Discourse connectives vs. paraphrases: what influences likelihood and form of reference in the absence of verb semantics?

Jennifer Spenader (j.spenader@ai.rug.nl)
Institute for Artificial Intelligence
University of Groningen

Likelihood of reference

Verb semantics influences the likelihood of subsequent reference to verbal arguments. Consider the verb like (1).

(1) a. John_{Experiencer} likes Bill_{Stimulus} because he ...(preferred continuation: he = Bill)
   b. John_{Experiencer} likes Bill_{Stimulus} so he (preferred continuation: he = John)

The subject of like has the semantic role of Experiencer, while the object is the Stimulus. Like and similar verbs belong to the class of Experiencer-Stimulus verbs (ES). In continuation experiments, where participants are asked to write continuations of sentences/clauses after certain verbs followed by a connective and (often) an ambiguous pronoun, as in (1), participants show a strong and robust tendency to continue with a reference to the Stimulus argument: in (1-a), he = Bill. But connectives also influence continuation likelihood. Stimulus continuations are preferred after because but Experiencer continuations are preferred after so, see (1-b).

ES verbs and Stimulus-Experiencer verbs (SE) like amaze, that have Experiencer objects and Stimulus subjects are all State verbs. Stevenson et al. (1994; hereafter Steve&co) and Fukumura & von Gompel (2010; hereafter F&vG) found semantic role continuation preferences hold for State verbs with consistent effects for connectives, regardless of the grammatical role of the Experiencer or Stimulus. However, not all verb groups that show consistent continuations biases find an interaction with connectives. Steve&co found choice did not influence Transfer verbs(e.g. give) or Action verbs (e.g. hit). Transfer verbs are consistently biased to Goal/object continuations, and Action verbs to Patient/object continuations. An explanation may be that connective-verb semantic interactions only occur when connective meaning relates to verb meaning. For State verbs the Stimulus is the cause, so it is not surprising that after because most continuations refer to this semantic role. The opposite is the case for so (used with Cause-so-Effect sequences). The connection between Agent/Patient roles or Source/Goal roles and causal connectives may be more indirect with Action and Transfer verbs.

Nonsense verbs can be used to study the contribution of connectives separate from preferences stemming from verb biases. Given this background, our first research question is:

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

The question above is framed in terms of connectives. But connectives are simply explicit evidence of a more primary coherence relation between two clauses. Recent continuation studies examining verb semantics and connective preferences (e.g. Rohde et al. 2006; 2007) have shown that biases are based on coherence relations, not connectives, and connective function has to be disambiguated to clearly identify biases. This leads to our second research question:

- Is it the coherence relation that is associated with certain focussing properties, or the connective itself?

At issue is an additional research question: Do expectations interact with referential form choice? For verb semantic biases previous work disagrees as to whether or not continuation likelihood influences referent accessibility. Recall that Transfer verbs are biased to Goal continuations. In a continuation experiment with Transfer verbs without connectives, Arnold (2001) found that speakers used pronouns more than proper nouns to refer to Goal continuations, suggesting that a strong likelihood of reference correlates with referential forms associated with higher levels of activation (e.g. pronouns, rather than names). However, F&vG found the opposite effect with State verbs: arguments consistent with continuation preferences were not more likely to be referred to with pronouns. F&vG conclude that verb-based expectations are separate from referent accessibility. They did however find that grammatical role affected form choice: continuations with subject antecedents tended to be pronouns. Based on these results our final research question is:

- Do arguments consistent with continuation biases have a greater chance to be referred to with a pronoun?

To answer these questions, we did a written continuation experiment with nonsense verbs. We had three Connective Types: because, so and but. We used because and so because these have been extensively used in previous continuation tasks and this allows useful comparisons. We also included but because a previous experiment (Ehrlich 1981) argued that it has the ability to reverse continuation preferences found with because. However, it is not clear if this characterization is accurate: but tends to mark cause-effect sequences, like so, but with but the effect is unexpected (e.g. It’s raining but I didn’t get wet. Perhaps but patterns with so because of similar causal structure. To check whether it is the coherence relation or the connective associated with biases, we also used three unambiguous paraphrases for each of the three connectives: and as consequence of the fact that, and as a result and and contrary to expectation.
Method

We created 24 simple transitive sentence fragments with nonsense verbs. For each verb we created six versions each with because, so and but, and the three paraphrases e.g.

(2) Marcie knocked Jared because/so/buteland as consequence of the fact that, and as a result-and contrary to expectation....

Half the stimuli used an unambiguous female name as subject and an unambiguously male name as object. The other half reversed this pattern. Connective Type (because, so or but) was a within subjects factor, and Connective Form (connective vs. paraphrase) was a between subjects factor. We created six balanced list. 96 native speakers of English recruited via Amazon’s Mechanical Turk service completed the task online. Participants were asked to write a natural continuation to each sentence.

Only continuations beginning with reference to the subject or object were analyzed. We excluded cases where neither character was referred to, where both were referred to with e.g. they, and numerous other errors or non-target responses (so interpreted as Purpose or adverb). Gender identified the antecedent for pronoun continuations.

Results

Subject or object continuations? We used mixed-effect logistic models for analysis. The maximally optimal model for predicting argument continuation preferences showed a significant effect of Connective Form (connective vs. paraphrase), and significant interactions between Connective Type and Connective Form, with participant and items as random effects. There was no independent effect for Connective Type.

There was an overall preference for object continuations, but traditional connectives lead to significantly fewer object continuations (78.7%) compared to paraphrases (96%). Further, in the connective condition, because had significantly fewer object continuations (69%) than but (84%) or so (83%). Strikingly, there were no differences between Connective Type in the paraphrase condition.

Names or pronouns? We again used mixed-effect models to examine the referential form choice. The maximally optimal model had Continuation Role (subject or object) and Connective Type and a random effect of participant. Connective Form had no significant effect on continuation form. There were no significant interactions.

In general, pronouns were much more frequent than names (76%). Further, 91% of subject continuations were pronouns, while only 73% of object continuations were pronouns. Examining Connective Type, because was significantly more likely to be followed by a pronoun than but or so. Note that this is even the case in the object continuations, where because had 77% pronouns while but and so had 71% and 72% respectively.

General Discussion

In answer to our first research question: connectives do contribute independent continuation biases separate from verb semantics. Because had significantly more subject continuations than but or so. But and so showed strong object biases. In answer to our second question, we found no differences between the different paraphrases, suggesting that traditional connectives have focussing properties, not coherence relations per se.

Finally, we found evidence that referential form choice did differ by connective, with because showing a higher rate of pronoun continuations than but or so, which seemed to pattern together. However, this does not seem to correlate with referential likelihood: objects are the most likely continuation, but subjects tended to be realized as pronouns. Even though because showed a significantly higher rate of subject continuations than the other connectives, and even though because also showed a significantly higher rate of subjects, there was no significant interaction.

Our results are similar to F&vG’s experiments on verb semantic biases. They found that 93% of subject continuations were pronouns, compared to 77% of object continuations. They conclude that referential choice is separate from verb semantics. Our conclusion is similar: referential choice is separate from connective based expectations, and also separate from even more general continuation biases.

Theoretically, the overall object continuation bias was unexpected. Multiple theories suggest a strong overall tendency for subjects to have previous subjects as their antecedents (e.g. the Parallel Preference Theory, Centering Theory). Corpus results also found subjects are more frequent continuations (Arnold 1998;1999), though results surely differ by coherence relation type. Our results are consistent with previous continuation studies finding object continuation preferences (e.g. Arnold (2001), with transfer verbs and F&vG with State verbs). Arnold attributes the discrepancy to a task effect. F&vG instead argue that it is additional evidence that referential choice is separate from grammatical role continuation preferences: by frequency alone, objects were slightly more likely as continuations in their study, but like our results, their participants used proportionally more pronouns with subject continuations than object continuations. These differences may actually also stem from what each theory is trying to predict. Theories of pronoun interpretation are trying to predict the antecedent of e.g. a given subject pronoun, and will prefer a subject antecedent. But this is very different from predicting likelihood of reference after a given sentence and connective.

We also do not know exactly how participants perceived the nonsense verbs: did they assume they were more like e.g. Action verbs, or did they consider them neutrally? Action verbs and Transfer verbs have a general tendency to focus on the endpoint of events, which is the object in these verb classes, so in this respect they pattern with our results. However, the fact that we did find different biases for be-
cause compared to so and but differs from Steve&Co’s results, and suggest that participants interpreted nonsense verbs differently from Action or Transfer verbs. Future work will have to investigate this further.