

Where Truth and Optimality Part. Experiments on Implicatures with Epistemic Adverbs.

Aim: The current paper investigates the derivation of scalar implicatures with the epistemic adverb *poate* 'maybe' in Romanian in the case of Romanian monolingual adults by means of two different tasks: a reward version of a truth value judgment task (i.e. a task where subjects reward characters depending on the truth value of their statements) and a reward version of an optimality judgment task (i.e. a task where subjects reward characters depending on whether what they say is the best description or not). We show that implicature rates are significantly higher in the case of the optimality judgment task (OJT) than in the truth value judgment task (TVJT), thus emphasizing an important methodological point: that results and, consequently, the theory one can build to account for them are to a large extent dependent upon the methods used.

Methodological Background: A lot of thought has gone into how to ask subjects questions so as to see whether they derive implicatures or not (Geurts & Poussolous 2009, Clifton & Dube 2010, Benz & Gotzner 2014, van Tiel 2016 a.o.). Various options have been embraced in the literature: from 'Do you agree with the character?' to 'Is the puppet right?', 'Does the puppet say it well?' a.o. An interesting discussion in Papafragou & Musolino (2003) points to the crux of the problem: "In our version, instead of asking subjects if the puppet is 'right' or 'wrong' (as in the original TVJT), we asked whether the puppet 'answered well' (i.e. *Apantise kala*, 'Did-(she)-answer well?'). This modification was made since we were interested in felicity, not truth." While running a TVJT (Crain & McKee, 1985; Crain & Thornton, 1998) calls for questions about right/ wrong or true/ false, agree/ disagree, running an optimality judgment task (OJT) calls for questions about adequacy/ appropriateness.

Literature on Epistemic Modality: Previous studies on the acquisition of epistemic modality (Hirst & Weil 1982, Noveck, Ho & Sera 1996, Noveck, Ho & Sera 1996, Noveck 2001, Ozturk & Papafragou 2015 a.o.) have focused on children and epistemic modals. Such studies have shown that children are sensitive to the relative strength of modals from very early on, being aware of the existence of a modal scale, but that 5-year-olds still have difficulties with modals, and it is only at Age 7 that the system is fully in place. The paradigm used in most studies is a version of the hidden object task, more specifically, the boxes paradigm: objects are hidden in boxes, and subjects have to evaluate statements about the certainty/ possibility of the location of a certain object/ animal based on current evidence. Interestingly, most of the previous experimental work showed quite high rates for implicatures with modals with adults. However, in an adaptation of Noveck (2001) conducted with epistemic adverbs on Romanian adults (X 2019), many adults were too cautious, rejecting statements about the certainty/ possibility of something they could not see.

Contrastive Experiments Using the Shadow Play Paradigm: Given the considerations above, we developed a novel *shadow play paradigm* (implemented in Penn Controller), where subjects have to reward a dragon for the statements he makes about the identity of a silhouette, on the basis of certain evidence. The reward task was inspired from Katsos & Bishop (2011). Importantly, unlike in the hidden object task/ the boxes paradigm, subjects can infer that the shadow must belong to an animal by looking at its silhouette and because of accompanying sounds (e.g. *woof-woof*, for a dog).

Goal: So as to test whether subjects are more sensitive to underinformativeness than to truth value in deriving scalar implicatures, we ran the same shadow play test in 2 different versions.

Participants: The right-wrong test was conducted on 64 native Romanian speakers, and the optimality test was conducted on 63 Romanian native speakers, recruited from 1st and 2nd Year Students at the Faculty of Foreign Languages, University of Bucharest.

Methodology & Materials: While the tests employ the same type of task (a reward task), the criteria for rewarding are different: truth value ("right-wrong") (in Test 1) and optimality ("best description") (in Test 2). Nevertheless, the tests rely on the exact same set-up. In the *shadow play paradigm*, the subjects are told there is a wizard who likes to play the shadow game with a baby dragon. In this game, various animals go and hide behind the curtain-but some of them may come in front of the curtain later on. The baby dragon has to say who he thinks the shadow belongs to. The subjects are told that they are supposed to reward the baby dragon with a big apple if what he says is right (Test 1)/ the best description (Test 2) and with a small apple if what he says is wrong (Test 1)/ not the best description (Test 2). There are various groups of animals of various colors: a control/ training group

of two bunnies and 4 testing groups of three animals each: dog, frogs, cats, cows. We will now exemplify by reference to the group of dogs: see Figures 1, 2, 3, showing each a picture with the main silhouette, a small image with the animals in front of the curtain, and a small image with all the animals in the game. The small image on the left (ALL ANIMALS) is always present for subjects to easily access the initial situation and in order to prevent processing difficulties because of memory load. *Scenario 1* (where all dogs go behind the curtain) ensures that subjects have in mind the set of animals (*the referential domain*) that is *at issue*, rather than all the animals in the world, or the other sets of animals they have been familiarized. Subjects are supposed to reason that the animal whose silhouette they see must be a dog, not a cat, not a cow, not a frog. *Scenario 2* (where one animal comes back in front of the curtain, in this case, the yellow dog) tests the subjects' understanding of alternatives, their ability to reason that the situation has two possible outcomes: either the silhouette belongs to the red dog or it belongs to the blue dog. *Scenario 3* tests whether subjects are able to reason that the silhouette can only belong to the blue dog, given that there are two animals in front of the curtain now.



Figure 1. Scenario 1



Figure 2. Scenario 2



Figure 3. Scenario 3

There are 59 sentences (3 training Ss, 14x4=56 test & control Ss containing *poate* ('possible') or *sigur* ('certain') presented in a randomized manner (see 1). The randomization was applied both within the same group of animals, and across groups (changing the order in which the animals are presented). The test contains a balanced number of Ss containing both epistemic adverbs so as to activate the modal scale $\langle possible, certain \rangle$ and trigger scalar implicatures. The key sentences for implicature detection are marked in color.

(1) a. It is possible that it is a dog. (UNDERINFO)// It is certain that it is a dog. (OPTIMAL)// It is possible/ certain that it is a cat. (FALSE) b. It is possible that it is the red/ blue dog.(OPTIMAL)// It is certain that it is the red/blue dog. (OVERINFO)// It is possible/ certain that it is the yellow dog. (FALSE).c. It is possible that it is the blue dog. (UNDERINFO)// It is certain that it is the blue dog. (OPTIMAL)// It is possible that it is the red dog. (FALSE)// It is certain that it is the red dog. (FALSE).

SCENARIO 1 (1a)				SCENARIO 2 (1b)					
SPOSSIBLE1 UNDERINFO	SCERTAIN1 OPTIMAL	SPOSSIBLE2 FALSE	SCERTAIN2 FALSE	SPOSSIBLE3 OPTIMAL	SPOSSIBLE4 OPTIMAL	SCERTAIN3 OVERINFO	SCERTAIN4 OVERINFO	SPOSSIBLE5 FALSE	SCERTAIN5 FALSE
SCENARIO 3 (1c)									
SPOSSIBLE6 UNDERINFO	SCERTAIN6 OPTIMAL	SPOSSIBLE7 FALSE	SCERTAIN7 FALSE						

Discussion of the results: The results show a significant task effect in the derivation of scalar implicatures with *poate* 'possible' when comparing the truth value ("right-wrong") judgment test to the optimality ("best description") judgment test. Whereas in the truth value judgment task, 30.3% speakers rejected underinformative Ss, with 14 consistent speakers (>5answers out of 8), in the optimality judgment task, there were 66.67% scalar answers, with 41 consistent subjects. A mixed effects model with Scalarity as variable, Task type as fixed effect, Item, Participant, Order as random effects, points to a statistically significant task effect ($\beta = -2.76, SE = 0.47, Z = -5.76, p < 0.01$). The current experiments thus show the importance of methodology in research: adults are sensitive to implicatures of scalar alternatives only when asked the adequate question.

