SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pages 194–195.
Los Angeles, California, 21–23 September 2011.

Optional visual information affects conversation content

Richard Andersson
Lund University Cognitive Science
Kungshuset, Lundagård
SE-222 22, Lund, Sweden
Richard.andersson@humlab.lu.se

Jana Holsanova
Lund University Cognitive Science
Kungshuset, Lundagård
SE-222 22, Lund, Sweden
Jana.holsanova@lucs.lu.se

Kenneth Holmqvist
Lund University Humanities Lab
Helgonabacken 12
SE-221 00, Lund, Sweden
Kenneth.Holmqvist@humlab.lu.se

Abstract

The language processing system is opportunistic and makes use of several information sources, if available. One extensively tested source of information is the visual modality. We now know that we can use the visual context to disambiguate structurally ambiguous sentences (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995), and that visually inferred agent statuses bias our assignment of thematic roles (Knoeferle, Crocker, Scheepers & Pickering, 2005) . Furthermore, we use the visual information to predict upcoming material by exploiting the semantic links between the visual object and its linguistic counterpart (Altmann & Kamide, 1999; Kamide, Altmann & Haywood, 2003).

Tracking the use of visual information in linguistic tasks has also been performed in non-stereotypical lab settings, using real objects and somewhat plausible contexts. Brown-Schmidt & Tanenhaus (2008) used a non-computer-based task and unrestricted dialogue to examine the developing restriction of the referential domain by the use of linguistic and visual information. Hanna & Tanenhaus (2004) examine visually mediated perspective-taking by having a confederate pose as a cook and using the participant as the cook's assistant. Tracking the gaze of the participant revealed that when the cook named an object he needed, objects close to the cook were only considered if the cook had his hands full. This showed that the participants used a source of visual information to facilitate perspective-taking and restrict the domain of referential targets in order to disambiguate the statement.

However, despite these innovative experiments, we believe that the use of visual information may be unfairly tested using situations which demand the use of visual information. For example, either by demanding references to visual objects, or by presenting visual information on a monitor which participants have to sit in front of. Therefore, it is hard *not* to use the presented visual information, and as such, unsurprising that we find that interlocutors are so good at exploiting visual sources of information. Although there exist many language situations that are inherently visual in their task, for example fetching objects for someone or describing a route, we argue that many common language situations have visual information present, but that the use is not explicitly required. As examples, imagine somebody asking you about what you think of their city, or discussing the wedding couple at a wedding reception. Such situations have available and relevant visual information to help generate appropriate responses (e.g. by referring to some impressive landmark, or the dress of the bride), but the communication seldom forces you make

explicit use of it. We wonder whether the presence of such a "shared visual experience" (Gergle, Kraut & Fussell, 2004) is exploited if it is optional and occurs as part of an unrestricted dialogue.

We report the first results of a breadth-first study on the use of optional visual information in an unrestricted dialogue task. Although the dominant focus of current language—vision research is explicitly on producing referential expressions or resolving the same, we are open to more subtle uses of visual information. Our hypotheses are four:

1) Access to visual information results in more deictic expressions (explicit referencing)
2) Access to visual information inspires more to talk about, resulting in more words per utterance, and/or more utterances per conversation topic.
3) The effects in H1 and H2 will wear off over time, as the novelty of the static image reduces.
4) Utterances produced in the presence of visual information will differ in its information content, as information is offloaded or incorporated to/from the present visual information.

These hypotheses were tested using 48 pairs of participants, discussing 8 topics each, drawn randomly from a pool of 48 topics. The presence of an image (the shared visual information) was manipulated (presence/non-presence). The conversations were transcribed to standard orthographic text and then analyzed.

Our results indicate, at this stage, surprisingly little support for the non-referential use of visual information. Only hypothesis 1, that added visual information would result in more deictic expressions, received support from the statistical analysis ($p < .01$).

We interpret the main result as meaning that the use of visual information when producing or resolving referential expressions is a robust practice in normal language situations, and this is likely to continue even in situations when the use of available visual information is not explicitly required. However, if it is really the case that visual information is employed in situations not involving explicit referential expressions, then the measures tested in this study fail to capture this effect.

References

Altmann, G.T.M. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247-264.

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, 32(4), 643−684.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language and Social Psychology, 23*, 491-517

Hanna, J. E. & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cognitive Science*, 28:105-115.

Kamide, Y., Altmann, G.T.M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye-movements. *Journal of Memory and Language,* 49, 133-159.

Knoeferle, P., Crocker, M.W., Scheepers, C., & Pickering, M.J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. Cognition, 95, 95-127

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.