Quantifying Interpretation Strategies for Embedded Implicatures in a Markov Decision Network

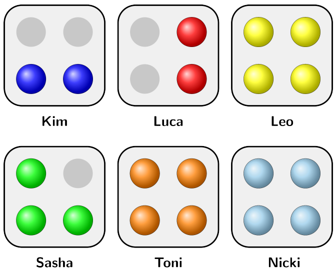
Anton Benz
Leibniz-Centre General Linguistics, ZAS, Berlin, Germany
benz@leibniz-zas.de

Background. All contemporary formal theories, to some extent, build upon [1, 2] and assume that Quantity implicatures are inferred by negating stronger structural alternatives, e.g., by the negation of alternatives '*All found some*' and '*Some found all*' for an utterance of '*Some children found some of their marbles*'. Analyzing data from an interactive experiment on embedded and unembedded implicatures, Benz & Gotzner [3, B&G] found that for some sentences only a minority of participants arrived at the predicted readings. It remained unclear whether there is a sub-population that systematically derives implicatures by negating stronger alternatives, and whether other theories can claim an equally large or even larger proportion of the population. The present study addresses these questions by modelling different versions of quantity-based interpretation strategies and competing strategies proposed by B&G as paths in a Markov decision network. Interpreters choose between these paths. By estimating latent probabilities at the decision nodes, the proportion of the data each theory explains can be quantified.

Data. We use data from a larger experiment that replicates [3] for English. It consists of (i) a production task and (ii) a comprehension task, see Tab 1. In speaker role, participants described a picture using a limited set of quantifiers (among them *all*, *some*, *none*, *some but not all*) with up to 5 sentences of the form 'Q_i *found* Q_j *marbles*'; in listener role, they had to reason which type of picture the speaker saw (by choosing rewards that distinguish seven different worlds). The experiment ran online in asynchronous mode, i.e. experiments were run in groups of 4, the production data were stored and then presented to the next group for interpretation. 162 participants (Prolific) contributed 3402 produced and 5670 interpreted messages.

Modelling. First, additional interpretation strategies were introduced. B&G showed that literal descriptions simplified by reducing *some but not all* to *some* and by removing sub-sentences with the outer quantifier Q_i = '*none*' are reliably understood. A hypothetical *negative* rule was added that retains only sub-sentences with the outer quantifier '*none*'. Reversing these rules allows inferring literal state descriptions from utterances. We call these *decoding* strategies. Then, a **Markov Decision network**, see Fig 1, was defined with interpreters choosing between **random** interpretation, **semantic**, **Q-implic**atures, and **Decoding**. **Q-implic**atures were sub-divided into **lexical** modification, **UE-Impl**icature (only considering alternatives in UE contexts) and **DE-Impl** (adding *none all* $\rightsquigarrow$ *some some*). **Decoding** has also been sub-divided into variants, the main ones being (basic) only the original B&G-rules and (basic+neg) with the additional negative rule. If an interpretation strategy did not lead to a unique solution, a choice between several repair strategies (e.g. *random* interpretation) had been added. A Bayesian version of the MDN was fitted to the aggregated data to obtain credible intervals.

Results: The fitted network shows that it is most likely that only a small proportion of participant derives implicatures by negating stronger alternatives, see Tab. 2. Participants negating stronger alternative either restrict it to lexical modulation or add the rule for DE-contexts. The majority of people follows a decoding strategy, either the basic version or one with additional +neg-rule (with some uncertainty about 'basic' due to the low effective sample size).



some found all,
some found some

✓ exists group found all ■
✓ exists group found sbna ▪
○ exists group found none □

Table 1: Examples of picture (left), produced sentence (middle), and inferred interpretation (right).

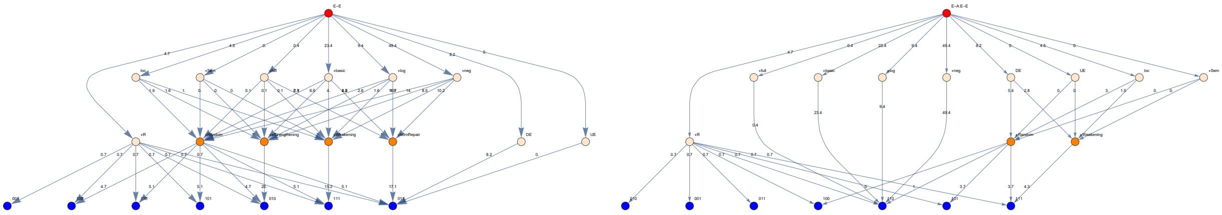


Figure 1: Parts of the MDN: '*some found some*' (left) and '*some found all, and some found some*' (right). The networks of all messages were combined into one MDN (13 free parameters) and fitted to the interpretation data.

Strategy	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
random	1.00	10173	0.05	0.00	0.04	0.05	0.06
semantic	1.00	11379	0.00	0.00	0.00	0.00	0.01
Q-Implic	1.00	7846	0.13	0.01	0.11	0.13	0.16
Lex-Mod	1.00	7563	0.05	0.01	0.02	0.05	0.08
UE-Impl	1.00	4865	0.00	0.01	0.00	0.00	0.04
DE-Impl	1.00	6708	0.08	0.02	0.04	0.08	0.11
Decoding	1.00	8654	0.82	0.01	0.79	0.82	0.84
full	1.00	5998	0.02	0.04	0.00	0.00	0.14
basic	1.01	473	0.28	0.09	0.01	0.29	0.44
+logRed	1.01	391	0.03	0.05	0.00	0.00	0.18
+neg-Dec	1.00	2206	0.49	0.08	0.32	0.49	0.64

Marble Scenario

Table 2: Bayesian posterior probabilities of interpretation strategies for the MDN (# free parameters = 13).

References

[1] Horn, L. R. (1972). *On the semantic properties of the logical operators in English* [Doctoral dissertation, Indiana University].

[2] Horn, L. R. (1989). *A natural history of negation*. University of Chicago Press.

[3] Benz, A., & Gotzner, N. (2021). Embedded implicature: What can be left unsaid? *Linguistics and Philosophy*, *44*, 1099–1130. https://doi.org/10.1007/s10988-020-09310-x