Do discourse connectives influence likelihood and form of reference in the absence of verb semantics?

Jennifer Spenader

June 2013
Discourse connectives

1. Mary admires Betty *because* she publishes so much.
Discourse connectives

(1) Mary admires Betty because she publishes so much.
(2) Mary admires Betty so she cites her work in all her papers.
Discourse connectives

(1) Mary admires Betty because she publishes so much.
(2) Mary admires Betty so she cites her work in all her papers.
(3) Mary admires Betty but she would never admit it.
Discourse connectives

(1) Mary admires Betty because she publishes so much.
(2) Mary admires Betty so she cites her work in all her papers.
(3) Mary admires Betty but she would never admit it.
What is the nature of discourse connectives?
What is the nature of discourse connectives?

- Are they simply convenient, handy explicit markers of coherence relation type?
What is the nature of discourse connectives?

- Are they simply convenient, handy explicit markers of coherence relation type?
- Or do they have focusing functions?
  - Is it the connective or the coherence relationship that has a focusing function?
Verb semantic biases

(4) Mary  admires Betty  because she . . . 
Experiencer .  Stimulus

(5) John  kicked the  ball to Bill  and he . . . 
Source .  Theme . .  Goal

- Certain semantic verb groups show strong biases for continued references to be to a certain argument (implicit causality)
- Experience-Stimulus verbs (ES): tend to have continuations referring to the Stimulus argument (with because)
- Transfer verbs (send, mail, throw) tend to have continuations to goal arguments
Connectives interact with verb semantic biases

(6) Mary appreciates Betty because she . . .
    Experiencer . Stimulus

(7) Mary appreciates Betty so she . . .
    Experiencer . Stimulus

- Stimulus continuations are preferred after *because* but
  Experiencer continuations are preferred after *so* (Stevenson et al. (1994) and Fukumura & van Gompel (2010))
Reach out and poke someone.
(8) Mary admires Betty because/so she . . .
Experiencer . Stimulus

- Verbs have baseline preferences
Verbs have baseline preferences (but this is maybe an incorrect perception remaining from early work)

Connectives poke them further

- **because** marks effect-cause
  - the cause of the admiring event is the Stimulus
- **so** marks cause-effect
  - the effect of the admiring event is on the Experiencer

Can we reduce connective effects simply to interaction with causal meaning in the verb?
What about with other verb groups e.g. transfer verbs?

- Stevenson et al. (1994) found no effect of connective choice on source-goal verb continuations
  - Consistent with idea that the connective interacts with the verb meaning to affect biases
  - If there is no relationship between verb meaning and connective, no effect
What about with other verb groups e.g. transfer verbs?

- Stevenson et al. (1994) found no effect of connective choice on source-goal verb continuations
  - Consistent with idea that the connective interacts with the verb meaning to affect biases
  - If there is no relationship between verb meaning and connective, no effect
- In continuation experiments *without* connectives: differences depending on coherence relationship type (causal vs. endpoint)
  - (Arnold 2001 and Rohde et al. 2006)
- Can we test IC biases without coherence?
Don’t pee in the pool. The pool is for members only.
Don’t pee in the pool. The pool is for members only.

- Even with simple juxtaposition, we impose a relation
- Rohde et al.: Classify continuations by coherence relation type and the pattern with cases when there was an explicit marker
  - it wouldn’t help to use a period or and
  - continuations would have to first be classified
(9) John admires Bill because he_{Bill} can juggle.

(10) John admires Bill because he_{John} is easily impressed by card tricks.

- (9) and (10) have different truth-conditions
Activation / Accessibility vs. Likelihood Predictability

(9) John admires Bill because he$_{Bill}$ can juggle.

(10) John admires Bill because he$_{John}$ is easily impressed by card tricks.

- (9) and (10) have different truth-conditions
- For the speaker, choice between a name or pronoun both give truth-conditionally equivalent statements
John admires Bill because he$_{Bill}$ can juggle.

John admires Bill because he$_{John}$ is easily impressed by card tricks.

- (9) and (10) have different truth-conditions
- For the speaker, choice between a name or pronoun both give truth-conditionally equivalent statements
- **Referential form choice** is more of a real choice than continuation choice
  - Outside of experiments that is...
Does predictability = accessibility = reduced forms?

- **Evidence for predictability = accessibility = reduced forms**
  - Arnold (2001): With transfer verbs, the preferred Goal continuations were expressed with pronouns

- **Evidence against:**
  - Fukumura & van Gompel (2010): arguments consistent with continuation preference were not more likely to be referred to with pronouns
  - Instead, continuations with subjects antecedents tended to be pronouns
  - Concluded: Verb-based expectations different from referent accessibility form choices
Two research questions:

- Do connectives independently influence the likelihood of continuing a discourse with a subject or an object in the absence of verb semantics?
Two research questions:

- Do connectives independently influence the likelihood of continuing a discourse with a subject or an object in the absence of verb semantics?
- Do arguments consistent with continuation biases have a greater chance of being referred to with a pronoun?
Can Nonsense words factor out verb semantics?

(11) Marcie knoffied Jared *because/-but/-so* . . .

- If subjects have no biases to go on, entirely dependent on connective?
4 Experiments: Common method

- 24 items, participants
- 8 examples each of three connectives, because, so and but
  - Test because and so because they have been studied most
  - Test but because it is believed to reverse expectations and perhaps focusing
- Subjects and objects different genders balanced across connectives
  - Makes pronouns unambiguous, e.g.
    
    (12) John spefted Mary but she ...

- Participants (paid, AMT) asked to complete sentences in a natural way without humor.
- Threw out non-target responses (e.g. ellipsis, plural pronouns, other references, etc.)
Experiment 1: Nonsense verbs
45 participants
Exp 1: Results: type of continuation

Table: Exp1: Percentage subject continuations by connective

<table>
<thead>
<tr>
<th>connective</th>
<th>percentage subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>because</td>
<td>30% (n = 93)</td>
</tr>
<tr>
<td>but</td>
<td>25% (n = 75)</td>
</tr>
<tr>
<td>so</td>
<td>16% (n = 46)</td>
</tr>
</tbody>
</table>

- Object continuations dominate (contrary earlier corpus data but typical for continuation experiments)
- *Because* has significantly more subject continuations than *so*
- Not unexpected, but effect is small
Exp 1: Role of continuation

Table: Percentage Pronouns for continuations

<table>
<thead>
<tr>
<th>connective</th>
<th>subjects (n=89)</th>
<th>objects (n=183)</th>
<th>but (n=39)</th>
<th>objects (n=160)</th>
<th>so (n=46)</th>
<th>objects (n=191)</th>
</tr>
</thead>
<tbody>
<tr>
<td>because</td>
<td>95%</td>
<td>85%</td>
<td>74%</td>
<td></td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>but</td>
<td>95%</td>
<td></td>
<td></td>
<td>74%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>so</td>
<td>91%</td>
<td></td>
<td></td>
<td>80%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Pronouns dominate
- Significant effect of role: subjects more likely to be pronouns
- Significant effect of connective: because more likely to have a pronoun than but
- No significant interaction
Summary Results Experiment 1

- Connectives do contribute independent continuation biases
  - Significant higher chance of subject continuation with *because*
- Connectives also contribute referential form biases
  - *because* more likely to be followed by a pronoun than *but*
  - which has a comparatively higher rate of name continuations
Problems with the experiment

- We don’t get close to balanced data
- We get very few names, hard to make form comparisons
- Solution: make task to continue with reference to marked argument

(13) *Julie* lorged Eli because ...
Experiment 2
24 participants
Same nonsense words
Same 3 connectives
Half marked for subject continuations and half for object continuations
Experiment 2 Results: Nonsense words with marked arguments

- People don’t always follow instructions: 248 objects, 219 subjects
Experiment 2 Results: Nonsense words with marked arguments

- People don’t always follow instructions: 248 objects, 219 subjects
- Names and Pronouns similar: 206 pronouns, 251 names
Experiment 2 Results: Nonsense words with marked arguments

- People don’t always follow instructions: 248 objects, 219 subjects
- Names and Pronouns similar: 206 pronouns, 251 names
- **But:** Participants choose either to almost always answer with a pronoun or with a name
Experiment 2 Results: Nonsense words with marked arguments

- People don’t always follow instructions: 248 objects, 219 subjects
- Names and Pronouns similar: 206 pronouns, 251 names
- **But:** Participants choose either to almost always answer with a pronoun or with a name
nothing to see here ....
How do subjects interpret the nonsense words?

(14) *Albert* leffed Rachel but he still cared about her.

▶ left?
How do subjects interpret the nonsense words?

(14) *Albert* leffed Rachel but he still cared about her.

▶ left?

(15) *Julie* orgued Eli because she thought he was cute.

▶ ogled?
How do subjects interpret the nonsense words?

(14)  *Albert* leffed Rachel but he still cared about her.

▶  left?

(15)  *Julie* orgued Eli because she thought he was cute.

▶  ogled?

▶  What words could we use that are completely uninterpretable?
Experiment 3: Nonwords, with temporal subordinate clauses

(16) Theresa kqbped Ronald before the train to the airport was about to leave, because . . .

(17) Harry xvdzed Jennifer while the debate continued in the classroom, because . . .

- Changes: Temporal subordinate clause (follows Fukumura & van Gompel 2010)
- Non-words
Results Experiment 3: Nonwords, with temporal subordinate clauses

<table>
<thead>
<tr>
<th>connective</th>
<th>subjects</th>
<th>objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>because</td>
<td>67% (n=107)</td>
<td>33% (n=52)</td>
</tr>
<tr>
<td>but</td>
<td>22% (n=28)</td>
<td>78% (n=96)</td>
</tr>
<tr>
<td>so</td>
<td>26% (n=34)</td>
<td>74% (n=92)</td>
</tr>
</tbody>
</table>

- Surprise! *because* significantly more likely to be followed by a subject than *but* or *so*
- Cf. Exp1 where subject continuations with *because* was 30%
- Almost all responses were pronouns :(

- Experiment 4
  - 24 participants
  - Same nonwords as Exp3
  - Same 3 connectives
  - Half marked for subject continuations and half for object continuations
Experiment 4: Nonwords, with temporal subordinate clauses, with marked arguments

<table>
<thead>
<tr>
<th>connective</th>
<th>names</th>
<th>pronouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>because</td>
<td>25% (n=44)</td>
<td>75% (n=130)</td>
</tr>
<tr>
<td>but</td>
<td>31% (n=53)</td>
<td>69% (n=116)</td>
</tr>
<tr>
<td>so</td>
<td>24% (n=40)</td>
<td>76% (n=126)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>role</th>
<th>names</th>
<th>pronouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>objects</td>
<td>34% (n=88)</td>
<td>66% (n=171)</td>
</tr>
<tr>
<td>subjects</td>
<td>20% (n=49)</td>
<td>88% (n=201)</td>
</tr>
</tbody>
</table>

- No effect of connective
- Grammatical role of antecedent significant: Referents with subject antecedents more likely to be pronouns
- 10 subjects varied their choices
Exp 4 Nonwords, Marked antecedents

Subject continuations significantly more likely to be a pronoun

<table>
<thead>
<tr>
<th></th>
<th>pronouns</th>
<th>names</th>
</tr>
</thead>
<tbody>
<tr>
<td>subjects</td>
<td>80%</td>
<td>20%</td>
</tr>
<tr>
<td>objects</td>
<td>80%</td>
<td>20%</td>
</tr>
</tbody>
</table>
Back to the research questions:

- Do connectives influence the likelihood of continuing a discourse with a subject or an object in the absence of verb semantics

- Yes, because leads to significantly more continuations with subject antecedents

- Strong preference for object continuations remains

- Do arguments consistent with connective continuation biases have a greater chance of being referred to with a pronoun?

- No

- Having a subject antecedent means a referent has a greater chance of being pronominalized
Back to the research questions:

- Do connectives influence the likelihood of continuing a discourse with a subject or an object in the absence of verb semantics
  - Yes. because leads to significantly more continuations with subject antecedents
  - Strong preference for object continuations remains
- Do arguments consistent with connective continuation biases have a greater chance of being referred to with a pronoun?
Back to the research questions:

- Do connectives influence the likelihood of continuing a discourse with a subject or an object in the absence of verb semantics
  - Yes. because leads to significantly more continuations with subject antecedents
  - Strong preference for object continuations remains
- Do arguments consistent with connective continuation biases have a greater chance of being referred to with a pronoun?
  - No.
  - Having a subject antecedent means a referent has a greater chance of being pronominalized
Observations and Directions for Future Research

- Troubling: why do continuation tasks lead overwhelmingly to object biases in contrast to corpus data?
  - Even if it disappears if you contrôlé for coherence, why don’t we see this reflected in production?
Observations and Directions for Future Research

- Troubling: why do continuation tasks lead overwhelmingly to object biases in contrast to corpus data?
  - Even if it disappears if you control for coherence, why don’t we see this reflected in production?
- Nonsense words and non-words lead to similar results for *so* and *but*
- Big difference with *because*, but due to temporal phrase?
- Continuations no longer related to neighboring lexical items
  - Instead, they seem to now be related to the subordinate clause
  - Was this not a result in Fukumura & Van Gompel?
Observations and Directions for Future Research

- Troubling: why do continuation tasks lead overwhelmingly to object biases in contrast to corpus data?
  - Even if it disappears if you control for coherence, why don’t we see this reflected in production?
- Nonsense words and non-words lead to similar results for *so* and *but*
- Big difference with *because*, but due to temporal phrase?
- Continuations no longer related to neighboring lexical items
  - Instead, they seem to now be related to the subordinate clause
  - Was this not a results in Fukumura & Van Gompel?
- Why do I get only names or only pronouns in many continuations (and can I prevent it?)
- Is it the connective, or the coherence relation contributes continuation biases?
  - Preliminary experiment using connective-paraphrases: with paraphrases biases disappear.
Observations and Directions for Future Research

▶ Troubling: why do continuation tasks lead overwhelmingly to object biases in contrast to corpus data?
  ▶ Even if it disappears if you control for coherence, why don’t we see this reflected in production?
▶ Nonsense words and non-words lead to similar results for so and but
▶ Big difference with because, but due to temporal phrase?
▶ Continuations no longer related to neighboring lexical items
  ▶ Instead, they seem to now be related to the subordinate clause
  ▶ Was this not a result in Fukumura & Van Gompel?
▶ Why do I get only names or only pronouns in many continuations (and can I prevent it?)
▶ Is it the connective, or the coherence relation contributes continuation biases?
  ▶ Preliminary experiment using connective-paraphrases: with paraphrases biases disappear.
▶ Can we learn something with more experimentation with non-words, or is it just fun?