What’s happening? Biases in the visual perception of event scenes and their descriptions

YULIA ESAULOVA, SARAH DOLSHEID & MARTINA PENKE
(University of Cologne)

Language provides its users with a number of possibilities when it comes to producing an utterance. The choices language users make are often driven by the environment they interact with and follow particular patterns. The present study examined visual and conceptual aspects of depicted event scenes as potential sources for systematic variations in scene descriptions.

There is some evidence that when speakers’ visual attention is directed to the patient and not the agent, they are more likely to produce passive voice sentences (e.g., Gleitmann, January, Nappa & Trueswell, 2007). At the same time, there seems to be a preference for spatial representations of events, such that agents are typically positioned to the left of patients (e.g., Chatterjee, Southwood & Basilico, 1999). This study investigated whether and how these two visual properties may interact affecting language production. In addition, a number of voice and word order variations have been explained by prominence effects of animacy, where animate entities are perceived as more prominent than inanimate ones and thus are more likely to be assigned subject/agent roles (e.g., Lamers & de Swart, 2012). Whether these prominence effects may be modulated by visual information was another question addressed in this study.

Native speakers of German (N = 44, mean age = 23.43 years, SD = 3.01) were tested in a picture description task while seated in front of a computer screen with an eye-tracker. The pictures depicted scenes where an animate agent performed an action on either an animate or an inanimate patient. Patients were situated to the right or to the left of agents (Figure 1) and half of them were preceded by a short visual cue (see Figure 2). Depicted referents were controlled in syllable length and were comparable in size, visual complexity and distance within which they were situated from each other across items. The portrayed transitive interactions involved no direct contact and the corresponding verbs had balanced profiles in terms of their likelihood to occur in active and passive voice frames. Participants were instructed to describe each picture using one sentence.

The results show that scenes with left- rather than right-positioned patients lead to longer speech onset times ($F(1, 43) = 6.46, p = .015$) and a higher number of passive sentences ($F(1, 43) = 5.48, p = .024$). In addition, passive utterances occurred more often for scenes with animate rather than inanimate patients ($F(1, 43) = 8.41, p = .006$). Animate patients and patients occurring to the left of agents were more likely to be fixated prior to speech onset compared to inanimate patients ($F(1, 43) = 64.56, p <$
.001) and those positioned to the right of agents ($F(1, 43) = 54.63, p < .001$). Cueing of patients did not have an effect on either speech onset times or the number of passives, however, the analyses of eye-tracking patterns revealed more initial saccades to patients after they were cued than when they were not ($t(43) = 4.83, p < .001$).

Our findings demonstrate that visual and conceptual properties of event scenes influence different aspects of language behavior. Both the initiation of utterances and the voice selection were mostly affected by the positioning of patients in event scenes, thus revealing participants’ bias to expect agents to the left of patients in visual scenes. Possible processes underlying left-agent preferences may involve an alignment between the linear representation of thematic roles and the serial planning of speech, thus relating to agent-first preferences (e.g., Jackendoff, 2002). Moreover, voice selection was sensitive to the animacy of thematic roles, so that more passive utterances were produced for scenes where both arguments were animate. Participants therefore displayed a bias to perceive animate roles as better fitting subject functions than inanimate ones. This is in line with prominence theories suggesting that animate entities are more likely to be realized as subjects in sentence-initial positions than inanimate ones. The discussion of findings integrates cognitive and linguistic models relating differences in linguistic output to attention and prominence effects.


