Embedded Implicatures in Comprehension and Production

Anton Benz and Nicole Gotzner (ZAS)

There has been a long theoretical debate on the existence of implicature in embedded positions (see Sauerland, 2012, for an overview). For example, does (A-E) *All of the girls found some of their marbles* implicate that (A-ENA) *All of the girls found some but not all of their marbles*? Previous experimental studies provided mixed results (Chemla, 2009; Geurts and Pouscoulous, 2009; Clifton Jr and Dube, 2010; Chemla and Spector, 2011; Potts et al., 2016; Franke et al., 2017). Gotzner and Benz (2018) were the first to show that, in situations satisfying strong Gricean requirements (cooperativity, competence, relevance), the percentage of subjects drawing the embedded implicatures can be at ceiling (96% for A-E). All studies report comprehension data only and concentrated on a few sentences. In our talk, we present a comprehensive study of embedded implicatures providing both production and comprehension data for about 100 sentence combinations. In addition, we provide an algorithmic production model that produces exactly those sentences with embedded *some* that can be as reliably understood as the corresponding sentences that literally express the embedded implicature.

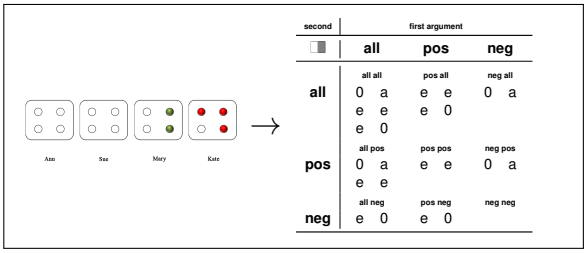
Methods: We have created an interactive game in which four participants take turns in (*i*) a production task and (*ii*) a comprehension task. The paradigm was based on the action-based task developed by Gotzner and Benz (2018). Our programmed system randomly paired two participants for a given production-interpretation trial. (*i*) The speaker had to describe a picture using a limited set of quantifiers (*all, some, none, some but not all, some and possibly all, any* in German) in up to five sentences. (*ii*) The hearer then received this message and had to reason which picture the speaker saw by choosing a set of rewards (distinguishing seven different worlds). Successful communication on a given trial was defined as a match between the hearer's interpretation and the world the speaker had seen. In order to gain more interpretation data on specific sentences of interest, we ran experiment versions with a confederate. In total, 38 native speakers of German took part in the experiment.

Results: The experiments provide a corpus of production and interpretation data of utterances composed of sentences of the form (X–Y) '*X of the girls found Y of her marbles*'. The data confirmed that sentences with embedded '*some*' can be understood as reliably as their literal counterparts (see table below). However, not all potential implicature are understood. For example, (E-E) *some found some* should implicate, according

world	critical strategy	#int	#prod	%int	%prod	literal strategy	#int	#prod	%int	%prod
□	A–N	112	158	100%	100%	A–N	112	158	100%	100%
▒	A–E	82	50	91%	96%	A–ENA	109	143	94%	99%
■	A–A	231	255	98%	100%	A–A	231	255	98%	100%
◧	E–E : E–N	51	37	96%	100%	E–ENA : E–N : N–A	17	27	94%	100%
◧	E–A : E–N	100	105	95%	99%	E–A : E–N : N–ENA	26	29	81%	100%
▒	E–A : E–E	54	38	98%	100%	E–A : E–ENA : N–N	16	21	94%	100%
◧	E–A : E–E : E–N	86	81	100%	100%	E–A : E–ENA : E–N	44	58	95%	100%

worlds: group found ■ (*black*): all; ▒ (*grey*): sbna; □ (*white*): none; **utterances:** A: *all*; E: *some*; *N*: *none*; ENA: *sbna*; **#int/prod**: abs. number of interpreted/produced utt.-tokens; **%int**: % of utt.-tokens interpreted as world; **%prod**: % of utt.-tokens produced in world.

to all theories, that not all found some. However, 42% of the time this implicature was not inferred, and 70% of the time it was violated in production.



	human interpretation	consistency with human production
literal	94%	99.85%
human	90%	95%
critical	95%	99%

Proportion of world-utterance-tokens explained:

# all	literal (UE)	critical	all strategies
1553	830 (53%)	1298 (84%)	1435 (92%)

Proportion of utterance-types explained:

# all	literal (UE)	critical	all strategies
97	36 (37%)	55 (57%)	78 (80%)

Figure 1: Basic coding strategies for *some-some&some-none* world and results of evaluation.

In our post–experimental analysis, we set out to characterize those sentences in our dataset (97) that communicate the state of the world as reliably as literal descriptions. Our assumption was that a crucial aspect is how the pre-verbal message is derived from visual input. We developed an algorithmic model inspired by computational (Reiter and Dale, 2000) and cognitive models (Levelt, 1989) of language generation. We defined ten different basic coding strategies that lead from (abstract representations) of visual input via message selection to utterance generation. We show that a subset of these strategies produces exactly those utterances that reliably communicate the state of the world. Fig. 1 shows an example in which 9 basic coding strategies are applied, from which in a second step utterances are generated (0→*none*, e→*some/some but not all*, a→*all*). For evaluation, we combine these strategies into a critical strategy. It compares to the literal and average human strategy as shown on the right side of Fig. 1. In particular, it is as successful as the literal strategy both with respect to average human interpretation and consistency with human production. The critical strategy accounts for 84% of all world–utterance tokens in the corpus, and 57% of the utterance types.

Conclusions and outlook: Our results extend those of Gotzner and Benz (2018) and show that embedded implicatures are reliably communicated when they are made relevant through an action-based task. Our interactive paradigm distinguished seven different worlds and tested about 100 utterance combinations. We have provided an algorythmic model predicting which of those utterances reliably communicate a given world. The critical strategy predicted by our model was the most efficient one and it was as successful as the corresponding literal strategy. Currently, we are extending our experimental scenario to downward-entailing and non-monotone quantifiers (e.g., sentences like *Exactly one girl found some of her marbles*), in addition to allowing for more varied response options in the production task. Preliminary results suggest that embedded implicatures also occur reliably in these contexts.

Chemla, E. (2009). Universal implicatures and free choice effects: Experimental data. *Semantics and Pragmatics*, 2(2):1–33.

Chemla, E. and Spector, B. (2011). Experimental evidence for embedded scalar implicatures. *Journal of Semantics*, 28(3):359–400.

Clifton Jr, C. and Dube, C. (2010). Embedded implicatures observed: A comment on Geurts and Pouscoulous (2009). *Semantics and Pragmatics*, 3(7):1–13.

Franke, M., Schlotterbeck, F., and Augurzky, P. (2017). Embedded scalars, preferred readings and prosody: An experimental revisit. *Journal of Semantics*, 34:153–199.

Geurts, B. and Pouscoulous, N. (2009). Embedded implicatures?!? *Semantics and Pragmatics*, 2(4):1–34.

Gotzner, N. and Benz, A. (2018). The best response paradigm: A new approach to test implicatures of complex sentences. *Frontiers in Communication*, 2(21).

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. The MIT Press, Cambridge, MA.

Potts, C., Lassiter, D., Levy, R., and Frank, M. C. (2016). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, 33:755–802.

Reiter, E. and Dale, R. (2000). *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge.

Sauerland, U. (2012). The computation of scalar implicatures: Pragmatic, lexical or grammatical? *Language and Linguistics Compass*, 6:36–49.