# A CROSS-LINGUISTIC INVESTIGATION OF QUD EFFECTS ON SCOPE-AMBIGUITY RESOLUTION

Yongjia Song (UC Irvine), Abimael Hernandez Jimenez (UC Irvine), Gregory Scontras (UC Irvine)
{yongjias, abimaelh, g.scontras} @uci.edu

Faced with a sentence like *Every horse didn't jump over the fence* as a description of a scenario as in Figure 1, adults readily endorse the utterance as a description of the scene, while children overwhelmingly reject it; however, systematic changes to the task setup lead to marked increases in children's endorsement rates (Musolino and Lidz, 2006; Viau et al., 2010). Savinelli et al. (2017) use a computational cognitive model of utterance endorsement in truth-value judgment tasks to analytically demonstrate how children's non-adult-like behavior likely arises from their inability to manage the pragmatic context, rather than from trouble grammatically accessing specific scope configurations. In an extension of this model, Savinelli et al. (2018) demonstrate how the same considerations can explain adult behavior, such that adults are similarly affected by pragmatic manipulations. We test a clear prediction of these models: manipulating the conversational goal (or Question Under Discussion; QUD) should lead to clear effects on utterance endorsement in a truth-value judgment task. In addition to investigating the predictions for English, we also investigate Spanish and Mandarin. We address two questions: first, whether the relevant languages allow for scope ambiguity in *every-not* sentences, and second, in cases where we believe there to be ambiguity, whether the predictions of the Savinelli et al. model hold.

**Design.** The English version of the Experiment was delivered via Mechanical Turk; the Spanish and Mandarin versions were direct translations of the English, and delivered via Prolific. Participants were introduced to a scenario in which Shark was organizing her storybooks according to specific criteria; the precise criteria depended on the QUD (e.g., whether *none* of the frogs jumped over the rock in Figure 3). The display featured bins labeled with the possible answers to the QUD, which served to illustrate Shark's goal structure. Elephant helped shark by reading a story and telling Shark about it: Shark would ask Elephant a question (the QUD) and Elephant would answer with the *every-not* utterance. Each story described a scenario like that depicted in Figure 1 in which some but not all of the agents completed an action, so Elephant's *every-not* utterance was true only under the inverse interpretation; participants adjusted a slider to indicate whether Elephant's answer was right. Participants completed only a single trial, where they were randomly assigned one of three QUDs: did *none* of the agents complete the action, did *all* of the agents complete the action, or how *many* agents completed the action; these three QUDs were the ones for which the model of Savinelli et al. (2017) makes concrete predictions.

**Results** appear in Figure 2, which plots endorsement rates for the *every-not* utterance for the English (n=263), Spanish (n=310), and Mandarin (n=79) versions of the experiment. Starting with English, we see a clear effect of QUD such that *all?* has higher rates of endorsement than *many?*, which has higher endorsement than *none?*—precisely the pattern of results predicted by the Savinelli et al. model. This effect is partially replicated in Spanish, where *all?* has higher endorsement than the other QUDs. In Mardarin, we fail to find an effect of QUD. Comparing across languages, we see that Mandarin has overall lower rates of endorsement than the other two languages, suggesting that English and Spanish allow for ambiguity in the *every-not* utterance, while Mandarin does not. This interpretation is further supported by the absence of a QUD effect in Mandarin: no amount of contextual support (in the form of a QUD manipulation) can lead the Mandarin sentence to felicitously describe a not-all scenario.

**Discussion.** We have found partial support for the information-theoretic modeling predictions from Savinelli et al.: English participants are more likely to endorse an ambiguous *every-not* utterance as a description of a not-all scenario when that utterance serves as a better answer to the operative QUD. In the case of *all?*, the *every-not* utterance provides a full answer (i.e., "no") under either interpretation; with *none?*, the *every-not* utterance is a particularly ineffective means of conveying that it is not the case that none of the agents completed the action. We discuss practical insight for experimental manipulations of QUDs, as well as the potentially puzzling absence of lower ratings for *none?* in Spanish.

**A note on the languages tested:** The literature on Spanish scope ambiguity is limited, but some work suggests that we should expect to find ambiguity in Spanish similar to that in English (e.g., Barberán Recalde, 2017). In contrast, Mandarin has been claimed to lack the ambiguity altogether (e.g., Huang, 1982; Scontras et al., 2017).

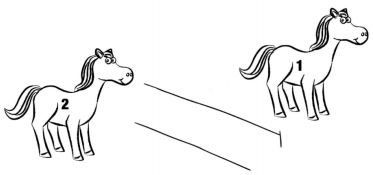


Figure 1. A not-all scenario in which one out of two horses jumps over the fence. In this scenario, the surface interpretation of the *every-not* utterance is false (it's not true that none of the horses jumped) but the inverse interpretation is true (not all of the horses jumped).

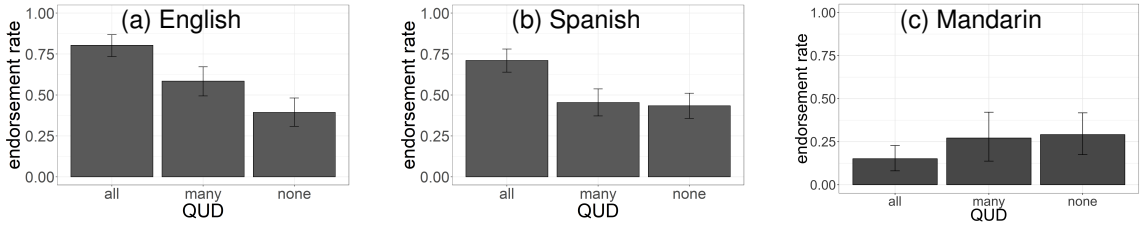


Figure 2. Endorsement rates for the *every-not* utterance in not-all scenarios in (a) English, (b) Spanish, and (c) Mandarin. Error bars represent bootstrapped 95% confidence intervals drawn from 10,000 samples of the data.

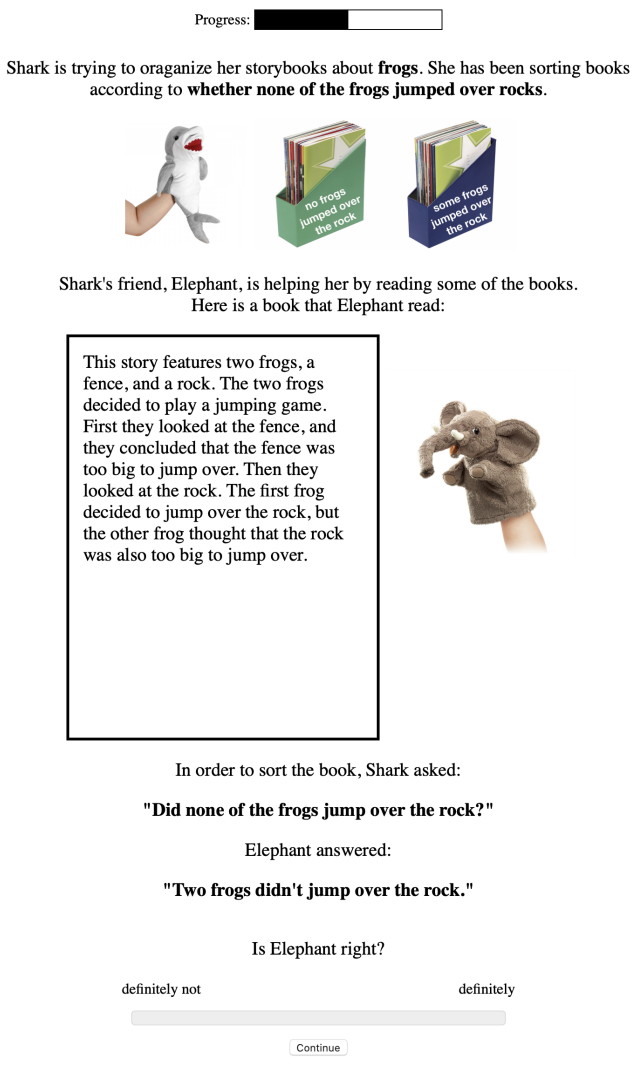


Figure 3. Example trial from the English experiment.

**References:**


Barberán Recalde, M. T. (2017). The interpretation of negated quantifiers in Spanish. In *EPiC Series in Language and Linguistics. Professional and Academic Discourse: an Interdisciplinary Perspective*, Volume 2, pp. 181–190.

Huang, C.-T. J. (1982). *Logical Relations in Chinese and the Theory of Grammar*. Ph. D. thesis, Massachusetts Institute of Technology.

Musolino, J. and J. Lidz (2006). Why children aren't universally successful with quantification. *Linguistics 44*, 817–852.

Savinelli, K., G. Scontras, and L. Pearl (2017). Modeling scope ambiguity resolution as pragmatic inference: Formalizing differences in child and adult behavior. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Savinelli, K., G. Scontras, and L. Pearl (2018). Exactly two things to learn from modeling scope ambiguity resolution: Developmental continuity and numeral semantics. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*, pp. 67–75.

Scontras, G., M. Polinsky, C.-Y. E. Tsai, and K. Mai (2017). Cross-linguistic scope ambiguity: When two systems meet. *Glossa 2*, 36.

Viau, J., J. Lidz, and J. Musolino (2010). Priming of abstract logical representations in 4-year-olds. *Language Acquisition 17*(1-2), 26–50.