Processing non-maximal readings of plural definites: A mouse-tracking investigation

Angèle Bernard – LLING ((UMR 6310) CNRS/Nantes Université)

Summary: We present an experiment using mouse tracking methodology to investigate the processing of non-maximal readings of plural definites. Schwarz (2013) found evidence from response times that non-maximal readings take longer to process than maximal readings, which is taken as an indication of a higher processing cost. We propose to use mouse-tracking as a way to investigate why these readings are more difficult. Specifically, we aim to assess whether no-maximal readings involve a two-step derivation vs. a one-step difficult derivation (Tomlinson et al. 2013). At the workshop, we'll present the final results of the experiment.

Background:

Non-maximality in sentences with plural definites: Sentences with plural definites allow what have been labeled «non-maximal readings», i.e readings that allow exceptions. To illustrate, a sentence like (1) is clearly a true depiction of a situation where all the circles are red, but it can sometimes also be accepted as a true depiction of a situation where some but not all of the circles are red.

(1) The circles are red.

Furthermore, the availability of these non-maximal readings is dependent on the context of the conversation, and specifically on what the question under discussion is, as shown in Malamud (2012)

Context: *Mary is leaving her house for a car trip with a friend. A few minutes after leaving, they see a storm coming in the direction of her house. She believes that she left at least some windows open and she knows that if at least one window is open, the house is not safe from the storm. Mary says:*

(2) Oh my, we have to go back, the windows are open !
⤳ *At least some windows are open.*

There have been few experiments on the nature of these non-maximal readings. Notably, Augurzky et al. (2023) and Romoli et al. (2024) have investigated the effect of context manipulation in the availability of non-maximal readings, showing that there was an effect of context, but the size of the effect was different depending on the linguistic environment of the plural definite. In terms of processing, the only insight goes back to Schwarz (2013), who found that, in non-maximal scenarios (i.e , where the sentence «The dots are red» is used as a description of a picture where some, but not all, of the dots are red), participants who chose to accept the sentence had longer response times than those who rejected the sentence. This was taken as evidence that accessing the non-maximal reading was associated with a higher processing cost, which was hypothesized by Schwarz as being the result of having to derive a pragmatically weaker reading from a literal, stronger, reading.

Theoretically, there are several proposals to account for non-maximal readings. Notably, we can distinguish between an approach where non-maximal readings are a result of a pragmatic weakening (Križ 2015, 2016), and approaches where they arise from a filtering mechanism, either of possible readings (Križ and Spector 2020) or of some of the alternatives that are used to strengthen the sentence (Bar-Lev 2021). These theoretical proposals can be linked to two different processing hypotheses that explain the difficulty of non-maximal readings. On the one hand, a weakening analysis, like Križ's, can explain the processing cost as a result of a two-step process involved in deriving the non-maximal reading (as was also suggested by Schwarz), first computing the literal meaning and then deriving the weakened meaning. On the other hand, under an ambiguity analysis such as Križ and Spector's, the processing cost of the non-maximal readings could be coming from having to resolve an ambiguity between the readings that make the sentence true and those that make it false. Therefore, the non-maximal interpretation could result not from a two-step process but from a difficult one-step process.

Mouse-tracking methodology: While finding longer response times can be a clue that a specific reading is more difficult, this measure cannot inform the source of the difficulty. Tracking mouse movements from participants during a forced choice task can give us more insights on the underlying processes involved in the decision making. Because motor responses are argued to be planned in parallel to cognitive processing and to be realised as soon as possible (Song & Nakayama 2006), researchers have drawn the hypothesis that reflects of cognitive processes can be found in mouse-trajectories.

On the question of sentence interpretation, mouse tracking has been primarily used to investigate processing of negative sentences (Dale and Duran 2011, Maldonado et al. 2019, ao) but the question of the processing of scalar implicatures was also investigated by Tomlinson at al. (2013) for the scalar item «some» in English and by Sauerland et al. (2017) for disjunctive items in Japanese. The type of design used in aforementioned studies is always similar: Participants have to click on a [START] button at the bottom of their screen to display a stimuli (either a sentence, or a sentence image pair) that they have to then evaluate as 'true' or 'false' (or, 'good' or 'bad') by clicking on one of the two buttons that appear on the top corners of the screen. The assumption behind all these studies is that mouse trajectories reflect cognitive processing and so mouse-tracking in a good tool to investigate whether a reading involves a two step derivation. The hypothesis is that the serial computation of the literal meaning first and the pragmatic meaning second is inducing a change in decision during the judgment task, which should be reflected by the mouse-trajectory in a a lot of deviation towards the competitor response. Several measures can be taken on the mouse trajectories, such as the area under the curve (AUC), the maximal deviation point etc... The actual methodological question of what measures best diagnose if a change of decision has happened is a complicated issue (Maldonado et al. 2019). In their study, Tomlinson et al. argue for using the AUC as a measure of overall difficulty and the deviation from the medial axis towards the competitor response (Xneg) as a measure of change in decision (see Figure 1 for a visualization of the prototypical expected mouse-paths for a difficult one step-process and for a difficult two-step process).

While Tomlinson et al. report finding evidence for a two-step process leading to the pragmatic reading of sentences with the scalar item «some», the results are less clear for pragmatic readings of conjunctive items in Japanese in Sauerland et al.

We will use the same reasoning applied to non-maximal readings. The apparent extra difficulty to access non-maximal readings found in Schwarz (2013) could have several explanations. On the one hand, it could be a result of some form of a two step mechanism, whereby speakers first compute a maximal (universal) interpretation and then weaken the meaning of the sentence (via a pragmatic mechanism like the one proposed by Križ). On the other hand, it could come from a one step process where speakers are entertaining several interpretations for the sentence and the forced choice in the task pressures them to resolve ambiguity (that could be better explained in a framework such as Križ and Spector's). Looking at the mouse-movements of participants that judge a sentence true in a non-maximal scenario can be useful in helping distinguish between the two possibilities. If the two-step interpretation is what happens here, we should find clear signs of deviation towards the button [FALSE] before the mouse goes back towards the button [TRUE]. Conversely, if the difficulty comes from having to resolve ambiguity, we should not find clear patterns of deviations but we can expect an overall difficulty reflected in a bigger AUC than the true control items.

Methods: Procedure: We adopt a similar design to Tomlinson et al. and Sauerland et al. The experiment is implemented using the JavaScript framework JsPsych, and is run on browser. In the critical phase, each trial starts with a screen showing, in the middle, an array of shapes of different colors; on the top left and right corner, the buttons [TRUE] and [FALSE] (in an order randomized by participants); and at the bottom a button labeled [SHOW THE SENTENCE], which is the only clickable button at this point. Once the participant clicks on this button, a sentence of the form «The [shapes] are [color]» appears on the screen and participants have 6 seconds to chose whether they think the sentence is true or false by clicking on one of the top buttons (See figure 2 for an illustration of the procedure). Participants can chose freely whether they think the sentence is true or false. To enforce a quick response from participants, a timer sound sets off at around 3 seconds, and we added progression bars on the sides of the screen, which visually indicate how much time they have left to answer. We want to make the participants answer quickly so that they start moving their mouse right away. Mouse-paths are recorded as a list of triples that contain the x and y coordinates of the mouse and the time t after the beginning of the trial.

Materials: All stimuli are composed of pairs of sentences and images that are arrays of 9 shapes. In the critical items, participants see a pair where the predicate is true of 6 out of 9 of the shapes mentioned in the sentence. We chose to make the experiment rather short, with three items of the critical condition. This is because we think the expected mouse patterns are likely to show more in the first trials of the experiment, when the participants are not yet used to the task, and because we plan to run the experiment online, recruiting participants via the platform Prolific, which gives us access to a large pool of participants. We also added two types of control items, a true control where the sentence is clearly a true depiction of the picture, and a false control where the control is clearly false (3 items for each controls). Since we don't want the true controls to be easier to judge because of the presence of only one color compared to the critical trials where there are two colors on the picture, we add an extra shape in the control trials (See figure 3 for examples of each stimulus). For all trials, the images are generated automatically during the experiment and the placement of the shapes in the pictures is completely randomized.

Analysis: Data preprocessing: In order to analyze the mouse-paths, the data must be preprocessed. Since the experiment is run online and participants can have different screen and mouse resolution, so we will remap all trajectories so that they all start and end at the same point, and normalize them on time so that they all are composed of 101 points, which will allow us to compute summary statistics. These transformations will be performed using the R package *mousetrap* (Wulff et al. 2023).
Planned analysis: We will compare the critical true trajectories against the control true trajectory. First, we want to see whether the findings from Schwarz replicate by testing looking at the response times. Further, comparing the AUC from critical true and control true will hopefully provide further evidence that non-maximal readings are indeed more cognitively difficult to process. Second, we'll investigate whether non-maximal readings involve a two-step process by comparing the Xneg for both conditions.

Pilot results: We ran a first pilot on 19 participants, mainly to assess whether we could get enough true answers from the participants, as they are the responses we are interested in. Surprisingly, the design elicited a majority of true answers (62%, se=10), with 9 participants consistently answering true across all three trials, which was not expected given that Schwarz reports only 30% non-maximal readings in his study (Figure 4). As for the response times, the critical trials seem to take slightly longer, although the dataset is too small to make significant conclusions (Figure 5). For now, we collected too little data to extract reliable descriptive statistics from the mouse tracking trajectories. The full results will be presented in the workshop.

References:

Augurzky, P., Bonnet, M., Breheny, R., Cremers, A., Ebert, C., Mayr, C., Romoli, J., Steinbach, M., & Sudo, Y. (2023). Putting plural definites into context. *Proceedings of Sinn Und Bedeutung, 27*, 19-32. https://discovery.ucl.ac.uk/id/eprint/10189579/

Bar-Lev, M. E. (2021). An implicature account of homogeneity and non-maximality. *Linguistics and Philosophy*, *44*(5), 1045-1097.

Dale, R., & Duran, N. D. (2011). The Cognitive Dynamics of Negated Sentence Verification. *Cognitive Science*, *35*(5), 983-996. https://doi.org/10.1111/j.1551-6709.2010.01164.x

Križ, M. (2015). *Aspects of homogeneity in the semantics of natural language : University of Vienna dissertation*.

Križ, M. (2016). Homogeneity, non-maximality, and all. *Journal of Semantics*, *33*(3), 493-539.

Križ, M., & Spector, B. (2021). Interpreting plural predication : Homogeneity and non-maximality. *Linguistics and Philosophy*, *44*(5), 1131-1178.

Malamud, S. A. (2012). The meaning of plural definites : A decision-theoretic approach. *Semantics and Pragmatics*, *5*, 3-1.

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods*, *51*(3), 1085-1101. https://doi.org/10.3758/s13428-018-01194-x

Romoli, J., Sudo, Y., Augurzky, P., Bonnet, M., Breheny, R., Cremers, A., Ebert, C., Erbach, K., Steinbach, M., & Mayr, C. (s. d.). *Putting plural definites into context*. Consulté 2 août 2024, à l'adresse https://lingbuzz.net/lingbuzz/007987/current.pdf

Sauerland, U., Tamura, A., Koizumi, M., & Tomlinson, J. M. (2017). Tracking Down Disjunction. In M. Otake, S. Kurahashi, Y. Ota, K. Satoh, & D. Bekki (Éds.), *New Frontiers in Artificial Intelligence* (Vol. 10091, p. 109-121). Springer International Publishing. https://doi.org/10.1007/978-3-319-50953-2_9

Schwarz, F. (2013). Maximality and definite plurals-experimental evidence. *Proceedings of Sinn und Bedeutung*, *17*, 509-526. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/358

Tomlinson Jr, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some : Scalar implicatures are understood in two steps. *Journal of memory and language*, *69*(1), 18-35.

Appendices:

Figure 1: Expected two-step vs. generally difficult two-step trajectory (Taken from Tomlinson et al. 2013)

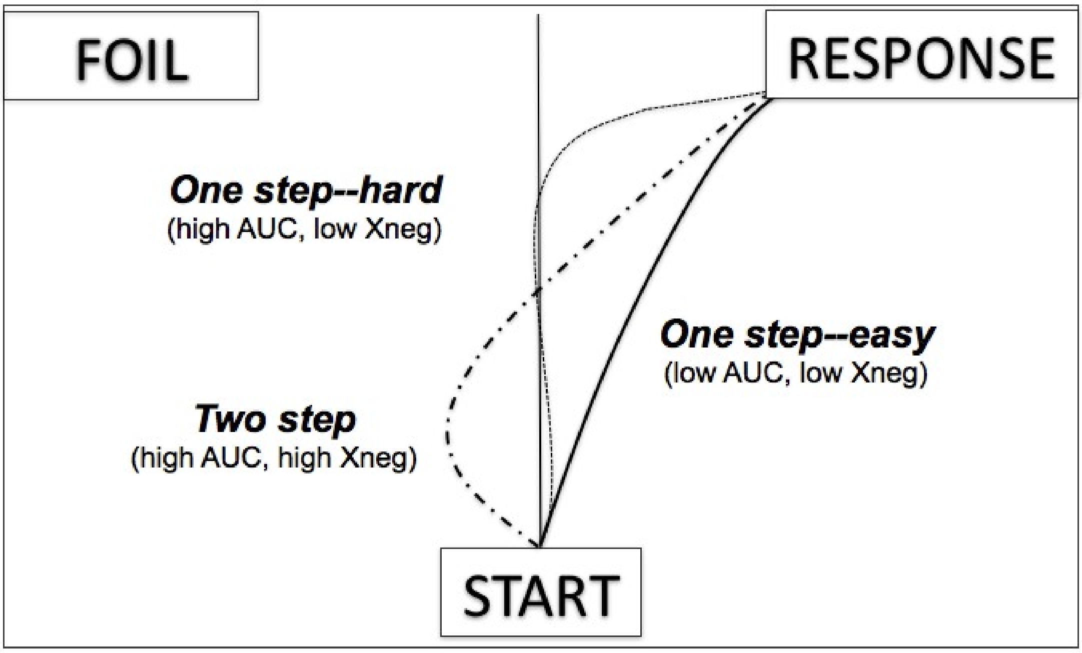

Figure 3: Examples of stimuli for critical, control TRUE and control FALSE trials:

Critical | **Control True** | **Control False**

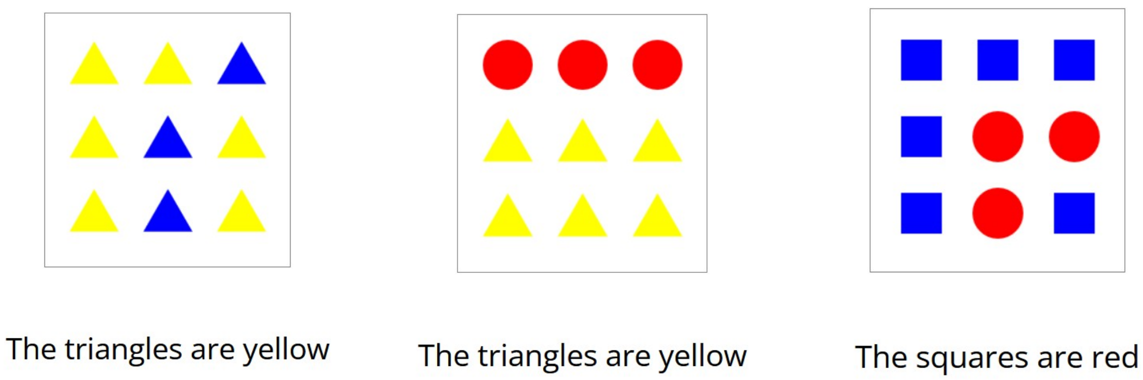

Figure 2: Experimental procedure

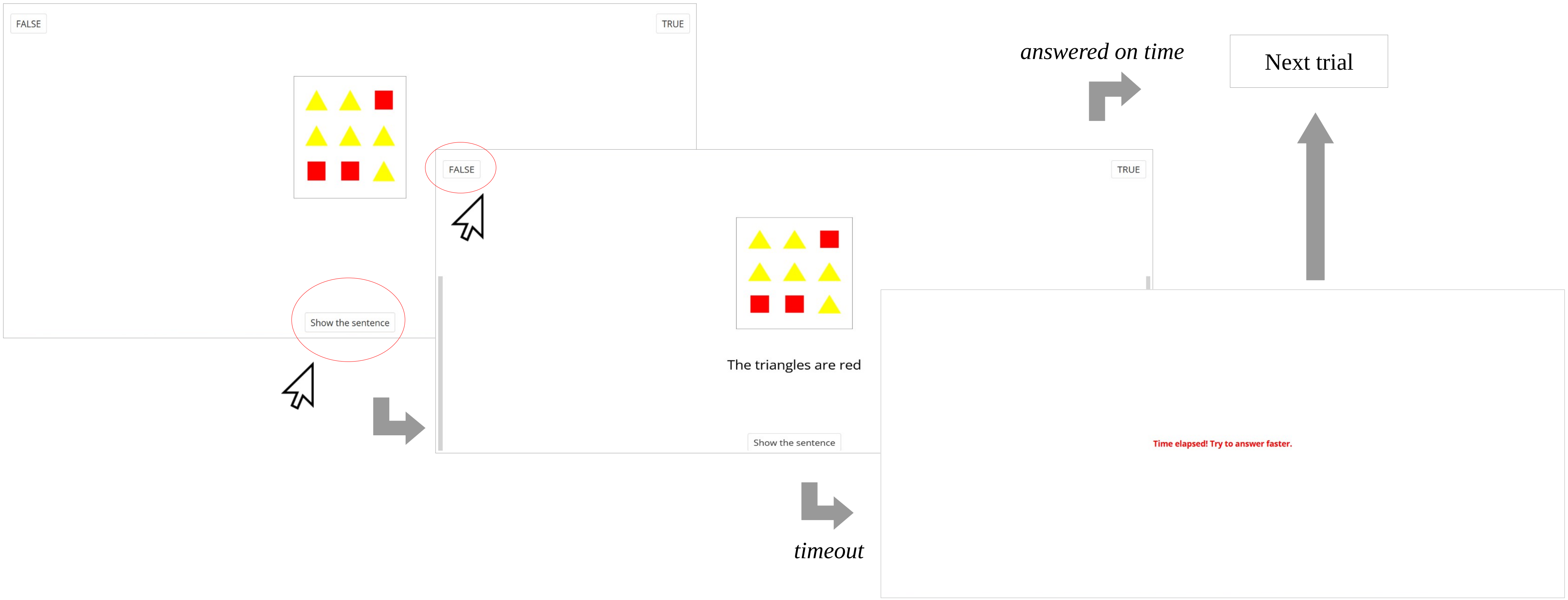

Figure 4:

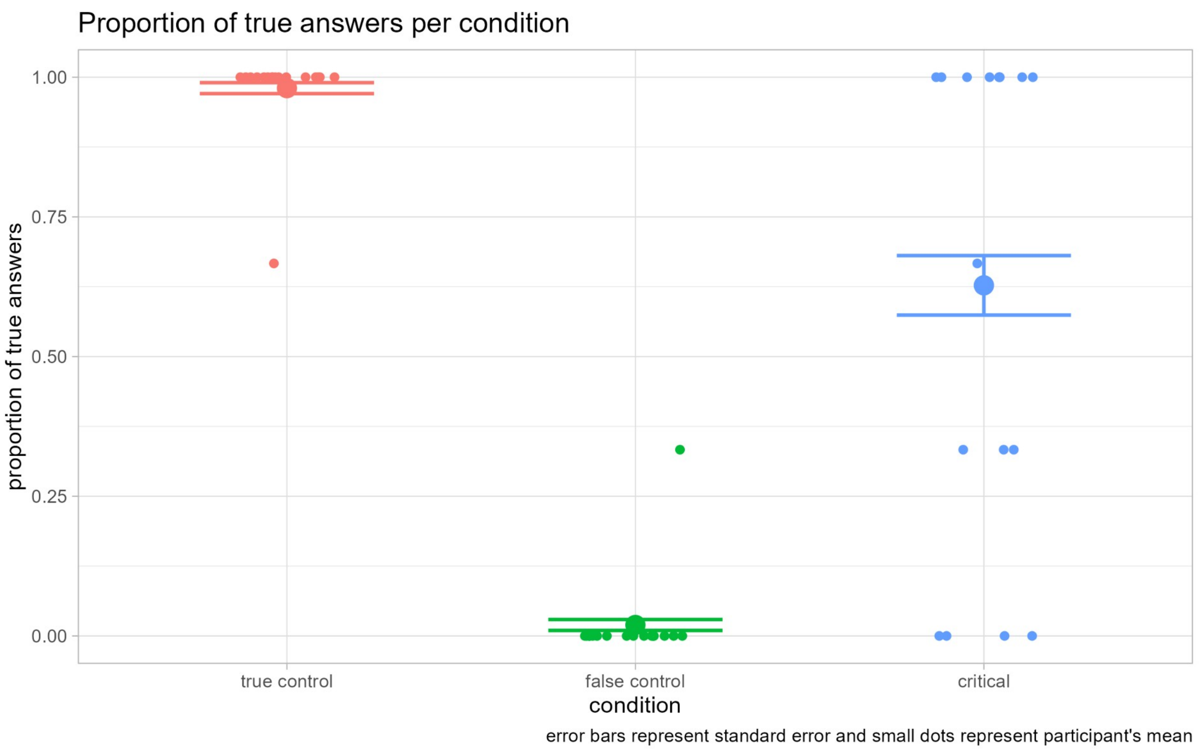

Figure 5: