Sentence processing is "good enough": Evidence from sentence-video matching

Gaurav Kharkwal & Karin Stromswold (Rutgers University, New Brunswick)
kharkwal@rutgers.edu

Sentence processing; Language representation; Sentence-video matching

Debate exists about the depth at which people process sentences, with some people suggesting that parses are "good enough" for the task being performed (e. g., Ferreira & Henderson, 2007). We investigated the nature of syntactic representations formed in two experiments.

Experiment 1. Twenty-five native, monolingual English speakers watched captioned videos of geometric objects (circles, squares, rectangles or ovals) "chasing" each other and performed a sentence-video matching task. As shown in [1], captions were active sentences. The verbs *chase, flee, lead,* and *follow* were used because all can describe the same visual scene, albeit from different perspectives. In half of the trials, captions did not match the video because the NPS were reversed. In half the trials, the objects had 2 eccentrically placed dots ("eyes") and in half, the dots were centered.

1a. The square is following/chasing the circle.
1b. The circle is leading/fleeing the square.

Participants were accurate (94.9% correct), with a mean RT of 2945 ms (*SE* = 63 ms). Participants were faster in conditions were the objects had eyes ($F(1,24) = 8.17$, $p=.009$). Verb Perspective had a significant effect, with participants faster on "front-perspective" verbs (*lead, flee*; $F(1,24)=7.35, p=.012$). Performance was unaffected by Verb (Bayes Factor, $BF = 58.15$, $p_{null} = .98$) and Caption Veracity ($BF=4.09$, $p_{null} = .80$).

Experiment 2. Twenty-four native, monolingual English speakers participated. The experiment was identical to Experiment 1 except half of the sentences were passive. Also, the verbs *guide* and *trail* replaced *chase* and *flee* because *flee* does not passivize.

2a. The square is followed/trailed by the circle.
2b. The circle is led/guided by the square.

Participants were just as accurate (95.7%) as in Experiment 1, but were 30% slower (*M*=3821 ms, *SE*=60 ms). Notably, their RTs were affected by Verb ($F(3,69) = 12.21$, $p<.001$), Voice ($F(1,23) = 29.87$, $p<.001$), Caption Veracity ($F(1,23) = 15.73, p = .001$), and Verb Perspective ($F(1,23)=25.82$, $p<.001$), but NOT by presence of eyes ($BF = 1.72$, $p_{null}=.63$).

Discussion. The fact that participants were about 30% slower in Experiment 2 suggests that people may have processed sentences differently in the two experiments. The slower speed in Experiment 2 cannot be the result of a speed-accuracy tradeoff as participants were equally accurate in the two experiments. The slow speed also cannot be the result of averaging RTs of actives and passives because participants were 21% slower on actives in Experiment 2 than Experiment 1. Furthermore, it cannot be that participants in Experiment 1 (and not Experiment 2) only parsed the first noun and matched it with the trailing (or leading) object, because inclusion of perspective shift verbs would have resulted in 50% accuracy.

In Experiment 1, because all sentences had a canonical NVN structures, participants could have done a "rough" parse (e.g., Townsend & Bever, 2001). Because of inclusion of passives in Experiment 2, that strategy is inadequate and a more detailed analysis in necessary. A sentence-video matching task cannot be performed until both the sentence and the video have been processed. In Experiment 2, but not in Experiment 1, a detailed analysis may have resulted in sentence processing being the limiting factor, thus overshadowing the effects of visual parameters. Consistent with this account, in Experiment 1, both the visual parameter (Eyes) and the slightly broader linguistic parameter (Verb Perspective) had significant effects, while in Experiment 2, only the linguistic parameters (Voice, Verb, Verb Perspective, and Caption Veracity) had significant effects. Taken as a whole, these results are consistent with "good enough" models of language processing in which detailed representations are formed only when necessary.

Ferreira, F., & Henderson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, 1:71-83.
Townsend, D., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules.* Cambridge, MA: MIT Press.