edinburgh at the time, was in the audience on the opening night SO afterwards the Director invited Bruce to join him and some members of the cast for a drink in a pub in the Grassmarket.	○	○
Bruce, who was in Edinburgh at the time, was in the audience on the opening night BECAUSE afterwards the Director invited Bruce to join him and some members of the cast for a drink in a pub in the Grassmarket.	○	○

Everything looking okay? Hit submit to record your answers for this trial. To share additional comments about this trial, please click here.
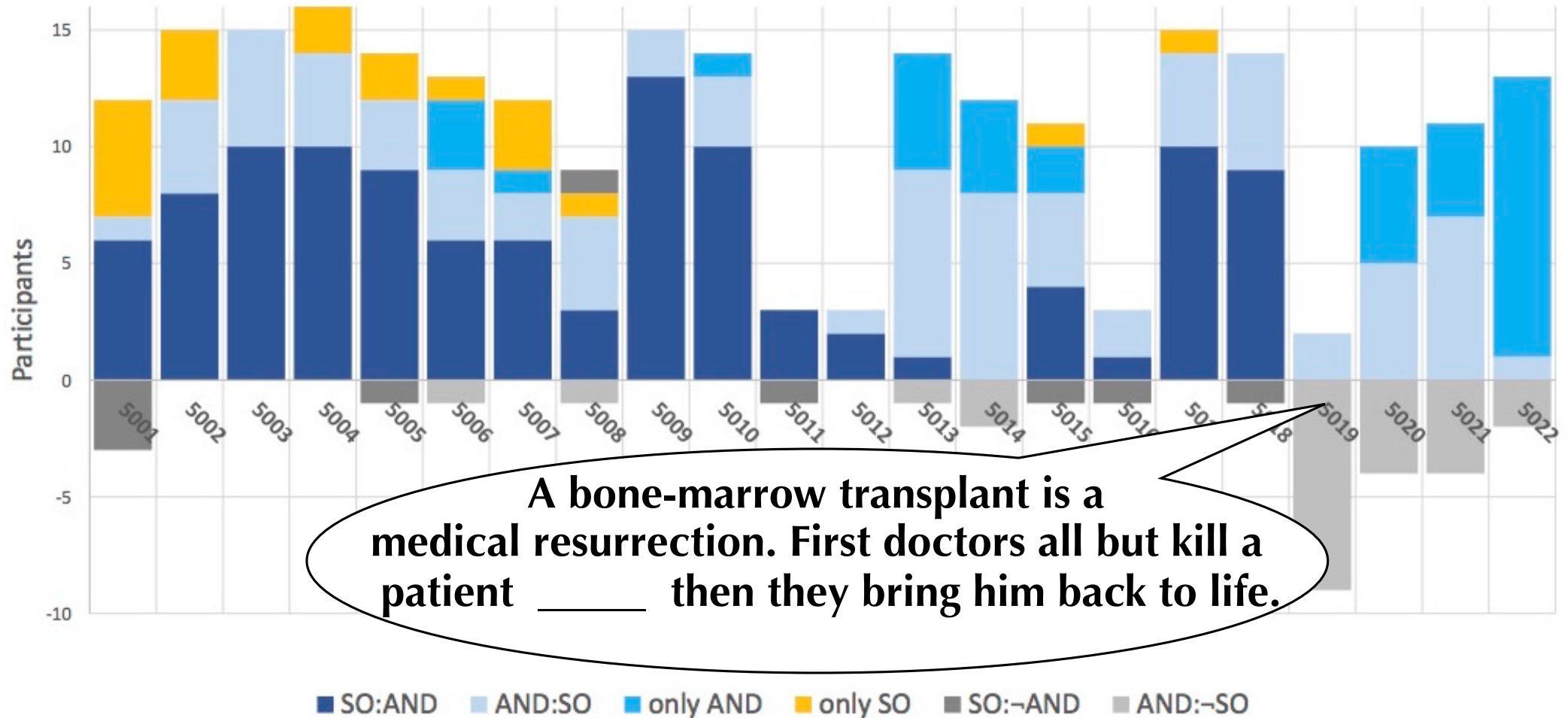
[Submit]

Results: predicted contingent substitutability AND:BUT

→ Some passages favored BUT as best choice (red,pink); others favored AND as best (blues). Most passages yielded patterns fulfilling the two realizations of contingent substitutability (colorful tall bars)

Results: predicted contingent substitutability AND:SO

→ Some passages favored SO as best choice (dark blue/yellow); others favored AND as best (light blues). Largely consistent with contingent substitutability.

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

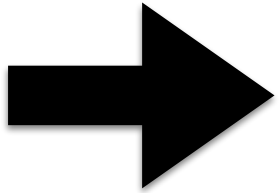# Results: exclusivity violations for cross-category conjunctions

▶ **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

Observed patterns - How & Why

▶ **How to uncover these patterns?** sufficiently large sample sizes and choice to not dismiss disagreement among participants

▶ **Why do these patterns emerge?**

   ▶ To the extent we have an account, no uniform explanation holds for all passages for all adverbials.

   ▶ Different explanations for different adverbials — patterns of substitutability can crosscut semantics and function.

   ▶ Certain patterns reflect BOTH the coherence relation signalled by the discourse adverbial and an additional coherence relation 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)

Summary

▸ **This study**

    ▸ Do participants endorse a conjunction even when an explicit adverbial is already present? **Yes, often**

    ▸ Do response patterns confirm Knott's predicted contingent substitutability pairs? **Yes (AND:BUT, AND:SO)**

    ▸ Can Knott's exclusivity pairs be violated? **Yes BECAUSE:BUT, BECAUSE:SO, BUT:OR, OR:BECAUSE, OR:SO**

▸ **What does substitutability tell us:**

    ▸ Sometimes conjunctions **share sense categories** (Knott 96)

    ▸ In other cases, alternation between two conjunctions reflects **different simultaneous sources of coherence**

Thanks!

Is OR always licensed by 'otherness'?

It seems to me that writing short stories is a discipline in conciseness, like an implosion. The novel instead is like an explosion. The first requires fitting an idea into the smallest space and is technically more difficult. The difficulty in the latter is how to stop expanding your runaway characters, situations, interactions unless you are comfortable with the idea of writing a series of novels. I do not know if I am making enough sense, _____ in other words , I find writing short stories easier to write because they are finite.

→ dispreference for OR (originally BUT)

Different attachment points

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

Author='because'
13 Participants='because'
11 Participants='but'

Exclusivity

Bob set about cleaning the house.

$$
\left\{
\begin{array}{l}
\textit{To start with,} \\
\checkmark \textit{ To begin with,} \\
\checkmark \textit{ First,} \\
\# \textit{ For one thing,} \\
\# \textit{ Furthermore,} \\
\# \textit{ And,}
\end{array}
\right\}
$$

he swept the floors; next he washed them; and lastly, he tidied the cupboards.

Television is bad for us.

$$
\left\{
\begin{array}{l}
\textit{For one thing,} \\
\checkmark \textit{ To begin with,} \\
\checkmark \textit{ Firstly,} \\
\# \textit{ Moreover} \\
\# \textit{ And} \\
\# \textit{ First}
\end{array}
\right\}
$$

it kills creativity; in addition it promotes an unhealthy kind of 'crowd mentality'.