Ontological Category as a Cue to Set of Alternatives in Child-directed Speech

Lindsay Hracs, *University of Calgary*

Issue: The acquisition of *only* has been the subject of much systematic investigation (see Crain et al., 1994; Kim, 2011; Notley et al., 2009; Paterson et al., 2003; among others). Although results from these studies vary, it is widely agreed upon that interaction between the phonological, syntactic, semantic, and pragmatic factors involved in interpreting utterances with *only* make it a complex learning problem for children to solve. This corpus study investigates the role of pragmatics by asking if children are exposed to ambiguous utterances containing *only* and whether overt context in Child-directed Speech (CDS) is a reliable source of disambiguating information.

Background: Studies examining the acquisition of *only* are primarily concerned with the difficulty children have when the presence of *only* results in multiple interpretations. Consider example (1).

(1) Norma **only** [BAKES [PIE]]$_{\text{FOCUS}}$

The evaluation of an utterance containing *only* depends directly on the Focus of that utterance. In (1), the entire VP, including the embedded NP, is focused giving rise to multiple different interpretations based on a single structural configuration: (i) Norma does not bake anything but pie, and (ii) Norma does not do anything with pie but bake it. Crucially, the presence of *only* signals a restriction on a contextually-defined set of alternatives and the ambiguity arises due to the possibility of different sets of alternatives depending on the constituent in Focus.

A set of alternatives must be comprised of members that are the same "semantic type" as the Focus. The approach adopted here, Conceptual Semantics (Jackendoff, 1983, 2011, i.a.), formalizes the notion of semantic type through the assumption that the central elements of the conceptual system are ontological categories such as ACTION, STATE, THING, DIRECTION, AMOUNT, and PROPERTY. Regarding (1), interpretation (i) corresponds to a set of alternatives made up of THINGs, e.g., {pies, cookies, cakes}, and interpretation (ii) corresponds to a set of alternatives made up of ACTIONs, e.g., {baking, frying, throwing}. Such an analysis provides a practical approach to corpus annotation that is grounded in independently motivated aspects of cognition.

Methodology: A data set was constructed from a subset of corpora from the CHILDES databank (MacWhinney, 2000). The data set contains 490 instances of *only* in CDS. To reduce the chance of task effects, only data from naturalistic situations were included. Utterances containing *only* along with the 20 preceding and following context utterances were extracted from the data set using NLTK *v*. 3.5 (Bird et al., 2009). Target utterances were manually coded as either 'ambiguous' or 'not ambiguous'. Next, context utterances were reviewed to determine whether a set of alternatives was present. Cases which were coded as having a set of alternatives overtly available were then further analyzed to establish the ontological category of the set of alternatives. Lastly, the utterance was evaluated against the set of alternatives to determine whether or not it could be disambiguated.

Results and discussion: Overall, results show that overt linguistic context is a reliable source of disambiguating information in CDS. Regarding the presence of ambiguity, 46.5% of the utterances containing *only* in CDS have multiple possible interpretations. This suggests that the difficulty children have assigning interpretations to ambiguous utterances with *only* is not due to a lack of experience with such structures. Moreover, 82.0% of the ambiguous utterances in CDS can be disambiguated on the basis of overt sets of alternatives made up of the same ontological category as one of the items in Focus. Ultimately, a learner's search for the 'right' entity will be greatly simplified if it is restricted to entities of the same ontological category. Future work seeks to determine if learners attend to the linguistic context available in CDS and if they use this information as a cue to the interpretation when utterances containing *only* have multiple interpretations.

References

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. Sebastopol, CA: O'Reilly Media Inc.

Crain, S., Ni, W., & Conway, L. (1994). Learning, parsing and modularity. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 443–467). Hillsdale, NJ: Lawrence Erlbaum.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (2011). Conceptual Semantics. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning, vol. 1* (pp. 688–709). Berlin, DE: De Gruyter Mouton.

Kim, S. (2011). *Focus particles at syntactic, semantic and pragmatic interfaces: The acquisition of only and even in English* (Unpublished doctoral dissertation). The University of Hawaii at Manoa, Honolulu, HI.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk (3rd ed.)*. Mahwah, NJ: Lawrence Erlbaum.

Notley, A., Zhou, P., Crain, S., & Thornton, R. (2009). Children's interpretation of focus expressions in English and Mandarin. *Language Acquisition*, *16*(4), 240–282.

Paterson, K. B., Liversedge, S. P., Rowland, C., & Filik, R. (2003). Children's comprehension of sentences with focus particles. *Cognition*, *89*(3), 263–294.

A list of corpora included in this study can be made available upon request.