Results

timo roettger

12/17/2019

Results

Looking at the fixation patterns across time (see Fig. 1), it becomes clear that listeners are biased to look at the picture that was the target of the trigger sentence, namely the picture that corresponds to the already mentioned animal and object (i.e. the red line (given NP1 - given NP2) starts higher than expected by chance and the light blue line (contrastive NP1 - contrastive NP2) starts lower). This initial bias is, however, quickly modulated by the incoming speech information.

Fig. 2 displays the aggregated fixation preferences for the animals (= 1st NP) and the objects (= 2nd NP) across three relevant time windows: the **early** window prior to the 1st NP and its prosodic profile, the window spanning the onset of the **1st NP** (= "thingy") up to the onset of the 2nd NP, and the window spanning the onset of the **2nd NP** up to the onset of the adverb.

Looks to the 1st NP

Looking at the fixation preferences to NP1 (top row), there is compelling evidence that listeners use information of the prenuclear region in the 'NP1 accented' condition (left panel). There is a compelling increase in looks to the target referent from the early window ($\hat{\beta} = 0.28$) to the NP1 window ($\hat{\beta} = 0.37$). The difference of the posterior distributions is compelling ($\hat{\beta}(logodds) = -0.44$, 95% CrI = $[-0.66, -0.21]$). Thus, hearing a prominent pitch accent on NP1 leads to more looks to a referent that contrasts with the previously mentioned referent. In other words, the pitch accent is immediately locally interpreted. The same pattern can also be numerically observed for intonation contours with two pitch accents (right panel). Listeners are more likely to look away from the target referent when hearing a pitch accent on thingy: There is a *de*crease in looks from the early window ($\hat{\beta} = 0.75$), to the NP1 window ($\hat{\beta} = 0.69$). However, in this condition there is substantially more variability, leading to less certain estimates of this difference ($\hat{\beta}(logodds) = 0.28$, 95% CrI = $[-0.07, 0.64]$). Within these two time windows, there was no compelling change in fixation preferences for the 'NP2 accented' condition (middle panel) ($\hat{\beta}(logodds) = 0.22$, 95% CrI = $[-0.08, 0.53]$), indicating that the absence of a pitch accent on NP1 is not used (as much) to infer the givenness of the corresponding referent.

Whether listeners make inference about the discourse status of NP1 based on the prosodic information on NP2 (backwards inference) can be seen in the comparison between the second and third windows. When having heard an early pitch accent on NP1, listeners use the absence of a pitch accent on NP2 (left panel) as additional evidence for the givenness status of NP1, indicated by a compelling increase in looks to the target referent ($\hat{\beta}(logodds) = -0.73$, 95% CrI = $[-0.97, -0.49]$). The evidence for the other two conditions remains inconclusive, indicating only small differences as well as great remaining uncertainty (NP2 accented: $\hat{\beta}(logodds) = -0.26$, 95% CrI = $[-0.69, 0.16]$; Both NPs accented: $\hat{\beta}(logodds) = 0.21$, 95% CrI = $[-0.2, 0.68]$)

Looks to the 2nd NP

Beyond the local interpretation of prenuclear prosodic information (i.e. in inferring the discourse status of NP1), it is informative to look at whether listeners use the early prosodic information to infer the discourse status of the upcoming NP2 (bottom row). Only the '2nd NP accented' condition (middle panel) shows a change in preferences. The early window ($\hat{\beta} = 0.33$) exhibits less looks to the target NP2 than the NP1 window ($\hat{\beta} = 0.44$). The model estimates a compelling difference between these windows ($\hat{\beta}(logodds) = -0.44$, 95% CrI = $[-0.71, -0.14]$). In other words, when listeners do not hear an early pitch accent, they anticipate a contrastive NP2. While, within the microcosm of the experiment, an early pitch accent was highly predictable

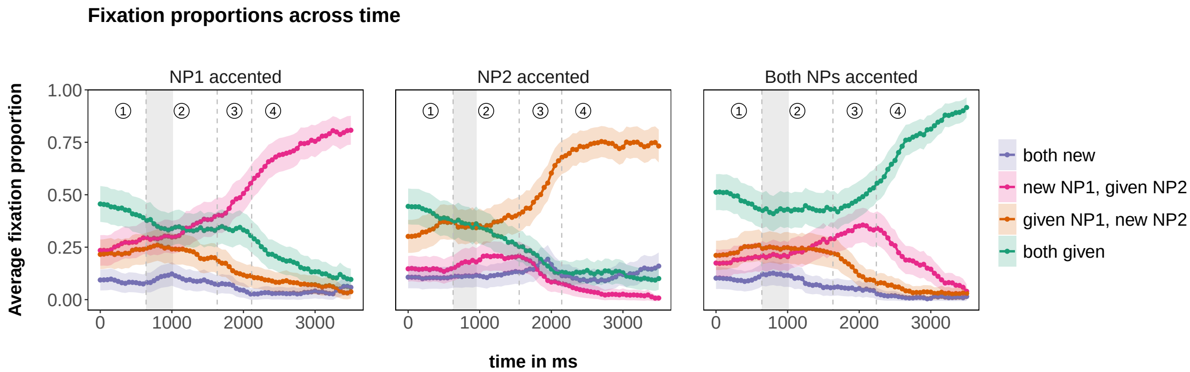

Figure 1: Mean fixation proprotions across conditions (panels), picture types (color) and time (x-axis). Error margin corresponds to the standard error calculatate by SD / (square root (number of participants). Dashed lines correspond to the acoustic landmarks plus 200 ms lag for the initiation of the saccade.

of the givenness of NP2, listeners did not seem to use that information to predict NP2's discourse status (NP1 accented: $\hat{\beta}(logodds) = 0.02$, 95% CrI $= [-0.21, 0.26]$; Both NPs accented:$\hat{\beta}(logodds) = 0.07$, 95% CrI $= [-0.2, 0.38]$) This is important as this indicates that listeners are not strictly rational comprehenders who greedily use whatever evidence they encounter. We come back to this interpretation in the general discussion.

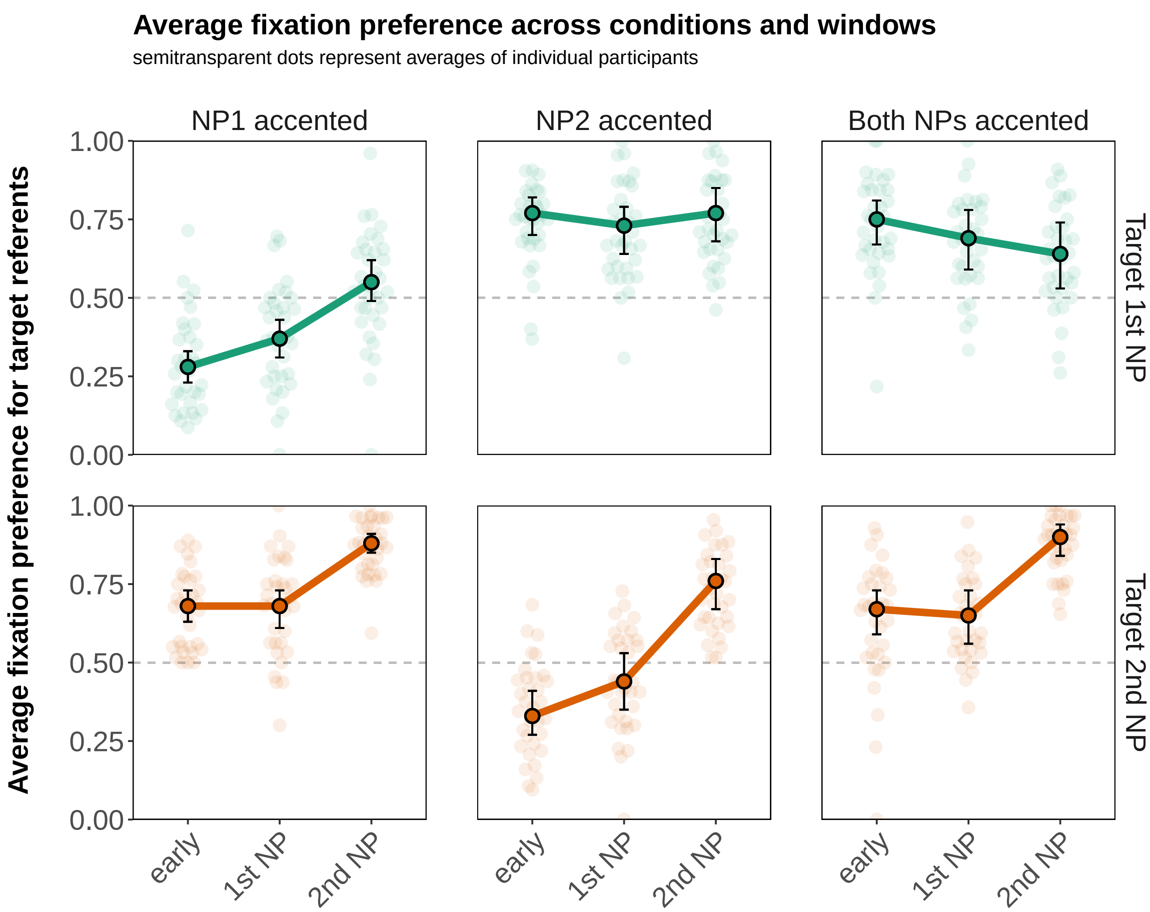

Figure 2: Aggregated fixation preferences across conditions (panels), time windows (x-axis) and referent type (color). Solid points and whiskers correspond to posterior mean and 95% CrI based on the models described above. Semitransparent points (and corresponding density) correspond to listener averages.