Eye-movement Research and the Investigation of Dialogue Structure

Thomas Clermont, Hendrik Koesling, Marc Pomplun, Elke Prestin, and Hannes Rieser¶

CRC 360 "Situated Artificial Communicators"
University of Bielefeld, Bielefeld, Germany

¶The talk will be given by Hendrik Koesling and Hannes Rieser.

Correspondence should be sent to:

Hendrik Koesling | Hannes Rieser

CRC 360 "Situated Artificial Communicators"
University of Bielefeld
P.O. Box 10 01 31
33501 Bielefeld
Germany

Tel.: +49-521-106 6054
Fax: +49-521-106 6011
E-mail: ihkoesli@techfak.uni-bielefeld.de

Tel.: +49-521-106 3666
Fax: +49-521-106 6011
E-mail: rieser@lili.uni-bielefeld.de

Eye-movement Research and the Investigation of Dialogue Structure

Thomas Clermont, Hendrik Koesling*, Marc Pomplun, Elke Prestin, and Hannes Rieser**

CRC 360 "Situated Artificial Communicators"
University of Bielefeld
P.O. Box 10 01 31
33501 Bielefeld
Germany

* `ihkoesli@techfak.uni-bielefeld.de`
** `rieser@lili.uni-bielefeld.de`

Abstract

In this talk we report about eye-movement research and the investigation of dialogue structure as it has been going on at several institutions at the University of Bielefeld (cf. the reports listed in the bibliography). Dialogues, one might argue, can be regarded as sequences of turns in which certain micro- and macro-structures can be distinguished. How can this be related to the investigation of agents' eye movements? Working with empirical data, transcripts, and video tapes of two-person, task-oriented dialogue reveals that agents do not behave as one would expect from the standpoint of abstract semantics and pragmatics. Most researchers do know about that, of course, at least since the Russell-Donellan-Kripke discussion about speakers' meaning and abstract meaning. Nevertheless, it is not trivial to find out what speakers do and to start to develop theories thereof. There has been little research in this area since the 1970s. As we will see, investigation of speakers' meaning is difficult because one has to develop precise methods for the observation of their doings. Although dialogues and video tapes provide only a rough idea of what is going on, let us start from there. The following observations should be uncontroversial with respect to these and similar data:

(1) Speakers select domains of interpretation and use them rather flexibly. This is important if we want to understand the use of definite descriptions, anaphora, and all sorts of relational expressions.

(2) Speakers describe things and situations frequently from an agent-related perspective.

(3) The use of descriptive vocabulary, especially non-literal expressions (tropes) and neologism, is induced by the domain under discussion. In short: Specific domains instigate specific wordings.

(4) The sequence of turns produced in describing the setup of an object depends on the ontological structure an agent "casts" over this object.

(5) Agents coordinate their wording in order to complete tasks more efficiently.

We refer to these observations as the "flexible domain constraint" (1), the "perspective constraint" (2), the "domain-description constraint" (3), the "ontology constraint" (4), and the "coordination contraint" (5), respectively.

If these constraints have some initial plausibility, it seems to be worthwhile finding out more about them. How can we do that? How can we, e.g., get more reliable information about the mechanisms of flexible domain selection? One answer is: We try to find out where an agent's focus of attention is while he produces speech, say the description of an object. The area spotted by attention can be taken as some sort of relevantly singled out domain leaving its traces in the language tokens produced. However, being in the mind, an agent's attention is not directly observable, hence we must look for its nearest equivalent and that is where his eyes rest upon. Perhaps one cannot maintain the latter in general, but it seems to be acceptable for tasks involving the description of objects *seen*. In short, we identify the focus of attention with sequences of clustered foveal fixations.

In our report we describe three experimental studies involving eyetracking and the findings they led to: A 2D-blocks-world study, a 2D-"airplane" study, and, finally, a 3D-"airplane" study. The currently used 3D-setting seems to be the most promising for future research. Our first 2D-study was based on task-oriented dialogues, in which an instructor told a constructor to build up a blocks world as shown in Fig. 1. In the scene used, the instructor had his blocks world presented on a computer monitor (hence 2D). Only he was integrated into the eyetracker-setting shown in Fig. 2.

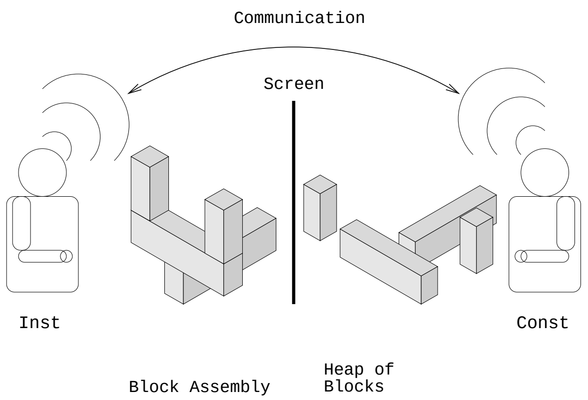

Figure 1: Blocks world

The 2D-study confirmed the intuitively set up constraints reported above. In addition, new and unexpected observations emerged: The instructor's eye movements are usually several construction steps ahead whereas the speech production (the directives produced) "lags behind". We may assume that the instructor's pushing ahead is connected with planning procedures. Hence we call this contraint the "asynchrony of planning and production contraint" (6). In case of production problems, however, especially problems concerning word finding or selection of syntax patterns, the focus of attention remains fixed on the

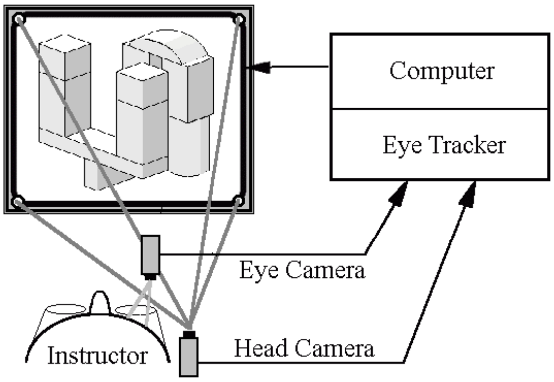

Figure 2: 2D-eyetracker setting involving the instructor of the task-oriented dialogue

object currently investigated, the object which gives rise to the problem. We call these cases "blocked focus movement" (7). Moreover, agents can coordinate their foci of attention: They do so by verbally controlling their eye movements ("coordination of focus contraint" (8)).

The first 2D-study also reveals the connection between focusing and discourse structure, especially the initiation of new turns by the instructor: In the default case where neither he nor the constructor face major problems, the instructor can proceed unimpeded and produce his next turn following the ontology constraint. At the heart of his new move there will, of course, have to be complex attitudinal states concerning the progress of the task on the constructor's side. (Although being a very interesting aspect, we will not further discuss mutuality and common ground in our report.) In non-default cases various sorts of "side tracks" (repairs, side sequences, back-tracking) are produced.

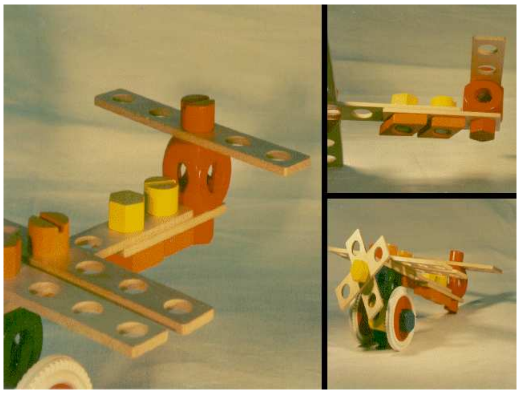

Figure 3: 2D-representation of various perspectives of a toy airplane serving as the basis for the instructor's directives

The second 2D-study (see Fig. 3) revealed that the perspective contraint and the ontology constraint are central for focus movement and discourse production. It also matters whether agents can freely rotate their objects of interest, since rotation of objects and eye movements are closely related. Furthermore, eye movements also act as a sort of anticipatory device for word selection.

Both 2D-studies to be reported yield only imperfect data concerning the coordination between instructor and constructor whilst they are carrying out their task. Therefore we have tried to develop a 3D-setting where the instructor's and the constructor's eye movements can both be recorded and matched with their speech production. As far as we know, this is the first 3D-eyetracker setting of this kind in operation (see Fig. 4). Using it we want to get a clearer understanding of the constraints (1)— (8) mentioned above. We also hope to show a video sequence in which the instructor's and the constructor's activities are integrated and provide an impression about how they organize their interaction. This, however, will very much depend on whether we can overcome the technical problems with the eyetracker equipment.

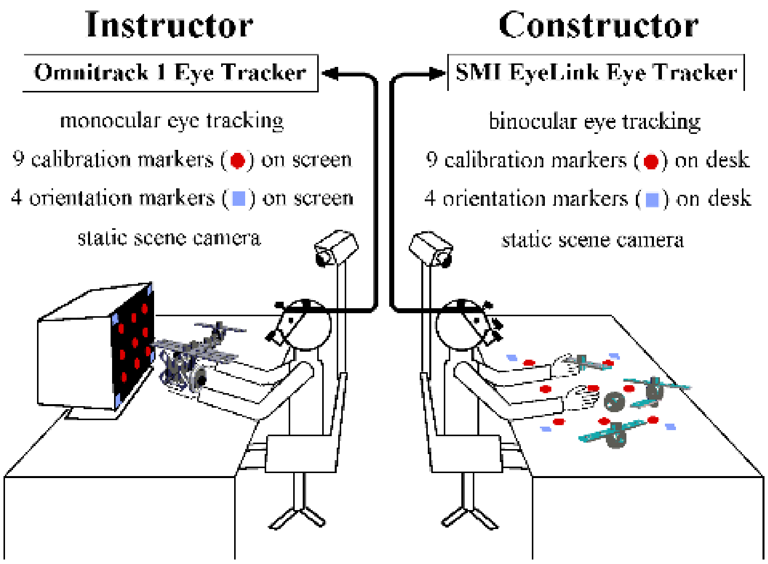

Figure 4: 3D-setting with eyetracking equipment for the instructor and the constructor

References

General:

[Asher, 1993] Asher, N. (1993). *Reference to Abstract Objects in Discourse.* Kluwer Academic Publishers.

[Chierchia, 1995] Chierchia, G. (1995). *Dynamics of Meaning—Anaphora, Presupposition, and the Theory of Grammar.* Chicago UP.

[Clark, 1996] Clark, H. H. (1996). *Using Language.* Cambridge UP.

[Just and Carpenter, 1987] Just, M. and Carpenter, P. (1987). *The Psychology of Reading and Language Comprehension.* Allyn & Bacon.

Selected Research Reports:

[Clermont et al., 1995a] Clermont, T., Meier, C., M., P., Prestin, E., Rieser, H., Ritter, H., and Velichkovsky, B. (1995a). Augenbewegung, Fokus und Referenz. Technical Report 95/8, SFB 360 “Situierte Künstliche Kommunikatoren”, Univ. of Bielefeld.

[Clermont et al., 1995b] Clermont, T., Meier, C., Pomplun, M., Prestin, E., and Rieser, H. (1995b). Focus and Reference. Videofilm on Eye Movements and Focussing. Videofilm. SFB 360 “Situierte Künstliche Kommunikatoren”, Univ. of Bielefeld.

[Essig, 1998] Essig, K. (1998). Messung von binokularen Augenbewegungen in realen und virtuellen 3D-Szenarien. Diplomarbeit, Technische Fakultät der Universität Bielefeld.

[Heydrich and Rieser, 1994] Heydrich, W. and Rieser, H. (1994). Public Information and Mutual Error. In Kunze, J. and Stoyan, H., editors, *KI-94 Workshops*, pages 110–2. Gesellschaft für Informatik: Saarbrücken. Workshop “Modellierung epistemischer Propositionen”.

[Heydrich and Rieser, 1995] Heydrich, W. and Rieser, H. (1995). Public Information and Mutual Error. Technical Report 95/11, SFB 360 “Situierte Künstliche Kommunikatoren”, Univ. of Bielefeld.

[Meier and Rieser, 1995a] Meier, C. and Rieser, H. (1995a). Modelling Situated Agents’ “Reference Shifts” in Task-Oriented Dialogue. In Dreschler-Fischer, L. and Pribbenow, S., editors, *KI-95 Activities: Workshops, Posters, Demos*, pages 318–21. Gesellschaft für Informatik: Bonn.

[Meier and Rieser, 1995b] Meier, C. and Rieser, H. (1995b). Modelling Situated Agents’ “Reference Shifts” in Task-Oriented Dialogue. Technical Report 95/11, SFB 360 “Situierte Künstliche Kommunikatoren”, Univ. of Bielefeld.

[Meier and Rieser, 1996] Meier, C. and Rieser, H. (1996). Perception, Focus and Resolution of Metonymy. In Gibbon, D., editor, *Natural Language Processing and Speech Technology. Results of the 3rd KONVENS Conference*, pages 305–9.

[Meyer-Fujara and Rieser, 1997] Meyer-Fujara, J. and Rieser, H. (1997). Zur Semantik von Repräsentationsrelationen. Fallstudie Eins zum SFB-Flugzeug. Technical Report 97/7, SFB 360 “Situierte Künstliche Kommunikatoren”, Univ. of Bielefeld.

[Pomplun et al., 1997] Pomplun, M., Rieser, H., Ritter, H., and Velichkovsky, B. (1997). Augenbewegungen als kognitionswissenschaftlicher Forschungsgegenstand. In Kluwe, R., editor, *Strukturen und Prozesse intelligenter Systeme*. DUV.

[Pomplun et al., 1996] Pomplun, M., Ritter, H., and Velichkovsky, B. (1996). Disambiguating Complex Visual Information: Towards Communication of Personal Views of a Scene. *Perception*, 25: 931–948.

[Pomplun et al., 1994] Pomplun, M., Velichkovsky, B., and Ritter, H. (1994). An Artificial Neural Network for High Precision Eye Movement Tracking. In Nebel, B. and Dreschler-Fischer, L., editors, *Lecture notes in artificial intelligence: Proceedings KI-94*, pages 63–9. Springer.

[Rieser, 1997] Rieser, H. (1997). Repräsentations-Metonymie, Perspektive und Koordination in aufgabenorientierten Dialogen. In Umbach, C., Grabski, M., and Hörnig, R., editors, *Perspektive in Sprache und Raum*, pages 1–26. DUV.

[Stampe, 1993] Stampe, D. (1993). Heuristic Filtering and Reliable Calibration Methods for Video-Based Pupil-Tracking Systems. *Behavioral Research Methods, Instruments, and Computers*, 25: 137–42.

[Velichkovsky et al., 1995] Velichkovsky, B., Pomplun, M., and Rieser, H. (1995). Attention and Communication: Eye-Movement-Based Research Paradigms. In Zangemeister, W. H., Stiehl, H. S., and Freksa, C., editors, *Visual Attention and Cognition*. Elsevier.