KogWis 07

Fachrichtung Psychologie

Gesellschaft für Kognitionswissenschaft

Cognitive Science 2007

***Proceedings of the 8th Annual Conference of the Cognitive Science Society of Germany*.**

Zimmer, H. D.

Frings, C.,

Mecklinger, A.

Opitz, B.

Pospeschill, M.

Wentura, D. (Eds.)

Saarbrücken

URL: http://psydok.sulb.uni-saarland.de/volltexte/2008/1185/

URN: urn:nbn:de:bsz:291-psydok-11856

Coordination in dialog: Alignment of object naming in the Jigsaw Map Game

PETRA WEIß[A], THIES PFEIFFER[B], GESCHE SCHAFFRANIETZ[A] AND GERT RICKHEIT[A]

[A]Fakultät für Linguistik und Literaturwissenschaft/SFB 673
[B]AI & VR Lab, Technische Fakultät
Universität Bielefeld[1]

People engaged in successful dialog have to share knowledge, e.g., when naming objects, for coordinating their actions. According to Clark (1996), this shared knowledge, the common ground, is explicitly established, particularly by negotiations. Pickering and Garrod (2004) propose with their alignment approach a more automatic and resource-sensitive mechanism based on priming. Within the collaborative research center (CRC) "Alignment in Communication" a series of experimental investigations of natural face-to-face dialogs should bring about vital evidence to arbitrate between the two positions. This series should ideally be based on a common setting. In this article we review experimental settings in this research line and refine a set of requirements. We then present a flexible design called the Jigsaw Map Game and demonstrate its applicability by reporting on a first experiment on object naming.

Motivation

"[H]umans are 'designed' for dialogue rather than monologue" – this is one of the central arguments brought forward by Garrod and Pickering (2004, p. 8) emphasizing the necessity for research on language use in interactive human communication. A central mechanism of successful communication is coordination. When people interact and communicate they need to tune their utterances on different levels, amongst which the lexical level plays a prominent role. The central questions are now, how coordination is achieved and how people build up shared knowledge during a dialog.

In psycholinguistics two main approaches exist addressing these questions. The collaborative model (Clark, 1996) proposes the idea of people building up a common ground, employing an explicit grounding process, e.g., by negotiation. In contrast, the notion of alignment (Pickering & Garrod, 2004) supposes that conversational partners build up an implicit common ground in an automatic and mechanistic way and fall back on utilizing full common ground only in cases of misalignment.

To develop an adequate experimental setting is of central importance for the CRC "Alignment in Communication", where besides lexical coordination also questions of coordination of sentence structure, spatial reference, prosody, gesture and gaze will be examined in a series of communication experiments. Data collected in these experiments will be used to build up a central corpus of interactive language data to function as the empirical basis of research in the CRC.

---

[1] Address: Dr. Petra Weiß, SFB 673, Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, Postfach 10 01 31, D-33501 Bielefeld, email: petra.weiss@uni-bielefeld.de, web: www.sfb673.org. Funded by German Research Foundation (DFG) within the Collaborative Research Center 673 (SFB 673) "Alignment in Communication". The authors want to thank Sara Maria Hellmann for her support in data analysis and Ulrike Well and the staff of the media lab (SCM) of Bielefeld University, especially Nora Johannesmann, for their technical support.

We discuss both psycholinguistic approaches and present successful experimental designs used in this line of research. We specifically highlight their particularities, but also their shortcomings with respect to our needs. This results in a set of requirements for an experimental design appropriate to investigate language coordination in interactive conversation in a natural but controlled way. Consequently, we propose the Jigsaw Map Game as a flexible design meeting these requirements and present first results from an experiment on the coordination of object naming.

Theoretical background

The collaborative model

Clark's (1996) idea is that language use is a joint action. Dialogical communication is a dynamic action. It is generated jointly by communicative partners and only successful when the action is coordinated. This way, language can only be analyzed adequately considering its social aspect. The basic setting for language use is spoken face-to-face conversation. Clark motivates this by giving three reasons: Spoken language is the most universal form of language use in all civilizations; there is no need for special skills in normal conversation and spoken language is essential in language acquisition. Clark and Brennan (1991) characterize face-to-face conversation by immediacy, no use of media leading to evanescence, recordlessness and simultaneity as well as full control of the participants what gets done and how. The less of these features are fulfilled the more demanding the communication between the interlocutors should be because they need more specialized skills and procedures to communicate adequately.

To the question, how people get coordinated to communicate successfully, the collaborative model proposes the idea of a common knowledge base, called *common ground*. "Two people's common ground is, in effect, the sum of their mutual, common, or joint knowledge, beliefs, and assumptions" (Clark, 1996, p. 93). Especially the idea of mutuality is very important. Every joint action is rooted in this common information, but only in the part people believe they share with others. What really is part of the common ground in a community can only be evaluated on a meta-level, but not by the community members themselves. Thus, there only exist individual assumptions about what the common ground comprises; this may lead to misunderstandings that have to be resolved explicitly.

In view of the fact that communication is a daily and natural human activity, the idea of conversation as a joint activity and the proposal of face-to-face interaction as the basic setting for language use seem plausible. On the other hand, the necessity to continuously update common ground information seems to be costly. The question arises, whether the explicit negotiation of information within the grounding process – though it is a typical phenomenon in conversation – really is a basic process. These and other critical points associated with the collaborative model led to the development of the alignment approach.

The notion of alignment

Pickering and Garrod (2004) suggest that dialog communication takes place easily and smoothly because basic interactive processing mechanisms permit coordination and adjustment of communicative partners in a cognitive less demanding way than explicit negotiation. They expand the preceding output/input coordination principle (Garrod & Anderson, 1987) to a mechanistic approach of interactive language processing emphasizing the necessity of aligned representations between partners on different linguistic levels.

The meaning of mechanistic in this context is twofold. On the one hand the authors propose priming as the central interactive processing mechanism leading to a coordinated knowledge base, i.e., a body of aligned representations. Simultaneous alignment on different linguistic levels of representation (phonetic, phonological, lexical, syntactic, semantic and the level of situation models) is based on the idea of percolation between the different levels, i.e., alignment at one level supports alignment on other levels (Garrod & Pickering, 2004; see also Branigan et al., 2000). On the other hand the approach is mechanistic because under default conditions coordination takes place in an automatic way without conscious control of the interlocutors.

To motivate their approach, Pickering and Garrod (2004) point out that the explicit computation of common ground is computationally too costly to be a realistic assumption for real life dialogs. Instead they propose that dialog participants usually only employ an *implicit* common ground – knowledge that is available to both speaker and hearer, but that does not include knowledge about the cognitive state of the other agent(s). Implicit common ground is usually inferred from information sources that are available to all agents (the immediate linguistic and situational context) and therefore provides a good approximation of *full* common ground, as Pickering and Garrod call by contrast the idea of common ground in the collaborative model. It is important to note that the authors do not totally deny the concept of full common ground and explicit grounding processes during conversation. But they advance the view that these concepts and processes cannot be the basic mechanisms for coordination in communication.

Based on the theoretical considerations so far our goal is to examine the question, how – under default conditions – coordination in dialog is reached. Before we will describe our own experimental paradigm developed to investigate into this question, we will give a short overview over some relevant experimental paradigms used in research on interactive language processing.

Language processing in interactive discourse

The empirical investigation of language processing in discourse has developed around a relatively small number of experimental situations (Garrod, 1999; Schober & Brennan, 2003). As language use in natural conversations cannot readily be controlled, some experimental designs have been developed to elicit semi-spontaneous dialog situations where some degree of control over the topic of conversation is possible.

Referential communication task studies

The referential communication task is primarily used to investigate lexical coordination. In an early study by Krauss and Weinheimer (1996) pairs of participants describe to each other a sequence of abstract shapes. The participants have fixed roles, one describes and the other tries to identify the figure. Krauss and Weinheimer demonstrate that the convergence on shortened names for low-codability shapes in the course of a conversation depends on the possibility of direct (verbal) interaction between the interlocutors, especially on getting concurrent feedback. Their results suggest that the referential process really depends on coordination. Clark and Wilkes-Gibbs (1986) show that the common ground between interlocutors is established as a consequence of coordinated actions among the interacting individuals across multiple linguistic exchanges and that this knowledge is consulted for the ongoing conversation.

Brennan and Clark (1996) suggest that for lexical coordination partners establish a conceptual pact during conversation, i.e., a temporary flexible agreement by partners to refer to and conceptualize an object in the same way. A conceptual pact is established by grounding; the partners mark having reached a conceptual pact by reusing the same or similar expressions in the ongoing conversation. Conceptual pacts are not restricted to a single dialog move; the more well established a pact is (the more often they have used it to refer to an object) the more likely it is to persist. In contrast to full common ground, it is not based on an explicit partner model, but may be shaped and maintained by the partner's feedback.

On the whole, results on default use of common ground information are partly inconsistent (e.g. Buhl, 2001). Therefore, the question arose under which conditions speakers are able or willing to consider common ground information. One important factor concerns the communicative situation established in the experiments. Experimental results show that speakers are more sensitive to the needs of visually co-present naïve listeners than to confederates (Schober & Brennan, 2003). This suggests, first, that the presence of a conversational partner with real needs matters and second, that people are sensitive to the behavioral naturalness of that person. These effects might be even stronger if interlocutors are allowed to interact in a real and natural way, when they are not restricted to speaking, but can nod, smile, point, gaze at each other and exhibit and place things (Clark & Krych, 2004). This being the case, employing confederates or not even using a second participant may result in findings that do not reflect the behavior in natural dialog situations.

Results gained within the referential communication paradigm illustrate that language processing is highly sensitive to the constraints imposed by interactive dialog, especially the importance of feedback and real interaction between communicative partners. However, whereas the referential communication task allows for a detailed analysis of referential processes, there always is a fixed role allocation between the communicative partners. One gives information and has a leading role and the other processes this information and has to fulfill a special task indicating correct or wrong interpretation, or one of them is even a confederate of the experimenter.

Maze game studies

In the maze game studies (Garrod & Anderson, 1987) two participants located in different rooms have to interactively get through a maze presented to them individually on a computer screen. The players, who are in audio contact with each other, can only solve the task by coordinating their moves via speech. They have to find a strategy of moving to positions where their partner has switch boxes in order to allow them a free route to the goal. As a result, the game elicits relatively free dialog sequences containing repeated location descriptions. Especially the analysis of how these descriptions develop during the course of the game leads to a number of insights about coordinated language use and interpretation (Garrod, 1999).

Though the dialogs turn out to be extraordinary varied across the sample of players, each description can be classified according to one of four basic schemes corresponding to a special combination of spatial conception of the maze together with a description lexicon. According to this classification, Garrod and Anderson (1987) analyze how the references developed during the course of the dialogs. Despite of the great variations across the whole corpus of dialogs any pair of interlocutors is very consistent in their choice of descriptions in any stretch of dialog. Conversationalists not only collaborate in establishing isolated references but also in formulating local description languages and contextual unambiguous dialog lexica. Furthermore, coordinating on a common description language is not simply a matter of sticking with the first reasonable scheme that emerges, but involves a much more extended history of development during which interlocutors explore different schemes in a coordinated fashion over a period of time. A central point of the analysis is that explicit negotiation does not seem to play an important role in either establishing a common description language or in fixing the language over the subsequent dialogs. If at all, negotiation can only be observed in later rounds of the game.

On the whole, the maze game studies highlight some of the ways in which language processing in a task allowing relatively free verbal interaction is affected by the demand of consensus in dialog. With both players sitting in different rooms, each of them presented with the maze on his/her monitor, the nature of the game demands that, though they have to cooperate, each of them pursuits his/her own goal. They have no common interaction space and no consistent conception of the world emerges during conversation. In contrast to the referential task there is no role differentiation between the partners resulting in a more interactive communication.

Map task studies

The map task was originally developed by Anderson et al. (1984) to explore properties of effective vs. ineffective communication in a pedagogic context. Subsequently, it has been used to investigate into a whole range of dialog-processing issues from reference and speech articulation to video mediated communication (Doherty-Sneddon et al., 1997).

Each participant has a map of an imaginary island containing a number of labeled landmarks. One participant has a route marked across the island, whereas the other just has the map containing the landmarks without the route. The task of one player is to communicate the route to the other player, so that she can draw the route on her map.

This basic setting offers the possibility of a lot of interesting manipulations, e.g., with respect to incompatibilities concerning the number or arrangement of the landmarks on the two maps. This allows to vary the degree of overlapping knowledge between communicators and to explore how they use different referential forms to signal their relative knowledge states. Particularly with this task an independent control of communicative success is possible by comparing the route drawn by the instruction follower with the route given on the map of the instruction giver.

The map task has contributed particularly in two ways to understanding of referential communication. First, in terms of how communicants deploy different linguistic devices to communicate about shared and unshared information (Anderson & Boyle, 1994). Second, in terms of how they combine multi-modal information in coordinating their productions and interpretations. Boyle et al. (1994) find that communicators who can see each other need much less amount of speech to convey the same amount of information as those who can not. Furthermore, communicators often look at each other during dialog segments where they are having problems in communicating, typically when there are incompatibilities between the two maps. Being able to monitor the face and the gaze of each other may help managing the exchange. Anderson et al. (1997) show that this monitoring does not reflect a moment-by-moment modeling of the listener’s actual use of the visual channel, but that it seems to reflect a more general assessment of the state of the mutual intelligibility of the conversation at that time. This result is consistent with the findings from the maze game studies, where communicators cooperate to establish a mutually acceptable description scheme but without explicitly modeling their partner’s knowledge state indicating that implicit processes are basic for interactive communicative coordination (Garrod, 1999).

Contrary to the maze game, in the map task again there is a clear role allocation of the partners like in the referential communication paradigm. The advantage here is that communication takes place about a special part of the world showing a landmark structure offering the possibility for structured spatial descriptions in a variety of communicative situations ranging from face-to-face dialogs to video- and computer-mediated communication. On the other hand, again no shared part of the world emerges during conversation.

Based on these considerations it becomes evident that for the examination of basic processes in natural face-to-face conversations none of the paradigms reported is sufficient on its own.

To examine the question how coordination in dialog is reached under default conditions requires a trade-off between natural communication and experimental control which is a big methodological challenge (Schober & Brennan, 2003). We try to overcome this problem by developing the *Jigsaw Map Game*, an experimental setting based on the referential communication task, combining elements of the maze game and the map task.

The Jigsaw Map Game

Following the lines of thought presented, we identify several objectives for our design of an experimental setting. The participants should

- … be naïve and not confederates.
- … be engaged in a face-to-face situation.
- … be able to directly perceive each other's behavior.
- … be able to communicate using several modalities.
- … be on equal terms regarding their roles.

The setting should allow for the most natural communication possible between the interlocutors under experimental conditions. At the same time means to control important parameters are needed. One should be able to

- … channel the course of the dialog.
- … control the task knowledge relevant for each participant.
- … induce specific conditions during the dialog.
- … assess data from several sensor devices, including eye tracking.

Encouraging natural face-to-face interaction

One of our primary objectives is to observe natural communication in dialog, so we decided to have a basic setting without any technical mediation, which is still flexible enough to allow mediation if needed. Our setting constitutes a face-to-face situation including physical objects which can be touched and handled by the participants. This promotes the presence of the participants within the setting (Lombard & Ditton, 1997). The objects should vary in shape, size, color or function to ensure a rich use of verbal and non-verbal communication. The interaction between the interlocutors and the objects happens in a distinct interaction space within a restricted area, e.g., defined by a desktop. On the technical level, the visual co-presence of objects and interlocutors in a face-to-face situation allows for multi-perspective video recordings as well as in-depth investigations using eye tracking, e.g., following the visual world paradigm.

Promoting multi-modal communication

From previous experiments (Kranstedt et al., 2006) we know, that a setting with a relatively small interaction space, may encourage a disproportionately high use of pointing or grasping gestures. This should be balanced in favor of more elaborate verbal object descriptions. For instance, the interlocutors can be made to introduce the objects into the interaction space, which is what we did in our first experiment. In the course of our setting each interlocutor has to ask her partner for specific objects she does not have access to. Each one can use verbal descriptions, iconic gestures or references to objects already in the setting sharing certain attributes.

Defining the task

With these decisions being made, most of the stated objectives have already been met. Previous work in the course of the CRC "Situated Artificial Communicators" had concentrated on complex construction tasks and tasks on object placement (Weiß et al., 2006). Inspired by the map task we decided to employ a combination of both types of tasks: the placement of objects according to a complex prescribed map (Fig. 1). The map is designed in such a way that some objects stand out as landmarks because of seize. These objects define the critical objects for the object naming task under examination in the experiment presented.

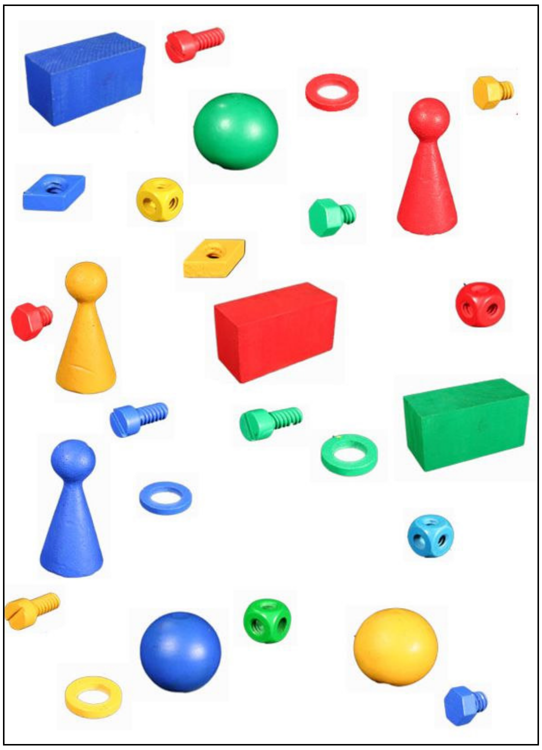


Fig. 1 Example for the full layout of the objects on the table: The critical objects for the naming task form landmarks because of their sizes and are expected to be referred to more often in the sense of spatial reference objects.

The goal of the game is that the communicators cooperatively build the common object arrangement shown on the map in Fig. 1. The cooperative character of the game emerges by the fact that none of the participants knows the complete object arrangement. Each partner only gets partial information concerning the constellation of maximal three objects (two already placed and a new one; see Fig. 2) at a time. She has to communicate to the partner what object the partner should pick next and where it has to be placed. This way the objects are related to each other, as in the construction task, but their relationship is focused on spatial relations as in the map task, leaving aside functional relations or role attributions. By placing the respective new object relatively to objects already placed we can differentiate between references identifying an object as a direct target object that has to be placed next and references identifying an object as an indirect reference object for the placement of the next target object.

Controlling the flow of the game and balancing out the roles of the participants

To control the task knowledge of each participant and at the same time channel the course of the dialog, we finally came up with the *Jigsaw Map Game*. Every participant is given an ordered set of small pieces of the map (Fig. 1), each showing only a small set of objects.

These pieces are carefully designed as in a jigsaw puzzle: They show exactly one new object in relation to several existing objects already in the interaction space (Fig. 2).

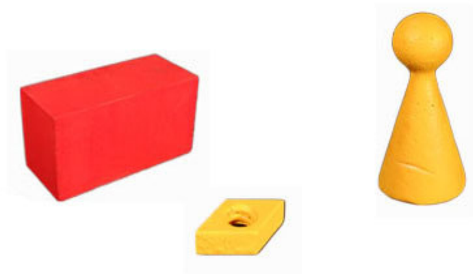
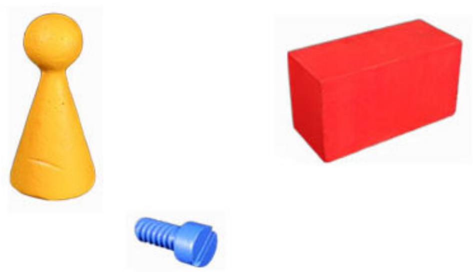
A	B
 1) A starts by drawing a piece of the puzzle showing a cuboid, asks B to fetch the "red box" from B's object pool and to position it within the interaction space. A's reference to the "red box" is a direct reference.	 2) Now B takes the next card and requests A to fetch and place a green bolt. In her instruction B might refer to the red cuboid as a landmark. This reference is an indirect reference.
 3) The next item A commands B to place is a yellow rhombus.	 4) Then it is B's turn to instruct A to place the yellow token.
 5) A then requests B to put the blue bolt into position.	 6) B then asks A to situate the red bolt.

Fig. 2 The flow of the game is controlled by providing small amounts of local knowledge. Each piece of the map (examples 1 to 6) shows a maximum of two known and exactly one new object. Please note that the two sets of cards show the setting from different perspectives: While in turn 5 for A, the blue bolt is right before or below the yellow token, for B the blue bolt is to the left and behind or above the yellow token.

The participants now draw their pieces of the puzzle in turns. Each turn, the one with the new part of the map has to instruct the other one to get the new object out of a personal box of objects (each one has a set of pieces of the puzzle describing new objects to be found in the personal box of the other one) and position that new object in relation to the objects already placed in the interaction space. Once this has been accomplished, they change their roles and the one having placed the last object draws the next piece of the map. A part of the flow of the game is shown in Fig. 2.

Set-up and recording technique

We want to conclude the presentation of the design of the Jigsaw Map Game with a concrete example set-up, as it has been used and tested in our first experiment. The two participants are placed face-to-face on the longer sides of a standard office desk (see Fig. 3). In between them, an interaction space is defined by two yellow lines. The full interaction space has been carefully chosen to be totally within grasping range of both interlocutors. To each side of the interaction space there is room for the piles of new and used pieces of the map. Aside each participant there is a small box with the personal selection of objects.

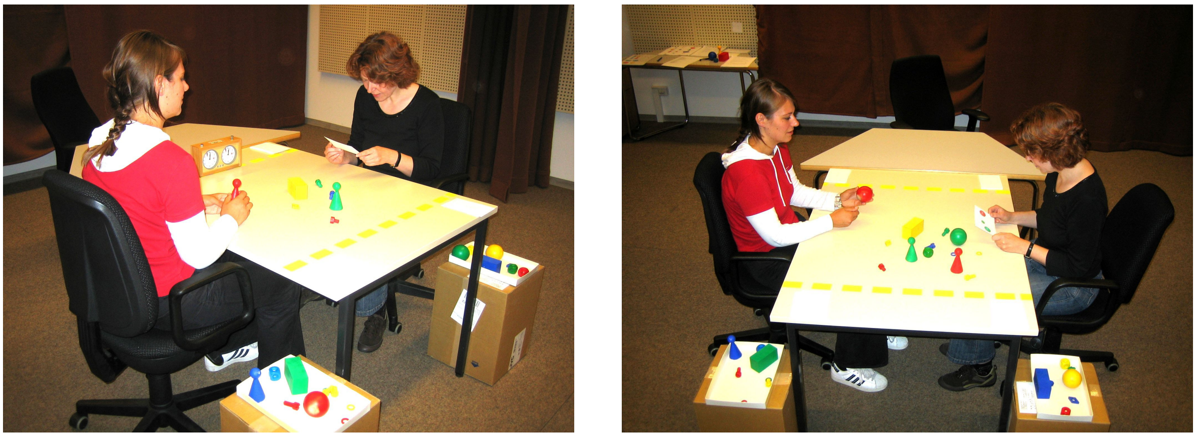


Fig. 3 The left picture has been taken from one of the cameras at the side of the setting with a frontal view on one of the interlocutors. On the right the perspective of the bird's eye camera showing the interaction space and both object pools has been taken.

The experiment has been recorded on video tape (Mini DV, Sony VX2100) from three perspectives, one corresponding to a bird's eye view (Fig. 3, right) giving a general overview of the setting and the other two from the front left and front right sides providing a uncovered view on gaze and mimics from a single participant (Fig. 3, left). Audio has been recorded separately using two microphones (KM-184), one hanging on the ceiling from above the interaction space and one placed in front of the table.

Alignment of object naming in the Jigsaw Map Game

The Jigsaw Map Game has been applied in a first experiment on the coordination of object naming (Schaffranietz, 2007). The central question of the experiment was on how coordination of object naming takes place under default conditions in an "ideal" face-to-face dialog. For such a situation we propose that coordination of object naming takes place fast and easily in a resource gentle way as put forward by Pickering and Garrod (2004). Furthermore, we were interested in the question how coordination occurs in a situation deviating from such an ideal face-to-face dialog regarding two aspects, i.e., differing *knowledge states* of the communicative partners and *cognitive load* during conversation. These considerations led us to the following design.

Experiment design

Concerning differing *knowledge states* between the communicative partners we were interested in the impact on the use of object references. The variation of the knowledge states was operationalized by priming the naming of all object shapes at the beginning of the experiment by learning. For the set of landmark objects, further referred to as the critical objects, in one condition both participants learned either the first or the second name, in a second condition A learned the first and B the second name (see Fig. 4). For all the other objects they learned one identical name.

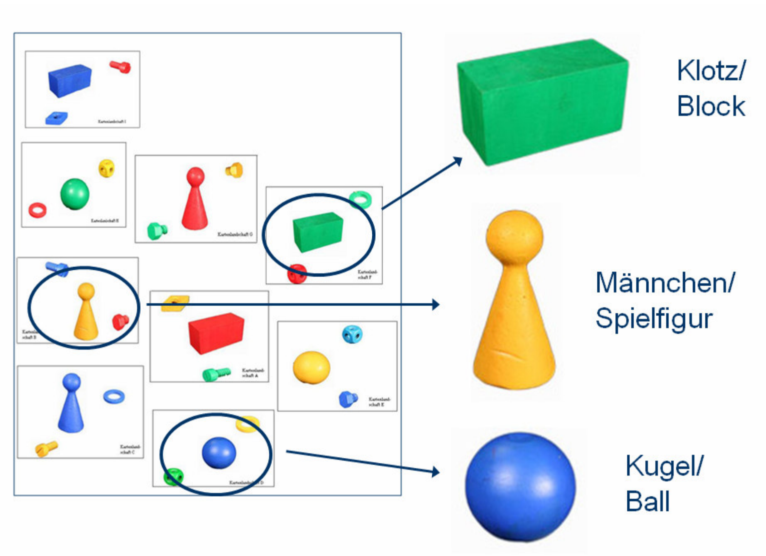


Fig. 4 Certain landmark objects define the critical objects for the object naming task. In one condition the participants learned different names (e.g. Klotz vs. Block) for the objects.

Causing the participants to explicitly learn names for the object shapes we can control the use of lexical references. Varying the learned names for the critical objects in the condition with different knowledge states we can provoke an explicit deviation of the preferred naming between the partners. Furthermore we are able to categorize the object names used during the dialogs into three categories *learned* name, name learned by the *partner* and *idiosyncratic* name, i.e., names not learned by any of the participants. With this procedure we can get to know in a controlled way which naming prevails. In this study, we chose common and therefore highly primed names for all objects so that the critical objects did not attract the attention of the participants solely because of a special or even funny naming. To vary the naming of the critical objects we chose two possible names with comparable plausibility (see Fig. 4). The plausibility to use the chosen names for the objects had been ensured by a small survey conducted with ten students which had been asked to name the objects spontaneously.

Regarding the influence of *cognitive load* on communication in dialog, Horton and Keysar (1996) report, that time pressure reduces the capabilities of speakers to take into account common ground information and behave partner oriented in their language production. On the other hand, results of Rummer et al. (2007) using the same experimental paradigm suggest, that the effects of time pressure on language production in dialog should be qualified by the degree of the interactivity of the communicative situation under consideration. On the basis of these differing results we were interested in the effect of time pressure on communication within the frame of the Jigsaw Map Game.

Cognitive load was varied by confronting one half of the participating pairs with time pressure operationalized by the experimental instruction and the presence of a chess clock, which the participants had to use during the game.

The two factors were varied between experimental pairs as between-subjects factors resulting in a 2x2 experimental design. To each of the four conditions, combining *knowledge state identical/different* and *cognitive load yes/no*, two male and two female pairs were assigned, i.e., 8 naïve participants for each condition. Thus altogether we tested 16 pairs and 32 participants respectively.

Assumptions

Our basic assumption is, that in an ideal dialog situation (represented here by identical knowledge states and no extra cognitive load) communicative partners should coordinate on common object names in an automatic and economic way. Participants should most frequently produce highly primed names for all objects, i.e., the names learned at the beginning of the experiment. Furthermore we expected nearly no explicit negotiations concerning the object references.

Furthermore, we are interested in the effect of differing knowledge states, i.e., differences in the naming of the critical objects. How do participants cope with the induced problems? According to the notion of alignment participants should frequently produce object names introduced by their conversational partner. Especially when referring to a critical object indirectly, in the sense of a spatial reference object, we expect participants to produce names previously used by the partner either for the very same object or for an object of the same shape. This effect may be modified by cognitive load imposed by time pressure.

Procedure

16 pairs of mutually unacquainted interlocutors of the same sex, mostly students, with an average age of 28, were tested in the multi-media laboratory of the Faculty of Linguistics and Literature Studies at Bielefeld University. They were paid for their participation in the experiment. One session took between 30 and 45 minutes.

The participants played the Jigsaw Map Game. Their specific task was to arrange interactively 27 unique objects (4 small bolts, 4 bigger bolts, 4 dies, 4 rings, 3 tokens with a cone shaped body and a sphere as a head, 3 cuboids, 3 spheres and 2 rhombuses) within the interaction space. The objects of the same shape and size were unique in their colors (red, yellow, green and blue). The order of the appearance of the objects on the pieces of the map and the colors of the objects were controlled and varied between sessions.

In the condition with different knowledge states, each participant learned a different name for the shape of the cones ("Männchen" or "Spielfigur"), cuboids ("Klotz" or "Block"), and spheres ("Ball" or "Kugel"). These objects formed the set of critical objects relevant for the analysis (see Fig. 4). The names had been ranked highest in a pretest where ten people had been asked to name the objects used in the experiment. The partners were separated during the instruction. Each was shown a tailored list of all objects and their naming within the game and told to familiarize with the objects.

Subsequently they were given a written instruction with a description of the game in two variants adapted to the condition cognitive load. They were also given time for questions to make sure they did fully understand their task. The instruction emphasized that the participants were requested to refrain from using pointing or grasping gestures and to concentrate on verbal descriptions. The kind of verbal communication had deliberately not been restricted. It has also been stressed, that the placement of the objects has not to be taken too picky as we wanted to reduce lengthy discussions and frustrations.

After the instruction participants were again shown the list of objects and requested to memorize their names. This was then checked in a first test where the object names had to be written down. This was immediately followed by a second test where the physical objects had to be identified verbally. The instruction was completed when the participants correctly remembered all object names. As a consequence the learned names in this experiment were primed strongly.

At the beginning of the game both participants were assigned their seats. Before they started a short example of two interaction turns was given by the experimenter, both to ensure they understood the task and to make a context switch from the object learning task.

The procedure of the game followed the description given above. After the game had been finished, the participants again were asked to complete the questionnaire about the object names. This was done to check, whether they were using the learned names or the ones established during the game. After this, they were asked to fill out a demographic questionnaire. At the end of the session participants again were asked to fill out the questionnaire about the object names, but this time they were instructed to write down the names they would normally have used.

Data preparation

We focus here solely on an annotation of how participants refer to the critical objects within their turns. First, the occurrences of the critical object names were classified according to their target. When a naming referred to the object to be placed, the reference was classified as being *direct*; when the naming referred to the object as reference object, the reference was counted as *indirect* (see Tab. 1). Second, the object references were classified according to the categories *learned* (names previously learned by the participant), *partner* (names previously learned only by the partner) and *idiosyncratic* (names learned by none of the participants).

Tab. 1 Excerpt from the session of pair number 4, the first nine turns of twenty-seven are shown. This game was started by VP8.

Pair 4	**VP7**		**VP8**	
Turn	***direct***	***indirect***	***direct***	***indirect***
1 cuboid			*"Klotz"*	
2		*"Klotz"*		
3				*"Klötzchens"*
4 token	*"Männchen"*	"Klotz"		
5				*"Kegelmännchen", "Klotz"*
6		*"Männchen"*		
7 token			*"Männchen"*	*"Männchen"*
8		*"Männchen"*		
9				*"Männchen"*
…				

Tab. 1 shows an example how critical object references were counted and classified. In the excerpt of the dialog of pair number 4 VP8 introduces *"Klotz"* for the cuboid in a direct reference in Turn 1. VP7 then aligns to this by picking-up the naming in the indirect reference (at this moment a bolt is to be placed). Interestingly VP8 then alters the naming into *"Klötzchen"*, which is the diminutive form of *"Klotz"*. But this does not affect VP7, who consequently sticks to *"Klotz"* in Turn 4, where the name *"Männchen"* for the newly introduced token is used. VP8 at first does not adopt this name, but is more specific by using *"Kegelmännchen"* (cone-token), still in later referential acts both participants seem to have aligned on using *"Männchen"* for the token.

Results and discussion

Tab. 2 provides an overview of the total number of references to critical objects split along categories and conditions. For a statistical analysis of the frequencies of the critical object references in the categories *learned*, *partner* and *idiosyncratic* we calculated the relative frequencies by counting their proportion relatively to the total number of critical object references produced by each participant. Then these relative frequencies were analyzed separately for each variable with an ANOVA with the two factors *knowledge state* and *time pressure*.

The factor *time pressure* never shows a statistically significant effect. This result is also substantiated by an analysis of the duration of the dialogs, where *time pressure* also had no significant influence. As a consequence we will restrict our report to the factor *knowledge state*.

Though we get statistically significant results concerning the variables frequency of learned and frequency of partner names, with these variables the problem exists that in the condition where both participants had learned identical names we cannot differentiate between learned and partner names on the lexical level. All names except idiosyncratic ones can be counted either as learned or as partner. Therefore these results cannot be interpreted until we find a way to differentiate between the two categories. For the variable frequency of idiosyncratic names the analysis showed no statistically significant effects.

Tab. 2 Number of references to critical objects. Under both conditions, identical learned names and different learned names, a similar amount of references can be observed. The percentages are relative to the total number of references to critical objects observed.

number of references	*direct references*			*indirect references*			total
	learned	*partner*	*idiosync.*	*learned*	*partner*	*idiosync.*	
identical names learned	74 (9.8%)	0 (0%)	4 (0.5%)	287 (37.9%)	0 (0%)	22 (2.9%)	**387 (51.1%)**
different names learned	67 (8.8%)	21 (2.8%)	5 (0.7%)	172 (22.7%)	93 (12.3%)	13 (1.7%)	**371 (48.9%)**
total	**141 (18.6%)**	**21 (2.8%)**	**9 (1.2%)**	**459 (60.6%)**	**93 (12.3%)**	**35 (4.6%)**	**758 (100%)**
			171 (22.6%)			**587 (77.4%)**	**758 (100%)**

Landmark use of critical objects

Based on the overall analysis, we were especially interested whether the idea with the critical objects forming the landmarks paid of in our setting. During the course of the study, the 16 pairs of participants had to place 144 critical objects (chances for direct references). They did so using 171 (+18.8%) direct references to critical objects, or 1.19 uses per relevant instruction (see Tab. 2 for a detailed overview). This was expected, as for introducing an object the participant with the piece of the map has to ask the partner to fetch the object out of her personal pool of objects. As the box is out of view for the one asking, the referring noun phrase employs the color and the shape of the object, thus using a direct reference. This we did not count as using a critical object as a landmark. Though, within each session for 26 of the 27 objects reference objects could have been used in the placement instructions, summing up to 416 potential uses of indirect references to critical objects. Effectively, the participants used 587 (+41.1%) references to critical objects. In other words: we found 1.41 uses of critical objects as landmarks per placement instruction. Thus, the strategy that the critical objects form landmarks to elicit a large number of verbal references to those objects is successful.

Direct vs. indirect use of names for object shapes

An analysis of the direct-indirect use with regard to the three categories of references shows that in direct references participants use the names they have learned independently of the fact if they have learned identical (m = .20, sd = .04) or different names (m = .19, sd = .08). If different names have been learned, there is only a small number of uses of partner names in direct references (m = .06, sd = .06). Thus the participants seem to stick to the names they were primed to use during the experimental instruction when introducing new objects.

This result is remarkable, as we expected participants to use more partner names when they had learned different names. This indeed also manifests in the overall data. For the variable frequency of learned names significantly more learned names are used overall, if participants had learned identical (m = .95, sd = .07) than if they had learned different names (m = .65, sd = .26; $F_{1,30} = 16.407$, $p < .001$).

Thus, participants use partner names mainly in indirect references. When they have learned identical names they use significantly more learned names (m = .74, sd = .08) than when they have learned different names (m = .46, sd = .22; $F_{1,30} = 21.357$, $p < .000$). And this effect can directly be explained by a higher amount of partner names used in the condition where different names have been learned when they refer to critical objects indirectly (m = .25, sd = .18) than directly (m = .06, sd = .06). There is no significant difference in the use of idiosyncratic names in both conditions.

With the result on the *direct* vs. *indirect* use of the critical names the idea of alignment is supported. If an object is named directly as target object the focus of attention is concentrated on this critical object that should be placed next. This situation is a direct recall of the instruction situation where the naming of the critical objects had been learned explicitly and therefore this leads to an explicit referring to the learned name. If the object is named indirectly as reference object the focus again is on the next target object that should be placed, whereas the reference object and its learned name are out of focus which may lead to an implicit and automatic take over of the partner name.

Hints on coping behavior

Finally, we were interested in the effect of the condition where we introduced differences in the naming of the critical objects. How do the participants cope with the induced problems? A thorough investigation of this question has to be deferred until all experimental sessions are transcribed and annotated. At this point we can only provide some hints.

First, we examined, if the induced differences in lexical preferences to name the critical objects lead to longer dialog sessions, e.g., because of an increase in negotiations. This is not the case: In the condition with identical names learned the mean session duration is 12.04 min and with different names learned the mean duration is 11.58 min (sd = 2.3).

Second, we analyzed if participants use references to the critical objects more often in the condition with different learned names for the critical objects. Our analysis shows that participants use significantly more direct critical names when they have learned different names (m = .26, sd = .08) than if they have learned identical names (m = .21, sd = .04; $F_{1,30} = 5.885$, $p < .05$). Despite of these differences we found barely any explicit negotiations in a first overview of the video recordings and, if at all, only in later stages of the dialogs.

The participants seem to compensate for the differences by increased redundancy in terms of repetitions when instructing the partner to fetch the object from her box. The opposite effect can be observed for indirect references. Here participants use significantly less indirect critical names when they have learned different names (m = .74, sd = .08) than if they have learned identical names (m = .79, sd = .04; $F_{1,30} = 5.885$, $p < .05$). When differences in the lexical terms used to refer to an object shape are an issue, participants are more careful when using those objects as reference objects.

The absolute numbers of references to critical objects for both conditions, identical and different names learned, are shown in Tab. 2. While there are differences in use depending on direct or indirect references, the total number of reference uses under each condition is comparable. This again confirms that the difficulties induced by different knowledge states do not lead to an increase in negotiations.

Conclusion

All in all, the Jigsaw Map Game has proved as a new capable setting for the investigation of natural interactive language processing in a controlled experimental way. It offers the possibility of real face-to-face interaction with two (or more) communicative partners. The roles of the partners can be varied from an equal to a strongly differing status regarding their knowledge, acting or goals. Furthermore, it allows for manipulations of the nature of the game, e.g., cooperative vs. competitive, and for a wide range of experimental variations concerning the communicative situation (direct face-to-face, mediated, remote, etc.) and the experimental materials. As the order of the turns within the game can be predefined, the games of different pairs of interlocutors can be easily compared. Finally, it offers many features that allow for a detailed linguistic analysis beyond lexical references, such as the analysis of sentence structure or the use of spatial reference systems. Furthermore, it offers the possibility to analyze prosody or gesture and to employ eye tracking techniques to get insights into relevant gaze information.

In the experiment reported we could get basic insights into the way how object references are coordinated in a face-to-face dialog. Particularly the results of direct and indirect object reference uses indicate that coordination of lexical references takes place fast and easily corresponding with the notion of alignment. However, to decide how local these effects of taking over the references of the partner are, a full annotation of the dialogs and more fine-grained data analyses including transition probabilities and distance measures are necessary.

Regarding the condition where both participants learned identical names for the critical objects we cannot tell lexically whether the naming has been learned or adopted from the partner. Here, we do not know so far whether participants use a name because of the strong priming during the experimental learning phase or because they aligned to the partner's use in preceding utterances. This differentiation is necessary not only for better analyzing the data but also on a theoretical level. If an object name is primed by the partner we can interpret it as an automatic take over in the sense of alignment. If a name is used because it is learned, this indicates a more explicit use to introduce a new object in the dialog. One idea to overcome this, is investigating whether there are differences on the phonetic level between interlocutors. If there are, and if alignment happens on the phonetic level, phonetic analyses could produce measures to see, if certain names are pronounced closer the way the interlocutor has pronounced the name before. Thus, it may be possible to conclude whether the name used by the partner has been adopted or not.

Concerning the influence of cognitive load we can draw no conclusions so far, since using the chess clock did not really cause time pressure. This is supported by the analysis of the duration of the dialogs as well as by informal remarks of some participants after their experimental session. Therefore, further experimental work should establish better methods to evoke time pressure and test other ways of inducing cognitive load, e.g., by using dual-task paradigms. This will also contribute to a deeper investigation of the relation between the concept of alignment and the necessity of establishing full common ground in communicative interactions.

References

Anderson, A. H. et al. (1997). The impact of VMC on collaborative problem solving: An analysis of task performance, communicative process, and user satisfaction. In K. E. Finn, A. J. Sellen & S. B. Wilbur (eds.), *Video-mediated communication* (S. 133-155). Mahwah, NJ: Erlbaum.

Anderson, A. H. & Boyle, E. (1994). Forms of introductions in dialogues: their discourse contexts and communicative consequences. *Language and Cognitive Processes*, *9*, 101-122.

Anderson, A., Brown, G., Shillock, R. & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.

Boyle, E., Anderson, A. H. & Newlands, A. (1994). The effects of visibility on dialogue performance in a co-operative problem solving task. *Language and Speech*, *37*, 1-20.

Branigan, H. P., Pickering, M. J. & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*, B13-25.

Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482-1493.

Buhl, H. M. (2001). Partner orientation and speaker’s knowledge as conflicting parameters in language production. *Journal of Psycholinguistic Research*, *30*, 549-567.

Clark, H. H. (1996). *Using language.* Cambridge: University Press.

Clark, H. H. & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: American Psychological Association.

Clark, H. H. & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62-81.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1-39.

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S, Garrod, S. & Bruce, V. (1997). Face-to-face and video mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, *3*, 105-125.

Garrod, S. (1999). The challenge of dialogue for theories of language processing. In S. Garrod & M. J. Pickering (eds.), *Language processing* (pp. 389-415). Hove: Psychology Press.

Garrod, S. & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*, 181-218.

Garrod, S. & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*, 8-11.

Horton, W. S. & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91-117.

Kranstedt, A., Lücking, A., Pfeiffer, T., Rieser, H. & Wachsmuth, I. (2006). Deixis: How to determine demonstrated objects using a pointing cone. In S. Gibet, N. Courty & J.-F. Kamp (eds.). *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers* (pp. 300-311). Berlin: Springer.

Krauss, R. M. & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*, 343-346.

Lombard, M. & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2). http://jcmc.indiana.edu/vol3/issue2/lombard.html.

Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169-226.

Rummer, R., Wintermantel, M., Schweppe, J. & Dirksen, R. (2007). *Wie lässt sich Common-Gound-Etablierung beim monologischen Instruieren verbessern?* Manuscript submitted for publication.

Schober, M. F. & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher & S. R. Goldmann (eds.), *Handbook of discourse processes* (pp. 123-164). Mahwah, NJ: Erlbaum.

Schaffranietz, G. (2007). *Zwischen Alignment und Common Ground: Ein Experiment zur Koordination von Objektbegriffen im Dialog*. Magisterarbeit. Universität Bielefeld: Fakultät für Linguistik und Literaturwissenschaft.

Weiß, P., Pfeiffer, T., Eikmeyer, H.-J. & Rickheit, G. (2006). Processing instructions. In G. Rickheit & I. Wachsmuth (eds.), *Situated communication* (pp. 31-76). Berlin: Mouton de Gruyter.