Verb Surprisal in the Visual World

Christine Muljadi, Christine Ankener, Les Sikos, Maria Staudte (Saarland University)
masta@coli.uni-saarland.de

The effort for processing a word in a given linguistic context is correlated with its surprisal and predictability [1]. That is, less expected words are more surprising and thus more difficult to process. Previous work has extended classical notions of surprisal to visually-situated contexts [2]. That paper presented a visual world eye-tracking study which examined whether manipulations of *referential entropy* (i.e., the number of potential referents in the visual display that were consistent with the target at a given point in the utterance) modulated the processing of the target noun. They did so by holding the auditory linguistic context constant (e.g., "Die Frau verschüttet gleich das Wasser." / English: "The woman spills soon the water.") while varying the number of objects in the display that were consistent with the verb (i.e., spillable things). Eye movements during the verb indicated that listeners anticipated (only) suitable objects. That is, uncertainty about the upcoming referent was reduced by exploiting the verb constraint. This parametrically decreased processing effort (as measured by the pupillometric Index of Cognitive Activity (ICA) at the sentence-final noun—but it did not affect processing effort at the verb. What was left unclear was why no effect of *referential entropy reduction* was found on verb processing. After all, verbs reduced uncertainty to a greater or lesser extent depending on the visual context, which also led to differences in anticipatory eye movements. One possibility is that the effect of referential entropy reduction is only manifest at nouns, because nouns serve as direct pointers to objects in the world, while verb selectional restrictions simply constrain expectations to a smaller set of possibilities.

The current study directly investigates this question using a setup similar to [2] and interrogative statements (in German) because they allow the verb to appear before the noun (e.g. "Tell me if the rose that is watered by the figure is located at the top / bottom / left / right / is missing."). Participants (N=32) viewed 40 experimental displays that varied the number of verb options (4, 3, 1, or 0) by depicting actions rather than objects, while participants listened to utterances related to the display (Fig.1). We recorded ICA, which has been shown to index processing effort across a range of linguistic phenomena [3], in addition to traditional eye movement measures.

Mixed modelling (maximal; Poisson distribution) of ICA values revealed that processing effort at the target verb was reliably modulated by the number of actions in the display that were consistent with the verb (Fig.2). More specifically, when only 1 verb-consistent action was displayed, mean ICA values were significantly lower (m=42.9, SD=13.1) than when 3 verb-consistent actions were shown (m=47.1, SD=11.2; ß=-.099, SE=.027, z=-3.70, p<0.01). 3 and 4 verb-consistent actions evoked similarly high ICA values (m=47.12 and m=47.08, respectively). Somewhat surprisingly, the 0 verb-consistent actions condition yielded the lowest ICA values (m=38.54, SD=15.4; wrt to 1 consistent action: ß=-.128, SE=.034, z=-3.76, p<0.01). This finding differs from [2], where the equivalent condition yielded the highest values. However, in the 0 condition in the current experiment, participants could already determine at the noun that the answer to the question (e.g., "Tell me if the rose….") could only be "Yes" if the question ended with "… is missing". Thus, listening to the verb was not useful in this case, thereby making the 0 condition the easiest at the point of the verb. Finally, statistical analyses further revealed typical anticipatory eye movements during the noun (i.e. looks only to likely upcoming actions/verbs), but as in [2], this did not appear to modulate processing effort at the noun.

These results indicate that visual context can similarly affect the predictability and surprisal of both verbs and nouns. We also replicate the lack of an effect on processing effort for the word that provides the constraining information (i.e. reduces referential entropy). Thus, regardless of word class, processing effort seems to correlate with situated surprisal but not with referential entropy reduction – here mediated by visual context and attention.
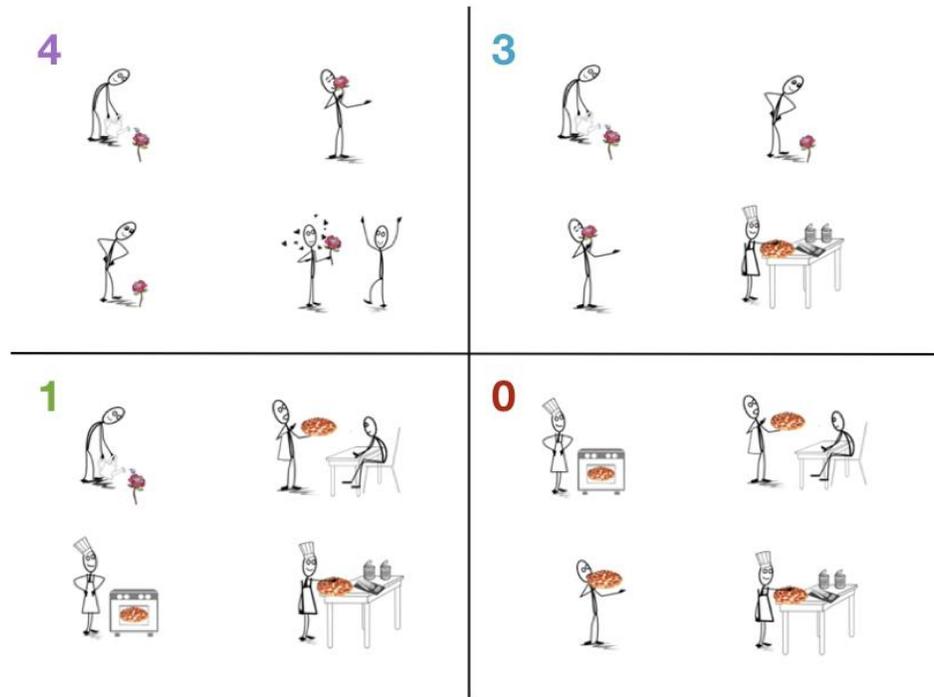
Fig.1: Example visual displays in all four conditions for the sentence: "Sag mir, ob die Rose, die von der Figur gegossen wird, unten ist." / English translation: "Tell me if the rose that is watered by the figure is located at the bottom". Upper left: 4 possible actions. Upper right: 3 possible actions. Lower left: 1 possible action. Lower right: 0 possible actions.
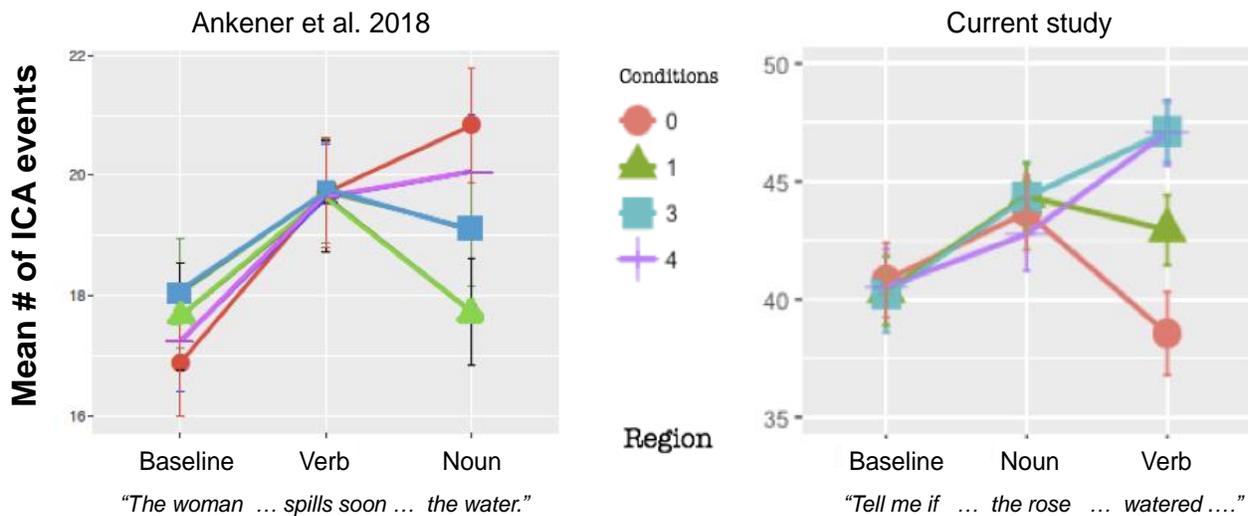

Fig.2: Comparison of mean ICA results (larger values correspond to greater processing effort) elicited during 3 time-windows (pre-noun baseline, noun region, verb region) and experiments. Left: Results from [2]. Right: Results from the current experiment.

References

[1] Hale, 2001. [2] Ankener, Sekicki & Staudte (2018). [3] Demberg & Sayeed (2016).