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.

SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pages 196–197.
Los Angeles, California, 21–23 September 2011.

How deeply rooted it turn-taking?

Jens Edlund
KTH Speech, Music and Hearing
edlund@speech.kth.se

Abstract

This poster presents preliminary work investigating turn-taking in text-based chat with a view to learn something about how deeply rooted turn-taking is in the human cognition. A connexion is shown between preferred turn-taking patterns and length and type of experience with such chats, which supports the idea that the orderly type of turn-taking found in most spoken conversations is indeed deeply rooted, but not more so than that it can be overcome with training in a situation where such turn-taking is not beneficial to the communication.

Wilson & Wilson (2005) propose that turn-taking is grounded in fundamental human cognitive processes, based in part on the observation that orderly turn-taking is present even in forms of dialogue where it need not be for communicative purposes:

> "To our knowledge, no culture or group has been found in which the fundamental features of turn-taking are absent. This is true even when the physical substrate of conversation is radically different from that of ordinary speech, as in the cases of sign language used by the deaf and tactile sign language used by the deaf-blind."

However, personal experience and discussions with colleagues and friends suggest that people's habits during text based chats may provide a counter-example: it is common for people in text based chats to type without waiting for their turn or waiting for a response. From introspection and memory, it seems that people who are quite used to maintaining text based conversations, and in particular those who are used to extended multi-party conversations such as in-game chats and IRC (Internet Relay Chat).

A possible reason for this could be that turn-taking makes little sense in a text-based chat. Typing is slow, and while one participant is typing, all others must sit inactive. When participant hits return and the message is revealed, all others must first read it, then whoever should respond will start typing, and the waiting game starts over. Furthermore, in case there are more than two participants, the issue of selecting the next speaker becomes severely complicated by the lack of gaze and gesture. If on the other hand turn-taking is abandoned, it is quite possible to maintain a conversation with two or more parallel threads, where one speaker narrates a story at the same time as another, so that they can both type simultaneously.

If these speculations are correct, they are compatible with Wilson & Wilson's statement. The turn-taking system we use in spoken interaction is indeed deeply rooted, and is not easily over-ridden even when the interaction is moved to a system in which turn-taking is not strictly necessary, and might even be detrimental. Following sustained use of such systems, however, users may learn more efficient patterns. This would be exemplified by two-party text-based chats.

It is also likely the process will be sped up by extended use of a system where traditional turn-taking is not only difficult but impossible, but that nevertheless functions well, by demonstrating forcefully that other patterns are possible. Multi-party text-based chats would exemplify this.

To explore this line of thinking, a pre-study in the form of a Google Documents questionnaire was sent to 80 people picked from the author's address list. The questionnaire contained questions on text-based chat experience and on turn-taking preferences. 38 people answered the questionnaire, 17 females and 21 males. There were no significant or even noticeable gender differences. All

those who answered had extensive experience with general computer use.

Three open questions were included: "For what purposes do you use text based chat? Please put down an example or two.", "Do you see any similarities or differences in the way you take turns when speaking and when you use text based chats? Please provide a few examples!", and "Do you see any similarities or differences in the way you use text based chats and email? Please provide a few examples!". At the time the answers to the open questions were compiled, 35 people had answered. The two most common purposes mentioned were *to stay in contact* (22/35) and *to ask brief questions* (15/35). The two most common similarities or differences to speech were *turn-taking* (26/35; mention as similarity as well as difference) and *timing* (14/35). The two most commonly mentioned similarities or differences to e-mail were the *level of formality* (22/35; e-mail more formal) and *presence* (12/35; presence required for chat).

The questions of real interest were embedded in a range of different questions about text based chats in order to make them inconspicuous. They were multiple choice questions phrased as follows:

(1) How frequently do you use text based chats? (Daily, Weekly, Monthly, More rarely)
(2) Do you use the multiple user/group chat functions? (No never, Yes occasionally, Yes, regularly)
(3) Do you prefer typing one message, then waiting for your chat partner to type a message, and so on in an orderly manner, or do you just type as you think of things and read whenever there is a response? (I prefer to just type as soon as I think of something, I prefer to take turns, I'm fine with both)

The hypothesis is that the answer to (3) should more commonly be "I prefer to just type as soon as I think of something" with participants who use text-based chats more, who have done it longer, and who are used to multi-party chats (such as in-game chats). "I'm fine with both" answers to (3) are omitted for space reasons, but they occur in all groups to a similar extent.

The answers to (1) and (3) support the hypothesis, in that a much larger proportion of those who use text-based chats often flaunts turn-taking:

	Chats weekly or more	Chats monthly or less
Prefers turntaking	6	8
Flaunts turn-taking	7	1

The same goes for the answers to (2) and (3), in that a larger proportion of those who regularly use multi-party chats flaunts turn-taking:

	No multi-party	Occasional multiparty	Regular multiparty
Prefers turntaking	9	5	0
Flaunts turn-taking	4	2	2

As these initial results seem promising, a larger survey in which a number of flaws revealed in the pre-study are remedied is in preparation, and will be made available to a much larger population. We are also seeking methods to test the results through analysis of chat data or possibly to verify them experimentally. The latter will be difficult, as removing the urge to take turns is seemingly a long process.

Acknowledgements

This work was supported in part by Riksbankens Jubileumsfond (RJ) under contract P09-0064:1-E, Samtalets Prosodi (Prosody in Conversation). Thanks also to those who took the time to answer the questionnaire.

References

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review, 12*(6), 957-968.