Environmental Constraints on Cultural Conventions:

A Case Study of Linguistic Coordination in Joint Problem Solving

Jonas Nölle

M.A. Thesis in Cognitive Semiotics

Supervisor: Dr. Kristian Tylén

Aarhus University, 2015

Stud. No. 201302592

Acknowledgements

First of all, I would like to thank my supervisor **Kristian Tylén** for his guidance, help with technical and organizational issues and his general openness that allowed me to pursue my own research interests. In regard to the experiment, I would like to thank the **Cognition and Behavior Lab (Dan & Josh)** for facilities and troubleshooting (e.g., when the wrong Java version kept reinstalling itself...), and further the **Interactive Minds Centre** for the Seed funding that allowed to pay my participants. I am also grateful that the **Studienstiftung (German National Academic Foundation)** continued to support me during my M.A. in Denmark.

I would further like to thank **Gregory J. Mills**, who developed the original DiET program that was used for this study, helped tremendously with the implementation of my own experiment into the code and gave important tips and insights. Furthermore, **Rafael Moreno** for support and assistance at the lab during data collection and **Riccardo Fusaroli**, not just for his help with the specific statistics used here, but letting me take his (and **Josh's**) incredible R course that enabled me to work on this level in the first place. **Stefan Hartmann** was also a reliable help, whenever I ran into programming problems.

Lastly, I wish to thank **Marlene, Juanito** and **Lucía** for motivational support and many insightful discussions as well as numerous other cognitive semioticians and friends who volunteered for the pilot study. Overall, the M.A. at the **Center for Semiotics** was a far more enriching experience than I could have ever imagined beforehand. This is thanks to **Svend, Peer** and **Mikkel** as well as **Frederik** and **Rasmus** whose academic competences, candor and most importantly, humor, make it the place to be. I am lucky and grateful for having met such great people during my stay in Denmark.

Tusind tak!

Summary

Background

What drives the evolution of cultural conventions? This broad question is the basis for the present study. More specifically, it shall be addressed how a specific kind of cultural conventions come about when we interact with each other: Linguistic routines. Language is one of the most fundamental cultural devices and used for all kinds of social purposes (Tylén *et al.*, 2010). It is most frequently used in direct face-to-face interaction, where we engage with one another to jointly coordinate our actions towards a joint goal (Clark, 1996). But how are we able to coordinate the myriad of processes required for successful conversation? As interlocutors, we have to be sure that we talk about the same things, relate what we are saying to the situation, to what we both know (or do not know), integrate gestures and produce our utterances without interrupting each other. This sounds like a whole lot of effort and yet, dialogue usually unfolds smoothly and effortlessly. Psychologists have attributed this to automatic processes: Pickering and Garrod (2004) suggest in their interactive alignment model that interlocutors mechanistically synchronize their language, which leads to mental alignment as well. When a construction is comprehended, it will also get activated for production – fully automatically and without the need to model the mental states of one's partner. This way, linguistic routines emerge spontaneously while we talk to each other. These routines can also last longer than the duration of the dialogue and even spread in a community (cf. Garrod & Doherty, 1994). They become conventions.

Looking at these conversational dynamics is interesting, because languages as such are nothing but sets of constantly re-negotiated conventions. Small local changes made by speakers can lead to global structures arising in a self-organized manner without any central control. Languages thus “evolve” as cultural complex adaptive systems through and in response to their usage (cf. Beckner *et al.*, 2009). The diversity we see among the world's languages is an example of different populations finding conventionalized solutions for the same collective problems, realized in different grammatical and semantic categories. But are these different conventions just arbitrarily chosen from several options? Could there be factors driving cultural evolution into specific directions?

Based on Gibson's (1979) theory of affordances, I suggest that environmental factors can, in fact, influence the salience of solutions when several alternatives are possible. Since communication never happens in a void and different cultures are situated in different environments, this can lead to selection. Specific conventions become more likely to emerge than others, simply because they are more environmentally adaptive, even though the communicative problems they are intended to address may be the same across cultures.

Aim

This thesis is a first attempt at experimentally testing a causal relationship between environmental factors and emerging linguistic conventions. This is meant to inform research both in dialogical communication and cultural language evolution. Recently, there has been increasing evidence that the diversity of the world's languages might be partly influenced by ecological factors (e.g., Majid *et al.*, 2004; Everett, Blasi & Roberts, 2015). However, most of these studies rely on correlational data. An experimental approach can complement this, since it enables direct manipulation of specific variables and the observation of direct effects. As a first test case, the present study is only concerned with linguistic routines in an existing language (Danish), but the insights gained can be used in follow-up research to test ecological constraints on linguistic structure more generally.

Methods and Procedure

An experimental approach was used to address the research question. The paradigm built on the "maze game", a joint problem-solving task that had been developed in experimental psychology to investigate linguistic coordination in task-related dialogue (Garrod & Anderson, 1987). Pairs of participants sat at different computers and had to jointly solve a series of mazes. The game was cooperative, since in order to reach the finish, players had to communicate positions to their partner that only they themselves could see. In previous studies, it was found that participants, while playing the game, started to establish and routinize linguistic conventions that relied on different strategies for describing locations. For instance, some pairs used a *PATH* scheme, where they would describe the path their partner should follow to get to the desired location. Others guided their partners by referring to specific *FIGURAL* configurations they recognized in the maze layout. Lastly, abstract strategies were also used,

where participants would dissect the maze, either into a MATRIX-like coordinate system along two axes or into parallel LINES to specify locations. These four description schemes (PATH, MATRIX, FIGURAL and LINE) varied among the pairs, suggesting that local interaction processes may govern language processing in dialogue. Furthermore, in a follow-up study, it was shown how these temporary routines could spread in a community of players and lead to group-wide conventions, e.g., the majority of pairs adhering to the MATRIX description scheme (Garrod & Doherty, 1994).

Adapting the paradigm thus enabled to test for potential pressures that could influence these routines. Three “environmental” conditions were added that were designed to afford for specific strategies more than for others: The “regular” condition featured neutral and square-like mazes. It was expected that pairs in this condition more likely rely on the MATRIX or PATH scheme that had been found to be most popular in previous studies. The “irregular” condition featured mazes that had very particular and salient figural shapes and configurations. It was predicted that pairs in this condition would more likely on the FIGURAL scheme. Finally, the “stratified” condition featured mazes with very salient horizontal displacements. It was therefore predicted that the LINE scheme would be more popular among participants solving mazes in this condition.

Another difference to the original study was that participants used a chat interface to communicate, which allowed testing several pairs at the same time and eliminated the need to transcribe dialogue data.

Results

Chat protocols from 33 pairs were coded in terms of description schemes and analyzed. It was found that participants established the same description schemes as in the previous studies and showed typical patterns of alignment and coordination. In terms of environmental influences, it was found that the description schemes were distributed quite differently among the conditions: In the regular condition, the picture was very mixed, with examples of pairs using PATH, LINE or MATRIX predominantly. In the “stratified” condition, LINE was by far the strategy used most frequently as dominant scheme and in the irregular condition, FIGURAL was the most frequent dominant strategy. The attraction strength of each description type per condition was calculated to test whether there were statistically significant differences between the conditions. It turned out that MATRIX was not a significantly stronger attractor in the regular

condition. This was due to the fact that *MATRIX* was very stable whenever it occurred. However, *PATH* was more attractive in the regular condition, as predicted. In the stratified condition, *LINE* was the strongest attractor and more attractive than in any other condition. Similarly, *FIGURAL* was significantly more attractive in the irregular condition than in the other conditions. This was the most remarkable difference, as *FIGURAL* was otherwise very unattractive and participants usually switched to other strategies after using it. The results suggest that there were two main pressures influencing the choice of description scheme: 1) The environmental context making strategies more attractive that were environmentally adaptive (like *FIGURAL* in the irregular condition or *LINE* in the stratified condition). 2) A general tendency to use an abstract *MATRIX* scheme – which could be the reflection of a cultural bias.

Conclusion

The experimental results indicate that manipulating the environment was enough to systematically influence the linguistic conventions participants developed while solving the mazes. Linguistic routines seem to be shaped not only by interactional dynamics and implicit priming mechanisms, but also by the external context. I suggest that this is the case because the affordance for the same solutions may differ in salience across environments. Some environments make a certain strategy very intuitive, while in others it may be less salient. There is hence a higher chance for a salient strategy to be selected and expressed in a linguistic convention. Based on this finding, I suggest expanding Pickering & Garrod's (2004) Interactive Alignment Model, by adding the environmental context. A model of language-processing in dialogue has to take into account that conversations are always situated, even when the surrounding environment is not directly related to the task at hand.

Finally, the insight that linguistic conventions may be shaped by environmental factors has some implications for the study of language more generally. If language is the outcome of cultural evolutionary processes, as many contemporary accounts consider (see, e.g. Steels, 2011 for a review), then the pressures found in the current study should apply to linguistic structure as well. Grammatical choices that build on different conceptual categorizations of the world can be seen as conventional strategies, just as the routines in the maze game. The diversity among the world's languages could therefore be more than a result of arbitrary choices. It is up for future research to uncover why some languages make specific conceptual choices that others do not.

Contents

Acknowledgements	2
Summary.....	3
Chapter 1: Introduction	9
1.1 Motivation and research question.....	9
Chapter 2: Background	14
2.1 Linguistic coordination and cultural conventions.....	14
2.2 Interactive alignment.....	21
2.2.1 Alignment: an automatic process	21
2.2.2 The mechanisms underlying interactive alignment.....	24
2.2.3 Routinization	27
2.3 Limits of interactive alignment and additional concepts.....	29
2.4 From routines to conventions: the concept of cultural evolution.....	33
2.4.1 Constraints on cultural evolution.....	37
2.5 The theory of affordances.....	40
2.6 Summary	42
Chapter 3: Case study	44
3.1 How environment affects linguistic coordination in the maze game	44
3.1.1 The maze paradigm.....	44
3.1.2 Findings from the original studies.....	46
3.2 Experiment: A variation of the maze game	50
3.2.1 Materials	51
3.2.2 Participants	57
3.2.3 Procedure.....	58
3.3 Results	59
3.3.1 Coding of chat protocols	59
3.3.2 Examples of different description schemes	60
3.3.3 Examples of alignment and synergy in the chat data	61
3.3.4 Impact of environment	66
3.4 Discussion.....	75
Chapter 4: General discussion	79
4.1 Taking the interactive alignment model out into the wild	79
4.2 Implications for the study of language evolution	82

Chapter 5: Conclusion.....	85
References	86
Appendix I – Mazes.....	96
Appendix II – Data.....	100
Appendix III : Instructions (Danish and English)	101
Appendix IV : Coding Scheme	103

Chapter 1: Introduction

1.1 Motivation and research question

“Language is a labyrinth of paths. You approach from one side and know your way about; you approach the same place from another side and no longer know your way about.”

– Ludwig Wittgenstein, *Philosophical Investigations*

Humans are characterized by culture – a vast amount of behavior and customs that are learned from other individuals and passed onto future generations in the form of tradition. Notably, there is high variation among different societies in terms of what is culturally “conventional”, i.e. how members of these groups go about certain things. This includes basic ways of living like dwelling (e.g., urban or rural), subsistence mode (e.g. hunter-gatherer, agriculture or industrial), but also more trivial practices, like hairstyle or clothing. It has been argued, e.g., by Lewis (1969) that conventions like these are in fact devices we use to act together effectively. In a social community, there are a lot of things we have to coordinate in order to function as a society. A simple example would be the problem of which side of the road to drive on to prevent crashes. To overcome this problem, different societies came up with regular behavior, i.e. conventions that they expect their members to conform to (e.g. right-hand traffic).

However, for different communities, these problems and their respective solutions might vary slightly or in some cases more strongly. Cultural groups are always situated in a specific environment that leads to specific coordinative demands. For instance, living on an island with sparse vegetation vs. living in a jungle may pose different challenges for obtaining food and thus requires different adaptations. Such environments can be referred to as “ecology”, “niche” or “Umwelt” (von Uexküll, 1921), in that they are more than just a space for the cultural group to exist in. Instead, both, a specific culture and the environment it is situated in, are part of a dynamical system that organizes itself through negative and positive feedback loops between its parts.

To illustrate this idea with an example, think of a community of early humans populating a forest on the verge of the Ice Age. As the climate becomes colder, it becomes necessary to build housing, for which they begin to cut trees and use the wood

as building material. Their Umwelt thus provides them with a problem, but also a specific set of solutions that could vary between environments (in the mountains, caves might have been more likely). It *affords* the creation of wooden housing.¹ If this strategy is successful, many houses will be built and the population will thrive (under the assumption that other factors, such as food-scarcity, do not inhibit growth). This results in a positive feedback loop where the community that came up with BUILDING WOODEN HOUSING grows, propagates this convention on the cultural level and – over time – possibly enhances it by improving structures, adding decorative elements or coming up with new techniques to cut more and more trees at the same time. It is not only the environment driving this change in the culture, but also the culture, which in turn changes the environment that, as the growing population is cutting more and more wood, turns from a forest into a flatland. This nonlinear phase transition leads to a breakdown of the loop. The humans cannot cut any more trees and have to either come up with novel cultural solutions, such as transporting wood from areas further away, or switching to other materials, like clay. In an extreme thought experiment, the land could have turned into an infertile steppe, where a nomad lifestyle might be more adaptive, leading the group to switch from BUILDING WOODEN HOUSING to BUILDING TENTS and TRAVELLING CONSTANTLY. This artificial example is of course highly simplified and ignores a number of factors, especially within the cultural group or the environment itself that could lead to a variety of outcomes. It is merely meant to illustrate the point that – similar to organism-habitat ecosystems that we can observe in the animal kingdom – cultural communities and their environment form an interrelated system where patterns of cultural behavior are driven by ecological challenges and in turn these cultural practices affect the environment (Figure 1).²

Fig. 1. Feedback loop between members of a cultural group and their local environment.

¹ The notion of affordances will be explored more deeply in chapter 2.4.

² This is not to say that culture in itself does not already resemble a complex ecosystem that evolves on different time and size scales, which its interdependent components act on (Wimsatt, 2013: 564).

Today, we are surrounded by technologies created by human cultural groups that have evolved over time, e.g., from simple canoes to large freighters as a solution of travelling and transporting cargo over sea (cf. Arthur, 2009: 16). It is clear that our cultural lifestyle has affected our environment, most dramatically even on a global scale: Our planet's climate is changing due to human technology (see, e.g., Hughes *et al.*, 2003; Solomon *et al.*, 2009). At the same time, this change is already feeding back into culture, in that, on the one hand, many industrialized societies have, e.g., decided to cut CO₂-emissions and increasingly promoted cultural practices of “sustainable living” (in fact, collections of various practices), while on the other hand, cultures were forced to adapt to these environmental changes: For instance, the subsistence activity that has, for millennia, been predominant among the First Nations of central British Columbia is dependent on weather patterns, and is thus highly affected by the effect regional climate change has on certain fish species (Jacob, Daniels & Hinch, 2010).

If such interdependencies exist, it should not be surprising that cultures situated in different climate zones came up with very different solutions to the same problems (housing, foraging, clothing etc.). But can such an environmental influence be extended to all kinds of cultural practices?

The main goal of this thesis is to address the question whether this is also true for a more immaterial cultural practice: **Linguistic communication**.

More specifically, I want to explore whether linguistic conventions that emerge spontaneously in dialogue are similarly dependent on environmental factors. The idea is that linguistic routines, i.e. recurring lexical structures we use to refer to the same things in interaction, are solutions to coordination problems just like other conventions such as the ones mentioned above. I will argue that language, above all, is used in social settings to coordinate with other people and achieve collective goals (like going for a coffee, solving a homework exercise together, deciding on a holiday destination, etc.). It has been shown in experimental research that people spontaneously develop novel linguistic strategies to achieve these goals. I will review these findings in chapter 2, where the focus lies on language as joint action and the different dynamics underlying its use in dialogue. I will discuss the interactive alignment model of dialogue, proposed by Pickering and Garrod (2004), which suggests that automatic priming mechanisms between interlocutors are the basis for successful com-

munication. Along with the authors, I maintain that people automatically imitate each other on several linguistic levels when they engage in conversation and that this leads to the synchronization of their mental representation of the task at hand. I will discuss how this leads to spontaneous routinization of linguistic practices and eventually community-wide conventions that spread and evolve via processes of cultural evolution. In addition, some alternative accounts are presented that highlight either egocentric processes in the speaker or structural complementariness on the level of the interaction, respectively. It will become clear that conversation is a complex endeavor that requires coordination on a variety of levels, and ad-hoc linguistic conventions designed to fit the task are an efficient solution to this problem. After discussing how these emerge in dialogue, I will turn to the main research question:

In which way can environmental constraints function as selective pressures on cultural conventions that emerge in interactions between individuals? Finally, I will show that the diversity of the world's languages could be the result of many such local interactions where (environmentally motivated) conventions arose in different places and times. The grammar and lexicon of a language provides a set of collective conventions that can be used efficiently to overcome communicative problems. They are also subjected to constant change as new ways of addressing the same issue are discovered, or as new problems arise that require a novel solution. Based on Gibson's (1979) theory of affordances, I argue that while the same coordinative problems might arise in different populations, the environment in which they are situated profile different solutions. In other words, some strategies are more favorable in one setting than another, because they appear more intuitive or are more effective there. In language, this can motivate the choice for one specific linguistic convention over a set of other alternatives.

In chapter 3, I will focus on an experiment that serves as a test case for this hypothesis. In a collaborative problem-solving task adapted from Garrod and Anderson (1987), participants were seated at different computers and had to communicate via a chat interface to guide each other through virtual mazes. Previous versions of the experiment illustrated how different participant pairs develop distinct linguistic routines, as they engage in this task. These routines are linguistic descriptions of locations that follow different underlying conceptualizations of the maze at hand. The present version of the experiment was designed to test whether the emergence of

these routines could be motivated by environmental factors. Three conditions were added to the paradigm, in which the mazes varied in their environmental layout to promote different strategies. They were designed in a way that made different conceptualizations of the maze more prominent. Whereas the “stratified” and the “irregular” condition featured mazes with salient horizontal displacements and salient elements resembling figural objects, respectively, the “regular” condition featured more neutral, square-like mazes. The task was the same in each condition. The results of the experiment indicate that linguistic strategies that were especially adaptive in the different conditions were much stronger attractors in these conditions and more stable over time. This suggests that although the task was identical, it was enough to place the participants in slightly different environments in order to influence which particular linguistic strategy they would converge on.

In chapter 4, I will discuss these results in more detail and consider the implications this has for the studies of dialogue and cultural language evolution. I will argue that external environmental factors need to be included in theories of natural language use, such as the interactive alignment model because, as my experiment shows, the environment can initially influence which strategies interlocutors eventually align to. I will also refer to the implications this has for the study of language evolution. If languages evolve culturally, we would expect the same kind of pressures to be at play during the emergence of grammatical categories. For instance, spatial language has been shown to rely on different conceptual strategies in different languages, some of which relate to very salient properties of the environment. This could very well be the outcome of similar processes of ecological adaptation and selection that were at work in the present study. Finally, I will propose some ideas for novel experimental paradigms that can be used to further investigate whether this is indeed the case.

Different lines of research have discussed innate structures, implicit cognitive learner biases or interactional dynamics as key factors shaping linguistic structure. This thesis is in essence an attempt to present external factors as an additional force that can motivate linguistic conventions by affording different cognitive strategies which in turn become expressed differently. The findings are therefore relevant for research on conversation, joint problem solving and language.

Chapter 2: Background

2.1 Linguistic coordination and cultural conventions

This chapter is meant to provide an overview about the nature of linguistic coordination as it is constantly taking place when we communicate in direct face-to-face interactions. Further, it will be shown in which way these local dynamics can give rise to more global, cultural patterns that emerge as an outcome of many such interactions. Following previous work, I will first argue that dialogue is the most fundamental type of language use. I will subsequently detail some of the mechanisms that determine the structure of dialogue, such as interactive alignment and interpersonal synergies, and finally give an account of cultural evolution being directly grounded in these processes, which at the same time function as the basis for the emergence of linguistic conventions.

2.1.1 The primacy of dialogue as joint action

As the most common and prototypical form of linguistic interaction in everyday life, dialogue lies at the heart of human communication (Pickering & Garrod, 2004). We learn language through this type of joint activity and keep using it in social interactions to maintain relations, establish or enforce cultural practices as well as acquire and transmit knowledge from and to others (Clark, 1996; Fusaroli & Tylén, 2012).

Although dialogue is a process we frequently engage in, it is far from trivial: Having to monitor her own utterances both with respect to the conversation as such and her conversational partner, the speaker simultaneously has to process her partner's contributions and the actual content of the dialogue. For instance, if we were to be engaged in a conversation, I would have to monitor whether what I am saying is actually relevant to you, whether you understand the words I am using, whether you know what or who I am even talking about – in short: I have to become a cooperative conversational partner. Paul Grice explored these aspects in his theory of conversational implicatures, whereby the rational speaker relates her speech acts to general features of discourse. The most important feature, according to Grice, is that conversations are not just successive sequences of disconnected remarks but require a cooperative effort and the recognition of a common purpose from both participants (Grice, 1991: 26). Furthermore, the Gricean “Cooperative Principle” and “Conversa-

tional Maxims”³ guide the speaker to make her utterance relevant to the current point of the conversation, just like they provide a basis for her conversational partner to reconstruct any conventional or non-conventional implicature contained in it (Grice, 1991: 25 ff.). They thus provide something like a set of rational guidelines that both interlocutors adhere to, in order to communicate successfully and efficiently.

Besides adhering to the Cooperative Principle and maxims, other factors play into people’s behavior in dialogue as well: Social relationships (symmetric vs. asymmetric; the personal history between interlocutors, etc.), cultural norms (terms and wordings to (not) use in specific situations) and the immediate or displaced context the dialogue is situated in (e.g., a conversation between a surgeon and her assistant in the emergency room is largely constrained by the task they are performing) constitute dialogue on the socio-cultural layer. In addition, lexical and syntactic choices have to be made at any moment in time, and paralinguistic cues (postural sway, eye-gaze, intonation, co-speech gestures) have to be coordinated and interpreted accordingly (cf. Goodwin, 2000).

This poses a set of numerous behavioral and cognitive coordination problems for interlocutors engaged in such activity. It is puzzling how, despite that, dialogue usually works flawlessly and easily without the need to reflect upon the mechanisms underlying its success. If one would attempt awareness of all the processes on all levels during conversation, the natural flow would probably be disrupted. In analogy to an introspective centipede, unable to move as it consciously attempts to coordinate all its legs, this has been referred to as a “centipede’s dilemma of social interaction” (Dale *et al.*, 2013: 44).

While Grice and other scholars like Austin (1962), Searle (1969) and originally Wittgenstein (1953) have correctly pointed out that linguistic meaning is strongly tied to the actual usage situation, their accounts are still centered around the individual mind in terms of intentional processing and production: The interlocutor is simp-

³ Conversational maxims are not moral norms, but rather rational guidelines that one expects every conversational partner to follow. Grice (1991: 26 f.) segmented these guidelines into five main maxims, here rephrased as imperatives: 1) Quantity: Include as much information, as is required, but not more. 2) Quality: Make your contributions genuine and not spurious. 3) Relation: Make your contribution appropriate to the immediate needs at each stage of the transaction. 4) Manner: Make it clear what contribution you are making and execute its performance with reasonable dispatch.

ly expected to recognize communicative intentions.⁴ Similarly, another line of research in psycholinguistics has also focused on individual cognitive processes. Brown-Schmidt and Tanenhaus (2008: 643 f.) refer to this tradition as the “language-as-product” tradition, because it “typically examines moment-by-moment processes in real-time language processing, using carefully controlled stimuli, scripted materials, and fine-grained online measures that are closely time-locked to the input” (cf. also Clark, 1996; Pickering & Garrod, 2004: 170). This includes how listeners constantly integrate input from the unfolding speech stream in their comprehension, e.g., to resolve temporary ambiguity (cf. Marslen-Wilson, 1973; Eberhard *et al.*, 1995).

While both the aforementioned pragmatist and these psycholinguistic accounts were fruitful in revealing some of the possible mechanisms underlying language in use (compared to linguistic accounts that merely focus on structure⁵), their reliance on rational maxims and intentionality appear insufficient to explain why dialogue seemingly works automatically and without any rational effort needed, considering additional issues like the coordination of who is to talk at which time, what exactly they are saying and how it is understood with respect to the context. Unlike written language, there is only limited time for planning and revision, because utterances are produced and have to be understood on the fly, constantly demanding to task-switch between comprehension and production (Garrod & Pickering, 2009: 293).⁶ Understanding conversation as a mere succession of illocutionary acts (assertions, questions, requests etc., cf. Austin, 1975) that are fully produced and understood is an idealization (Clark & Wilkes-Gibbs, 1986: 2). In reality, conversations are rather messy and include a big number of incomplete, elliptical utterances, misunderstood parts or corrections (e.g., Galantucci & Roberts, 2014).

Clark and Wilkes-Gibbs decided to move away from this “literary model of definite reference” that makes these idealizations. Instead, they propose a “collaborative model for the process of reference” (Clark & Wilkes-Gibbs, 1986: 3) that takes into account the limits of spoken conversation and techniques interlocutors rely on to

⁴ See also Sperber and Wilson’s (1986) “relevance theory” that focuses on the ability of interlocutors to make the right inferences at any moment of the conversation. They argue that our minds are adapted to rapidly and automatically account for beliefs and intentions of others as they come in dialogue. They consider these inferences as the basis for communication.

⁵ For instance structuralism (de Saussure, 1983) or generative grammar (Chomsky, 1980) put focus on syntagmatic and paradigmatic relationships between linguistic units or hierarchical dependencies, respectively.

⁶ The usual pauses between interlocutors are even below the standard 300 ms reaction time (cf. Dale *et al.*, 2013: 72).

overcome them. For example, speakers draw addressees into the reference process by using placeholder referents or dummy nouns (e.g., *ermm, what's-his-name?*) that the addressee is expected to fill in. Likewise, the addressee can repair reference by using different noun phrases (NPs) from the speaker. A mutual belief that the listener has understood what the speaker meant is needed to establish reference and this requires a collaborative effort. Both conversational partners may iteratively produce new versions of a NP until they decided on a version that both accept. "Accept" here means according to a "criterion sufficient for current purposes" that can range from high to low depending on the situation (Clark & Wilkes-Gibbs, 1986: 34). For instance, if I told you where to find the university, you would set your criterion high in a situation where you have asked me for the exact directions, whereas you would set it low if I would just casually mention its location while talking about studying in general.

Once utterances are mutually accepted (e.g., after repair, expansion or replacement) they become real "contributions" to the conversation (Clark & Wilkes-Gibbs, 1986: 35). Every contribution can be followed by a repair by the partner (cf. Schegloff, Jefferson & Sacks, 1977). Note that listeners often also accept utterances as contributions they have not fully understood. To some extent, conversation remains an uncertain endeavor, where it would be far too much of an effort for the listener to try to understand everything that has been said. Instead, it is only necessary to respond *as if* she had understood everything (even in insincere cases) as long as she feels that her comprehension is substantial or expects that the informational gaps will be closed later by the speaker. This is beneficial, because an idealized conversation, where everything is understood, would be highly impractical and inefficient, as after each clause the listener would interrupt the speaker to ask for clarification concerning the identity of referents and so on. Interlocutors have to recognize the responsibility they have in conversations to maintain the flow of information with just the right amount of uncertainty to still allow for successful communication.

Following these insights, it appears sensible to refer to linguistic activity as a form of "joint action [...] that is carried out by an ensemble of people acting in coordination with each other" (Clark, 1996: 3). Clark identifies face-to-face conversation as the most basic and common setting for language, since unlike written settings, it is universal to human societies and the only setting that does not require any additional skill (unlike, e.g., writing an article, acting in a play, lecturing etc.). This is because

other settings would lack the immediacy, medium or control of face-to-face conversation. “Immediacy” refers to the fact that participants can constantly see and hear each other and their surroundings in real-time without perceptible delay. The “medium” includes multimodal channels, such as speech, gestures and eye gaze, which are all evanescent (they fade and are not recorded) and make useful strategies available to interlocutors, such as interrupting, overlapping speech or responding *uh huh* and other markers of acknowledgement.⁷ Lastly, “control” means that both participants speak for themselves and jointly determine who says what and when, formulating their utterances on the fly as the conversation proceeds. Clark furthermore points out that children acquire language in this setting and are not able to use other media to learn their first language (cf. Clark, 1996: 7 ff.).

2.1.2 Common Ground and the creation of meaning

In order to establish meaning, the speaker has to be understood by the addressee, who in turn has to attend to him. This requires coordination, a form of joint action. They achieve this with the help of their “common ground”, the rich mass of cultural and episodic information they both share (and are aware of sharing). This includes background knowledge about the world both live in, the words they both understand as referring to the same concepts and shared experiences from the past (Clark, 1996, chapter 2). Furthermore, the immediate surroundings become part of the common ground that interlocutors rely on. Clark calls this “co-presence”, but I will adopt Tomasello’s notion of “joint attention” here (Tomasello, 2008: 78). Joint attention means that in the interactive context both partners are attending the same entities (e.g., objects or events they are talking about) and at the same time mutually aware that they are jointly attending these entities. According to Tomasello, this can be due to top-down processes (e.g., when both partners pursue a shared goal in the interaction) or bottom-up, e.g., when an event occurs and both partners know together that they have experienced it (e.g., both hearing a noise). Suppose that you are heading towards the train station with a friend. You reach the station, but the platform is still around the corner. Yet you can already hear the train arriving on the railway. You look towards your friend and say: “We have to hurry!”, to which your friend replies with a

⁷ Interestingly, repair markers that occur in conversations, like *huh* have very recently been identified as a pragmatic universal among the world’s languages, which strongly supports Clark’s arguments for face-to-face conversation as the most common and basic form of language use (Dingemanse *et al.*, 2015).

nod. This is only possible because you both know that you jointly attended the sound of the arriving train and your friend is immediately able to interpret your signal as referring to the approaching train as well as the action that he would like the two of you to perform in response. This example illustrates how common ground complements overt communicative acts: Everything that is part of the common ground does not have to be expressed/referred to explicitly, which allows us to even talk about absent referents (cf. Tomasello, 2008: 80). Every social activity that we engage in builds on a conceptual common ground (how is what we are doing related to what we are commonly doing in our cultural community?) and the current joint attentional scene (how are our surroundings related to what we are doing?). However, additional coordination is necessary in order to create meaning in dialogue. The speaker S, engaged in a joint activity with the addressee A, wants A to integrate new information relevant to the activity into her common ground in order to reach their shared goal. He can do so with a deliberate signal that includes his communicative intention (*We have to hurry!* in the above example). Signals can be words, gestures, noises, or more generally, “the presentation of a sign by one person to mean something for another” (Clark, 1996: 160, see also 155 ff.). According to Clark there are three types of signaling based on Peirce’s typology of signs:

1. **Demonstrating** a thing (which creates an icon), e.g. with a manual, facial or vocal gesture/intonation.
2. **Indicating** a thing (which creates an index), e.g. via pointing.
3. **Describing** a thing **as** something (which creates a symbol), e.g. via a linguistic expression.

In traditional pragmatic accounts (e.g., by Grice or Searle, as discussed above), signaling is viewed to be the sole use of symbols representing what the speakers mean and intend the addressee to interpret. It depends merely on choice and interpretation of symbols. According to Clark (1996, chapter 6) it is instead a mixture of these three signaling methods, where describing-as indeed involves symbols, but never works without indicating or demonstrating that rely on linguistic and non-linguistic resources. Language, in addition to being a single tongue (like, e.g., Danish or English), is here viewed more broadly, including every way and modality in which one person creates meaning for another (cf. Clark, 1996: 187 f.).

Clark (1996: 185) provides an example with George greeting Helen using the conventional meaning of *hello*. In this way, his action is **described as** greeting by the symbolic meaning of the word. His voice and eye gaze **indicate** himself as speaker, Helen as hearer and *now* as the time of greeting. His smile, intonation and open eyes **demonstrate** his enthusiasm. The single composite signal that arises from these three components can be interpreted by Helen as George, himself, now greeting her, Helen, enthusiastically. In this view, language (a symbolic system in the traditional sense) is embedded as single joint acts within a broader joint activity, like buying something at the supermarket, preparing a presentation together, planning a journey and so on. Signals coordinate actions at stages of these activities, enabling the participants to add information to their common ground that will bring them closer to their joint goal. Clark proposes an action ladder of four levels of joint action that comprise such an activity (see Clark, 1996: 148 ff.):

1. Execution and attention: The signaler S displays behavior and the addressee A perceives/is attending to it.
2. Presentation and identification: S presents signals (e.g., linguistic expressions or bodily gestures) to A and A identifies these signals.
3. Signaling and recognition: S signals that *p* (e.g., that S requests something) for A, and A recognizes that *p* from S.
4. Proposal and consideration: S is proposing a joint project (e.g, greeting reciprocation) to A and A is considering this proposal.

Level 4 is Clark's addition to classical pragmatic theories that focus on level 3, because here signaling receives its purpose "to further the joint activities people are engaged in at the time and to advance the common ground of the social group" (Carston, 1999: 170).

As we will come to see, linguistic conventions on the lexical level are forged exactly in these kind of extended usage situations where meaning is negotiated to pursue a joint goal. One way to coordinate action easily in such settings is through conventional procedures – routinized behavior (cf. Clark, 1996: 35 f.). But how are these established to begin with? In the following section, I will focus on the dynamic mechanisms that allow interlocutors in dialogue to form a coupled system with dis-

tinct properties on a higher level and how self-organization on this level leads to the emergence of conventions.

2.2 Interactive alignment

Pickering and Garrod (2004: 170) point out that while Clark and others engaged in the study of language as action have successfully managed to investigate processing in dialogue in natural tasks (embedded in joint activities), their sole focus remains on the goals and intentions of the participants, e.g., how common ground is built. This intentional view, which is based on a complex series of inferences (and inferences about inferences) as well as the constant modelling and updating of inferred belief-states, seems insufficient to adequately explain the simplicity that characterizes natural dialogue. In their view there is, in addition, a more mechanistic and automatic dimension to production and comprehension in dialogue, which they describe in their “interactive alignment model”. Interactive alignment refers to multi-layered processes that emerge when two people interact, effectively allowing for the direct and automatic alignment of representations and thus bypassing much explicit inferring.

2.2.1 Alignment: an automatic process

Imagine you are taking a walk in the park with a friend. Chances are that you are walking at a different pace than if you had taken the walk alone and the same is true for your friend: Somehow automatically, without thinking about it, you have aligned your steps.

Generally, alignment is something that has been described even for the most basic kinds of human interactions. Richardson and colleagues (2007) found that people seated in rocking chairs next to each other, similarly to a coupled oscillator system, would start to synchronize their rocking movements over time. This happened although the chairs had been manipulated in their natural periods, so that they required rocking in different strengths to achieve the same frequency. Even when participants were instructed to rock at their own pace while in presence of the other, synchronization took place, suggesting that this type of coordination is unintentional and fully automatic. The strength of such visuomotor coupling seems to be “mediated by attention and the possibility of detecting the relevant movement information” (Richardson *et al.*, 2007: 882). This lead to a difference in coupling strength depending on whether participants could directly focus on each other’s movements or only

peripherally. It is also supported by the finding that visual tracking can strengthen an individual's entrainment of an oscillatory stimulus, both when a wrist-pendulum is swung intentionally and unintentionally (Schmidt *et al.*, 2007). In social interactions we find similar patterns: Chartrand and Bargh (1999) describe what they call the "chameleon effect" – unintentional imitation of incidental actions that occur when we interact. These include facial expressions and behavioral mannerisms, like rubbing one's face or shaking a foot, which they were able to elicit in experiments by letting a confederate display this behavior in a standardized way across participants.

Findings like these suggest that joint actions do not need to be fully intentional or goal-directed (which, e.g., Sebanz *et al.*, 2006 had still argued): Instead we can engage in them automatically as soon as we interact. When another person is present, we seem to naturally attune our behavior to theirs and vice versa (cf. Garrod & Pickering, 2009: 295).

It should not be surprising that similar coordination is at work when we communicate with language that, as we have seen, can also be understood as joint action: And indeed, it was found by Branigan, Pickering and Cleland (2000) that syntactic structure is coordinated very much in the same way in dialogue. They had participants play a matching task with confederates, in which participants had to describe the meaning of cards depicting six ditransitive actions (e.g., GIVE or OFFER). The confederate was instructed to vary the syntactic structure of their description, which had an effect on the other speaker's subsequent descriptions. There was a strong tendency for participants to align with the confederate and produce the same syntactic form they had just heard. For instance, if the participants heard a sentence describing a card as *the chef giving the jug to the swimmer*, it would prime the participant to say *The cowboy handing the banana to the burglar* to the next card; whereas if the confederate produced *The chef giving the swimmer the jug*, first, the participant would instead tend to describe the next card as *The cowboy handing the burglar the banana* (cf. also Garrod & Pickering, 2009: 296). Branigan and colleagues argue that similar to syntactic priming effects found for individuals (e.g., Bock, 1986), "prior processing of a particular structure can facilitate subsequent use of that structure" (Branigan, Pickering & Cleland, 2000: B21). The experiment was designed in a way that the results cannot be explained by coordination of mental models or interpretations of specific lexical choices. The authors thus conclude that similar to the conceptual common

ground (Clark, 1996, see above) that requires semantic and lexical coordination, a common syntactic ground is established by activating shared syntactic structures. Cleland and Pickering (2003) extended the experiment discussed here, and showed that similar priming can be shown for complex noun phrases. Participants were more likely to use noun phrases with the same syntactic structures as those they had heard before. This tendency was enhanced when the head nouns were the same or semantically related in both sentences, but not if they were only phonologically similar. Lastly, Hartsuiker, Pickering and Veltkamp (2004) found that syntactical alignment in the card description task even works in cross-linguistic dialogue: When the confederate answered in Spanish, his syntactic choices (specifically passive vs. active voice) got reflected in the participants' subsequent English descriptions.

These experiments on syntactic alignment add to findings that were more concerned with the alignment of "situational models" and lexical choices that reflect different semantic conceptualizations in task-related dialogue. Situational models are "integrated mental representations of a described state of affairs" (Zwaan & Radvansky, 1998: 162; see also Johnson-Laird, 1983). They are multidimensional and encode time, space, causality, intentionality and reference of what is at hand in the current situation (individuals, objects, events). In other words, they represent what is going on in our mind when we understand a text. Several studies have been conducted that investigated how interlocutors align their situational models to solve a collaborative task. Garrod and Anderson (1987) found that participants playing a cooperative maze game align their description schemes of positions in the maze, i.e. the way they lexically represent different situational models of their current position in the maze (see chapter 3) for a very detailed description of this experiment). Likewise, Brennan and Clark (1996) found alignment on the lexical level, i.e. interlocutors using the same referring expressions to refer to specific objects they had to describe. In recent work, it has been shown that this lexical alignment is not only present in dyadic dialogue, but can also be found in triadic situations (i.e. more than two interacting participants), where participants do not only align with the addressee with whom they are performing a collaborative task, but also third persons (for alignment in overhearers see Behnel *et al.*, 2013). The presence of a "third voice" was shown to effectively disrupt lexical alignment to a specific interlocutor, because the speaker will align to the lexical items coming from this third source (Foltz *et al.*, 2015).

There have also been studies showing alignment on phonological levels like word pronunciation (Pardo, 2006), response latency and prosodic elements (Nilsenová & Nolting, 2010). Via coupling of production and comprehension, processes of alignment are happening on all levels of conversations. Let us now consider the accounts trying to explain the underlying processes.

2.2.2 The mechanisms underlying interactive alignment

Several scholars have highlighted different mechanisms that make alignment occur in conversation. One major theoretical account is the mechanistic approach by Pickering and Garrod (2004), who propose lower level priming mechanisms as the basis of interactive alignment. In their view, strategic behavior, as for example suggested by Brennan and Clark (1996), is only necessary for repair processes, e.g. if something is misunderstood. Generally, they see their model as somewhere in between the strategy-oriented language-as-action tradition (including pragmatic approaches and conversational analysts, like, e.g., Clark, 1996) and the language-as-product tradition that – originally rooted in generative linguistics – is largely concerned with the processing and production of isolated, decontextualized sentences (Pickering & Garrod, 2004: 170). The latter approach would therefore only deliver a mechanistic account for monologue, while the former is dealing with dialogue, but only interested in higher order processes (goals, the establishment of common ground etc.), not automated, low-level mechanisms of linguistic processing.

Their interactive alignment model (IAM) can be broken down into three basic main components:

1. Automatic priming mechanisms at all levels of linguistic representation and percolation between these levels.
2. An interactive repair mechanism that comes into play when alignment fails.
3. Alignment via explicit “other modelling”, which is used as a last resort.

The first process assumes that dialogue, as a collaborative effort, becomes successful, if both parties manage to construct similar situational models, i.e. their models become “aligned” (Pickering & Garrod, 2006: 204). This means plainly that both interlocutors share the same ideas of what they are talking about. According to the IAM this does not happen consciously, but via alignment at other levels of representation.

For example, interlocutors reuse phrases or grammatical constructions they have just heard to refer to the same entities, preserving specific form-meaning pairings. Alignment on the lexical and syntactic level is connected: Branigan et al.'s study (2000, see previous section) shows that syntactic priming is especially strong when the same verb was present in both prime and target sentences. Likewise, Cleland and Pickering's experiment (2003) demonstrates that alignment of noun phrase forms was enhanced for semantically related nouns, again suggesting that lexical alignment seems to promote syntactic alignment. These processes happen fully without any attribution of mental states or conscious decision from the interlocutors, simply by linking their various levels of linguistic representations. As these levels become coupled via priming, production and comprehension are not taking place in isolation anymore (as they perhaps would in monologue): What has been said is guiding the speaker and the speaker's utterance in turn constrains how the listener comprehends it, leading to a similar understanding of the situation at hand (cf. Pickering & Garrod, 2006: 213 f.). Interlocutors build a situational model containing the objects and entities they are talking about. At the same time, they activate relevant linguistic knowledge (lexical items, syntactic rules, phonological information) including their own perspective. When one of them produces an utterance based on this, the other will activate the same linguistic information, bootstrapping the construction of the same situational model (i.e. it does not matter whether an utterance is comprehended or produced). Understanding an utterance activates the same situational model that was used to produce it. In addition to this vertical relationship between linguistic knowledge and situational models, Pickering and Garrod (2004: 176) also propose horizontal links via alignment *on every level of representation*, including the level of situational models⁸. The interface between this is routinization, i.e. the use of fixed expressions for particular meanings. This allows decreasing the number of choices that have to be made in production during dialogue, like, e.g., which syntactic voice or synonym should be used. I will return to routinization in more detail after shortly considering the other processes involved in the IAM.

Since alignment is automatic and not conscious, it is still prone to error and can create misunderstandings (e.g. when both interlocutors align in using a lexical

⁸ However, they do not explicate how exactly it is that situational models become directly aligned *horizontally*, that is, other than via (vertical) connections to other (linguistic or other behavioural) levels.

item, but actually mean something different). This is where the second process, repair, comes in. Repair relies on what Pickering and Garrod call the “implicit common ground”. Similar to the traditional sense of the term common ground (e.g., by Clark, 1996, cf. previous section) this refers to inferences about what can reasonably be assumed to be known to both interlocutors on the basis of the evidence at hand (Pickering & Garrod, 2004: 178). The difference is that in the IAM, conversational partners do not have to track this information by maintaining a situational model containing both what they and their interlocutor know about the situation. Instead, the implicit common ground follows automatically from alignment. This works because, as the interlocutors become aligned, they produce and comprehend more and more of the same information that is consequently added to their situational models – extending their implicit common ground, without the need to represent the situational model of the interlocutor.⁹ The more costly establishment of a full common ground would only be needed when massive misalignment (e.g., talking about ‘the President’ and finding that, due to the fact that one of the interlocutors has been in a coma for a while, they have two different people in mind) needs to be repaired. Repair of the implicit common ground can be done via the interactive repair mechanism: First, it is checked whether the input matches one’s own interpretation. If this is not the case, a reformulation allows establishing implicit common ground again (cf. Pickering & Garrod, 2004: 179). Reformulations or “clarification requests” (Ginzburg, 2001) can be iterated by both interlocutors until they reach realignment and provide no longer have to consider the listener’s mental representation (see section 3.3.3 for an example). This kind of repair is interactive, because it requires feedback (e.g., *what did you just say?* or *eh?*).

The third process, explicit other modelling (Pickering & Garrod, 2006: 205), becomes only possible when the full common ground is strategically assessed. This will mostly be the case when the basic mechanism has failed. The listener then has to draw inferences about what the speaker meant. This can happen internalized or via explicit negotiation with the speaker (Pickering & Garrod, 2004: 180) and requires higher reasoning and what has been called a “theory of mind” (Premack & Woodruff, 1978) or “mentalizing” (Frith & Frith, 2006), i.e. the ability to attribute mental states

⁹ It should be noted that this definition of “implicit common ground” has been criticized, because it could very well be described in more simple terms as “shared knowledge”, which is clearly distinct from common ground in the sense of a meta-knowledge (see, e.g., Barr & Keysar, 2004).

to others. Garrod and Clark (1993) found that younger children (7-8 years) could not suppress automatic alignment processes and failed to abandon strategies that were unsuccessful for the task at hand. This finding points to the third, inferential process being rather costly and too hard for less skilled language users (Pickering & Garrod, 2004: 180).

In summary, the IAM highlights automatic priming processes operating at different levels of the conversation, where production and comprehension get coupled and alignment on one level leads to alignment at others. Coordinating linguistic structures thus leads to aligned situational models and a better coordination of the task at hand. In Pickering and Garrod's (2006) view, this is what leads to successful communication, and it could be one reason explaining why interpersonal decision making usually outperforms individual decision making (see, e.g., Bahrami *et al.*, 2012). Fusaroli and colleagues (2012) tested this by having participants solve a collaborative oddball detection task, where dyads had to make private decision, but were also able to discuss a joint decision by communicating with their partner. They found that dyads frequently aligned their expressions of confidence levels (an evaluation of which is required to find a joint decision) and that local alignment of these expressions led to global patterns of recurring lexical items. Further, the results showed that the more dyads aligned, the more successful they were in the cooperative task. Crucially, it was only alignment in task-related expressions that predicted better performance, not alignment in general. Indiscriminate alignment in fact led to less successful cooperation. This suggests that alignment is indeed beneficial to joint problem solving, but the linguistic items that are aligned have to be fit to the task at hand. The findings also show how practices emerge and evolve on the fly: "Through repeated linguistic interactions, global linguistic patterns seem to stabilize from local adaptive dynamics", as Fusaroli and Tylén (2012: 118 f.) put it. Let us now consider this routinization process in more detail.

2.2.3 Routinization

Language is full of routines, i.e. fixed expressions that occur at higher rates than the frequency of their sub-components would predict (Aijmer, 1996). Usually, these are idioms, stock phrases or semi-productive expressions. Kuiper (1996), e.g. showed that sports commentators very repeatedly rely on phrase-long routinized expressions, where they just have to fill in slots, like *X is in the lead*. The permanent storage of

these routines seems to allow much faster production in recurring situations, since they are accessed as a whole.

As we have seen, in dialogue, listeners pick up the utterances their conversational partner produces and subsequently reproduce it to refer to the same present entities or events. This process can be understood as routinization as well, because a new specific meaning becomes fixed, at least for the duration of the ongoing dialogue. Pickering and Garrod (2004: 182) therefore argue that routinization does also happen momentarily and not just by learning over a longer period of time and that the routines that emerge in dialogue are usually temporary and short-lived. Routines allow speakers to ignore many of the lexical, semantic or syntactic choices that would otherwise be necessary for constructing complex expressions, greatly enhancing production. This gives rise to repetition of forms that will activate the same representations and alignment of the interlocutors – they start using the same expressions with the same interpretations:

“A repeated expression (with the same analysis and interpretation) is of course aligned at most linguistic levels. Thus, if interlocutors share highly activated semantic representations (what they want to talk about), lexical representations (what lexical items are activated), and syntactic representations (what constructions are highlighted), they are likely to use the same expressions, in the same way, to refer to the same things” (Pickering & Garrod, 2004: 181).

An example for this routinization process is given in section 3.3.3, where data from the present study is presented.

These *ad-hoc* routines can be seen as some kind of “proto-conventions”, a first step towards a linguistic convention that is permanently tied to a specific usage situation. For now these routines are only momentary strategies that are used to deal with a specific situation (build a coherent situational model of whatever the dialogue is currently concerned with) and they most likely only stay around for the length of the dialogue. However, with the help of additional processes like entrenchment and group convergence, they can turn into full-fledged practices that a whole speech community adheres to. These processes and the tightly connected phenomenon of “cultural evolution” will be discussed after shortly considering the limits of the interactive alignment account as well as alternatives and theoretical additions.

2.3 Limits of interactive alignment and additional concepts

While the interactive alignment model provides a good picture of the mechanistic processes underlying dialogue, subconscious priming only takes place on one level of the dialogue. Some alternative and additional accounts focus on other levels or highlight other processes.

Keysar (2007), for instance, suggests that interlocutors are in fact egocentric when they produce and understand utterances. It has been shown that we frequently generate ambiguity to others, misconstrue referents and often stick to egocentric perspectives even though it leads to failure in an experimental task (cf. Keysar *et al.*, 2000; Wu & Keysar, 2007; see also Ferreira, Bailey & Ferraro, 2002). According to this line of research, people have a tendency for egocentricity in interpreting other's beliefs from early childhood on, and only adults can sometimes recover from this initial tendency (Epley, *et al.*, 2004). Other-centrism requires additional working memory compared to just sticking to one's own perspective (Lin, Keysar & Epley, 2010). It is therefore more economical for the listener to only take the speaker's mental states into account when misunderstandings occur. In this account, cooperation is not the default mode of conversation. Instead, interlocutors merely expect their partner to use their terminology, without taking into account mental states, simply because it is easier to produce. The difference between monologue and dialogue would then only be that the latter allows for overt feedback from the listener that helps the speaker clarify potential points of misunderstanding (Barr & Keysar, 2004). For this kind of exchange, no automatic alignment is needed, the interactive context that provides the repair mechanism is what makes dialogue unique.

This egocentrism-based interactive process, wherein a switch from one's own perspective is only used to correct errors could potentially lead to a lot of misunderstanding. Keysar (2007: 82) speculates that while it appears that miscommunication is rare, much of it might just go unnoticed. In a recent study, Galantucci and Roberts (2014) have shown that this is indeed the case: In their experiment, participants detected incoherent statements from an interlocutor just above chance level. A strikingly large number of participants was unable to notice at all when their chat conversations were crossed with other conversations leading to incoherent turns and descriptions of pictures they did not actually see. This is more in line with a messier account of communication, where communication is usually more noisy and error-prone and

where completely faithful communication requires additional effort. It might also explain why communicative interactions that heavily rely on faithfulness, e.g., between pilot and control tower, require strict protocols to ensure success (cf. Galantucci & Roberts, 2014: 5). There is some support from Foltz et al.'s (2015) triadic experiments for this egocentric account. Participants in their study broke alignment with the addressee and instead aligned to a third voice that they heard over headphones. One explanation could be a mix of implicit priming mechanisms leading them to align with the voice (interfering with interlocutor alignment) as well as the egocentric tendency to use the most recent term for convenience. By picking up the voice's utterances, participants saved the effort of coming up with their own description or having to remember that of their partner. An egocentric model would predict such behavior, but the authors admit that a strategic model could also account for the findings: Participants could simply consider the voice's contributions as more "important" coming from some sort of authority. Research has shown that socially asymmetric communication situations can make a difference in low-level conversational dynamics. For instance, participants in one study aligned lexically more to an interlocutor assumed to be a professor compared to one who was presented as a fellow student (Pirie, 2010). Also, participants in the triadic study showed more addressee-centered behavior in tasks where the risk of miscommunication was lower, suggesting a strategic addressee-centered process (Foltz *et al.*, 2015: 8). Both egocentric and addressee-directed processes seem to play a role in the dynamics of dialogue.

Another line of research is less concerned with the processes of the individuals communicating (and what motivates their individual coordinative actions) and more with the interaction as such. Dale and colleagues (2013) propose a dynamical systems account of social interaction that acknowledges that there is no central control mechanism orchestrating all sub-processes that contribute to it (working memory, social judgement, visual attention, language production and comprehension etc.). This means all these processes lead to the self-organization of interactions on a higher level as soon as face-to-face interaction takes place. When a dynamical system self-organizes, it experiences a reduction in its degrees of freedoms. For multi-part systems, e.g., the multiple muscles involved in executing an action, this synergetic behavior has been used to explain how it is possible that we can actually make coordinated

actions when the amount of possible movements seems so endless (e.g., Bernstein, 1967; Kelso, 1995).

We can observe similar behavior in the form of “interpersonal synergies” in dialogue: (Fusaroli, Rączaszek-Leonardi & Tylén, 2014). At the level of the interaction, dialogues can be viewed as emergent systems, where functional coordination between the interlocutors is taking place. Alignment is one type of functional coordination, since it induces “synchrony” (of linguistic behavior) to the system, similar to the synchrony that was found in the rocking chair experiment discussed above (see section 2.2.1). However, this cannot explain the whole structural organization of a successful conversation, because if we would just imitate one another, the conversation would lead nowhere and eventually break down. It was already mentioned that experiments have shown that it is important which linguistic items are aligned to be successful in a joint task that depends on communication, rather than “blind” or unselective alignment of any linguistic behavior (Fusaroli *et al.*, 2012). Therefore, it is not surprising that syntactic alignment was found less in a corpus of spontaneous conversation, compared to a corpus of conversations related to solving the “Map Task” (Reitter, Moore & Keller, 2006).

Besides synchrony, conversations also depend on **complementariness**, i.e. that both partners take on complementary roles. This means they have to have somewhat different situational models and produce different utterances that together allow them to distribute their resources to fit the task at hand (cf. Fusaroli, Rączaszek-Leonardi & Tylén, 2014: 149). When we talk to each other and coordinate turn-taking, we do not start to speak at the same time, which requires some kind of anti-phase locking instead of synchrony. Even speech-impaired patients can engage in complex conversations as long as this reciprocal coordination, this “rhythm” is maintained by others (cf. Dale *et al.*, 2013: 72 f.). Conversations are normally also organized in functional sequences: If you greet me, I respond. If you ask me a question, I give an answer, and so on. Earlier in this chapter I discussed how these patterns were long considered to belong to a shared set of assumptions that was given when people engage in a joint activity (e.g., by conversational analysts like Schegloff). However, proponents of the interpersonal synergy account claim that instead, during dialogue, these patterns evolve in response to the context. Conversations often start rather unstructured, but then over time, turns become more operationalized (often including

different roles for each interlocutor) to talk about what's at stake more efficiently. Mills and Gregoromichelaki (2010) showed how in the classical maze game, much less turns are usually needed to solve the task towards the end of the experiment. Over time, participants who are playing this cooperative game start to develop interactional routines, e.g., giving a position description of an important point in the maze, asking for clarification etc., and they start to distribute complementary roles: For instance, if one player's description cannot be followed by the other, she in turn will simply provide a description herself, and so on, until the goal is reached (this was also found in the present experiment, cf. section 3.3.3). Their interaction became increasingly more coordinated by profiling only relevant aspects of the task and flexibly coordinating roles, effectively reducing degrees of freedom and thus the need for elaborate utterances and negotiation.

The interpersonal synergy account highlights the role of **function** in dialogue. What is said is said for a specific purpose and not just to be understood (which is the main point of interactive alignment). In order to fit the present function, interlocutors have to take into account their context and goal; they have to select "relevant dimensions for alignment, distribut[e] roles and compensat[e] for mistakes" (Fusaroli, Rączaszek-Leonardi & Tylén, 2014: 150). Importantly, these dimensions and roles can only be understood in relation to the joint activity and the functional organization of the interaction as a whole. When the function becomes the object of coordination, there is a proper motivation to select from the multitude of possible behaviors and processes available those that best fit the activity (Louwerse *et al.*, 2012). This way, the sheer complexity of the conversation is reduced by structuring the dialogue via function-oriented routines. Of course, there is a multitude of functions that dialogues can serve; for instance, small talk serves more of a social function (establish or preserve affiliation), while the dialogues elicited in almost all of the experiments discussed in this thesis (including my own) serve to solve a joint problem. To fit these purposes, turn-taking and other recurrent interactional patterns have to be tightly coordinated. In a recent study, Fusaroli and Tylén (2015) analyzed how pairs communicatively solving a joint decision problem established routines via recurrent interactional patterns. They found that the more the overall conversation (not the independent contributions) was characterized by recurrence of lexical, prosodic and speech/pause sequences, the more successful was the pair. These measures at the

level of the interaction were a better predictor of success than interactive alignment/synchrony as well as individual self-consistency (using the same structural sequences, but analyzed for each individual) which were also measured. The authors conclude that alignment and synergies seem to capture different aspects that are employed to fit the task. It is therefore not a single mechanism that makes conversation work effortlessly.

In summary, the research reviewed in this section, highlights that conversations are far too complex to be described in fully mechanistic models. Social and contextual factors always play into human interactions – something the interactive alignment model (but also other psycholinguistic theories of production and comprehension) do at best marginally account for. Dynamical systems approaches to human interaction (e.g., Dale *et al.*, 2013) are attractive to describe this complexity, because they allow integrating multiple mechanisms, behaviors and contexts with concepts such as self-organization or synergies that account for the flexible adaptation of structures to an interactive task. As I will argue, especially affordances of the situation in which dialogue takes place can also have an impact on communicative strategies and processes. I will show how it is not only the task that puts demands on the interaction, but different affordances can further constrain which routines are selected for, in order to meet these demands.

2.4 From routines to conventions: the concept of cultural evolution

These processes of alignment, synergetic loops and coordination of routines are not just momentary, but continuously and restlessly shaping systems of social interaction on a larger, historical, time-scale. Throughout the lifetime of each and every human, there is an accumulation of endless interactions that keep affecting the whole community. If we would imagine a still pond that represents a current cultural state, every social interaction would be a raindrop landing in this pond, creating a circularly expanding wave. As long as there are interactions, it will keep raining and the surface of the pond never becomes a flat state, but will stay an ever-changing complex pattern of intersecting waves, where it is impossible to make out what change was caused by which drop. In other words: Socially defined cultural systems are permanent subjects to change, due to the ongoing dynamics between their sub-parts. Culture evolves as an outcome of many small local changes, without any central control unit prescribing what the set of conventions it is comprised of should be. Cultural practices are, in this

sense, emergent and constantly changing, similar to biological organisms. Therefore, processes of cultural change and conventionalization are often referred to as “cultural evolution” in analogy to biological evolution (e.g., Whiten, *et al.*, 2011).¹⁰

As we have seen, conventions can emerge in interaction, simply because there is a recurring need to solve the same coordination problems within a community and they provide ready-made solutions (cf. Lewis, 1969). This applies to normative cultural conventions, e.g., how we dress to attend certain events, or right-hand versus left-hand traffic, but also to more fundamental forms of culture, like language. Lewis (1975) argued that a language is nothing more than a set of conventions that a speech-community agreed on – trusting and expecting each member to use the same words to refer to the same meanings (“common knowledge”). This type of linguistic conventions then allows for speakers to solve coordinative problems within the community for which the use of language is required, e.g., buying a cup of coffee or asking for the way to a location etc. (cf. Tylén *et al.*, 2010). Earlier in this chapter, it was shown that automatic alignment can lead to routinization on several linguistic levels in an *ad-hoc* manner during this kind of problem coordination. But how do these temporary conventions spread within a community and live on as permanent linguistic conventions, once the dialogue is over? Garrod and Doherty (1994) tackled that question by adapting the earlier cooperative game from Garrod and Anderson (1987) in a version that compared isolated pairs playing the game with communities of players. They found that there are both, global and local processes governing the coordinative choices that players make. Local coordination between isolated pairs led to local alignment (routinization) of strategies that differed between pairs. However, when pairs were made up of varying partners belonging to a (simulated) bigger community, global patterns became visible, where precedence was an important factor affecting the choices made on the group level, whereas local salience was often overridden by the community’s conventions. Community pairs were first less coordinated than isolated pairs, but at later stages of the experiment became much more coordinated, due to the establishment of group-wide conventions. Once certain strategies spread to become common knowledge, the players’ choice of linguistic items was less affected by small local changes/salience than in the isolated condition. The authors attest this to “a global co-ordination model which assumes pressure to con-

¹⁰ Many scholars have also highlighted a *co-evolution* of culture and genes as the process that transformed *homo sapiens* into the successful species we are today (see, e.g., Richerson & Boyd, 2008).

verge with each new partner linked with dominance of the most commonly used past scheme” (Garrod & Doherty, 1994: 202). In addition to the pressure to coordinate their language with any partner (as in the isolated condition), there seemed to be a global pressure leading the participants to stick to the most common scheme they had previously encountered. A mechanism like that can be held responsible for how groups converge on a single convention and how this convention stabilizes, which might be true for any cultural practice. Tomasello (2014) describes it as the leap from “joint intentionality” to “collective intentionality” that first occurred in modern humans in our phylogenetic history: We became able to establish norms not only for how *you* and *me* do things, but also how *we*, as a cultural group, do them. In this sense, the history of local interactions can determine the global common ground, i.e. set of conventions that is established (cf. Tylén *et al.*, 2010: 57).

Once conventions are established, they do not necessarily remain fixed. There is a continuous potential for them to “evolve” – they can change to fit current demands more economically. This kind of cultural evolution has been proposed for language, amongst other phenomena. For a long time mainstream linguistics – influenced by theories of generative grammar – assumed that a set of fundamental rules which they refer to as universal grammar (UG) underlies all languages and “generates their grammars” (e.g., Chomsky, 1980). Any grammatical convention, like, e.g., word order, aspect or marking of subject/topic/case, could therefore be traced back to this UG. However, while formal grammars like the ones developed in the course of the generativist program provide good description tools, extensive fieldwork and comparative research with non-Indo-European languages increasingly challenge the idea of universals. The languages of the world do, in fact, display a striking variation across all linguistic domains; including morphosyntax or semantic categories that make it hard to discern universals (cf. Croft, 2010; Haspelmath, 2007). For example, some languages do not have a negative marker (like English *not* or Danish *ikke*); Tamil for instance deletes a tense morpheme from the positive instead (Pilot-Raichoor, 2010). There is also a wide range of different basic color terms that languages use, e.g., English uses *blue*, where Russian has two basic words *goluboy* and *siniy* to refer to darker and lighter blue hues (Winawer *et al.*, 2007). Other languages instead subsume green and all blues under one term, but yet another language, surprisingly, uses a basic term for turquoise (Davies, Sosenskaja & Corbett, 1999). There have also been dis-

cussions of some languages not even possessing the most basic fundamental property of UG proposed by its proponents (e.g. Hauser, Chomsky & Fitch, 2002), namely recursion (Everett, 2005), although this remains hotly debated. Either way, a simple glance at the World Atlas of Language Structures (WALS) easily reveals a large amount of differences for almost any grammatical property.

Evans and Levinson (2009) argue that this diversity is the outcome of cultural evolutionary processes. Grammatical choices, on this account, are simply structures that speech communities agree upon and share, similar to the emerging routines in Garrod and Doherty's community experiment. Simulations (Christiansen & Chater, 2008) as well as experiments with human subjects (Kirby, Cornish & Smith, 2008) have shown that implicit learning biases could then be responsible for diachronic change over time. This means that each generation of language users exerts some pressure on the specific forms that are preserved, modified or lost over time. In a way, one can contrast these diachronic, cumulative transmission dynamics governing language change as **vertical** processes with synchronous **horizontal** processes, taking place when people interact. Numerous Iterated Learning¹¹ studies (see Kirby, Griffiths & Smith, 2014 for a review) as well as agent-based models of cultural language evolution have shown how vertical processes can shape the form of languages: Compositionality, case-markers or agreement systems have been observed to emerge in such models (see Steels, 2011 for a review). It seems, again, that just like routines in dialog, these grammatical conventions seem to emerge to fit a specific function and to yield a successful outcome in a joint activity. For instance, Beuls and Steels (2013) show how in a population of agents playing a "naming game", agreement systems evolve to minimize parsing complexity and ambiguity of expressions. Agreement markers emerged from scratch or were recruited from existing words. Gradually, they also "eroded", i.e. they became shorter (e.g., *-uinbui* > *-uin* > *-u*), to reduce effort when playing the game. This shows how horizontal processes contribute to vertical dynamics at every point

¹¹ Iterated learning (IL) experiments rely on diffusion chain paradigms where subjects that are initially trained on an artificial language (often random and unstructured for the first generation) are tested in their knowledge of that language and, by doing so, generate output that becomes the input for the next generation. This allows us to observe structure that emerges or changes over time, simply in virtue of cultural transmission and learner biases. Such experiments have been done with simulations (Smith, Kirby & Brighton, 2003) and human subjects (Kirby, Cornish & Smith, 2008). They also test transmission and learning of other cultural practices, e.g., tool-making (Morgan *et al.*, 2015). Most of all, the IL approach has been proven to be an effective way of studying how learning biases can lead to specific structures evolving without any intention or control unit.

in time, because communicative interactions *per se* can already constrain what is symbolically realized (Healey *et al.*, 2007) and then over time, changes accumulate and a language evolves. In this model, grammar is the outcome of handling problems that arise when a symbolic communication system is built, similar to how routines in dialog are responses to a coordinative problem.

People interact with one another in social groups, where they communicate to solve coordinative problems. While solving these problems, alignment, interdependence and complementary lead them to routinize conventions and patterns on the level of the interaction. This happens again and again and they begin to rely on strategies that were previously successful over the history of past interactions. Over time, the routines spread in the community and a cultural common ground is established, e.g. the set of conventions that comprise a language and constrain its use in specific situations. In this view, languages are not fixed systems, but rather continuously shaped by self-organization and selection through interactions among its users (cf. Steels, 2011).

2.4.1 Constraints on cultural evolution

If we accept cultural evolution as the process shaping linguistic structure and leading to socio-culturally diverse conventions, there still remains one question: Are the resulting conventions arbitrary or driven by further, external factors? Put differently, are the different grammars Evans and Levinson (2009) describe to argue against universals simply the outcome of different groups finding different solutions for the same problem? Tylén and colleagues (2013: 57) argue that

“[...] it is only in concrete situations of contextualized language use that certain structures are selected for and realized in linguistic categories, construals, and conceptual models. The selective realization of some structural affordances rather than others gives socio-cultural variability to human meaning-making.”

This quote entails several concepts that are important in trying to answer the above question: **situation/context**, **affordances** and **construal/conceptual models**.

Let us consider them in relation to a concrete example: One prime argument that proponents of linguistic diversity and cultural evolution often put forward concerns spatial language. A rich body of comparative language data shows high variations in how topological relations, movements or distant object relations are expressed (e.g., Levinson & Wilkins, 2006, Majid *et al.*, 2004). This supports Levinson’s

(1996, 2003) theory of different Frames of Reference (FoR), which are employed in these languages to construe spatial relations. According to Levinson's theory, there are three distinctive FoR: relative, intrinsic and absolute. Consider the spatial scene depicted in figure 2 below:

Fig. 2. A spatial scene that can be described in to different coordinate systems (Frames of Reference) in English (adapted from Levinson, 2003: 25).

In English, it is possible to describe the position of the FIGURE (F, the *ball*) in relation to the GROUND (G, the *car*) in different ways: In (1), **intrinsic** properties of G (the functional *front*) determine the ball's location. In (2), a third coordinate, the viewpoint (V), is used to locate F in relation to G. V can either be the perspective of the sender or receiver, defining where *left* is and how F and G are arranged in space. This strategy is therefore called **relative**. A third strategy, the **absolute** FoR, locates F and G in respect to a fixed direction or landmark such as *west* in (3).

Curiously, while all FoR can be expressed in English, many languages employ only one or two out of the three (cf. Levinson & Wilkins, 2006). Tzeltal speakers, e.g., solely use an *uphill-downhill* system, relating all objects to the giant slope of their geographical region in Tenejapa, Mexico (the ball would be, e.g., *uphill of the car*). The Guugu Yimithirr (Australia) also rely on the absolute FoR, but use the cardinal directions. Interestingly, Tzeltal speakers appear to have problems recognizing left-right distinctions and Guugu Yimithirr can accurately show the cardinal directions at any time, exhibiting what has been dubbed a *mental compass* (Levinson, 2003). This, supported by evidence across other domains (Everett, 2013 for an extensive review), led to a revival of the linguistic relativity hypothesis and sparked discussions about how language and general cognition might influence each other.

It is important to note that this implies that the linguistic conventions speakers of a language adhere to are also conceptual choices that *categorize* and structure a *situation* in a specific way (in this case: its spatial relations). But what determines which strategy/*construal* is chosen?

Majid and colleagues (2004: 112) speculate that external factors like the environment (open vs. dense terrain or urban vs. rural), habitual actions (e.g., specific subsistence patterns) or global cognitive styles (individualism vs. collectivism) could mediate between language and cognition and influence the choice in FoR. They present data from 21 languages suggesting some (with such a small sample size statistically not verifiable) correlations between choice of FoR and ecological factors. For example, in tropical rainforests, the relative FoR seems to be more frequent, whereas desert dwellers favor the absolute.

If we think of culture as a form of adaption to solve a problem, this should not be surprising. Cultures are usually embedded in a specific environment providing them with a specific set of problems and *affordances* (see the following section). This could be one (of various) selective factors in cultural evolution that has been neglected. Small languages often rely on surrounding landmarks (LMs), like in Tzeltal (see above) or in Mian, where two nearby rivers are used to locate objects (Fedden & Boroditsky, 2012) and even to talk about time. Intuitively, a certain cultural strategy should be more adapted in a certain environment and get selected over time. *Left/right* may not be very helpful in the desert where absolute directions seem more reliable for orientation. In a dense jungle or urban area, however, turns can hardly be memorized and far sight is disabled; a relative/intrinsic system might become adequate.

Specific situations with specific affordances might thus promote one conceptual strategy over another, because it is more adaptive (or salient). In a simulation, this has been modelled for color: Linguistic color categories that emerged during interaction turned out to reflect the range of colors present in the stimuli the agents encountered in different regions (Puglisi, Baronchelli & Loreto, 2008). For real languages, it has been suggested that color categories could be correlated with different degrees of sunlight exposure in various regions (Lindsey & Brown, 2002). Very recently, it has also been suggested that tonal languages developed in more humid climates (Everett, Blasi & Roberts, 2015) and languages spoken in warmer areas with

more dense vegetation tend to have more vowels (Maddieson & Coupé, 2015). These could be acoustic adaptation in direct response to physiological affordances of the environment.

Correlations like that are hard to corroborate, especially because there could be hidden additional factors causing such results. Agent-based simulations, on the other hand, are just models of very simplified realities. It would therefore be desirable to test this role of environmental affordances in actual humans coming up with linguistic conventions in the laboratory. Garrod and Doherty already mention small differences in description schemes among subject pairs playing differently shaped mazes in their maze game. The authors note that “Presumably this reflects the salience of a particular scheme in relation to positions being described” (Garrod & Doherty, 1994: 185). The present case study directly follows up on this by focusing on this very environmental aspect. An experiment was designed that systematically adds even more salient environmental affordances to the original paradigm, to test whether this alone could influence the conventions that dyads come up with. Since environmental “affordances” have been mentioned quite a few times by now, I will give a short definition before concluding this chapter.

2.5 The theory of affordances

The concept of affordances and the term itself were introduced by James Gibson (1966, 1979) who was an active contributor to the field of ecological psychology. He and his colleagues pointed out that human behavior cannot be studied as an end in itself, since humans are inherently situated in an environment with which they are tightly interconnected.

Within this framework, affordances are functional relationships between an organism and its environment that unfold through perception. There are specific things the environment “offers” an organism to act upon. For example, a flat surface affords support. A small metal object affords grasping, but if an electric current flows through it, it also affords electric shock (cf. Gibson, 1979: 143). Affordances can thus be beneficial or harmful; they comprise all invariant properties of the environment (including other organisms within that environment) that are connected to a potential behavior. Perception is really important here, because it constraints affordances significantly: An agent has to perceive the affordance or he cannot act upon it. This

was famously illustrated by Duncker (1945) who posed the “candle problem”, a problem-solving task, to experimental subjects. The task was to fix and light a candle on a wall without wax dripping on the table below. Participants were provided with a box of matches and a box of thumbtacks to complete the task. Most subjects failed to see the affordance of the thumbtacks box as a supporting device for the candle that could be fixed to the wall. However, more subjects were able to solve the task when that affordance was made more perceptually salient by placing the tacks outside the box. An affordance is invariant, but it has to be identified as such (cf. Gibson, 1979: 138 f.).

Since perceptual salience can profile affordances, it is a good candidate for a selecting pressure towards a specific strategy when several are afforded by the situation. Consider people interacting to achieve the same mutual goal in different environments that possess the same affordances, but vary in how salient these affordances are: The salience causes them to recognize one ecological property over another and base their solution (e.g., a linguistic description) on that affordance (Figure 3).

Fig. 3. A schematic representation of a problem-solving process based on affordances. First, the problem is encountered. In three environments (A, B, C), the same three solutions are afforded, but differ in their salience. Strategies are then selected in response to the most salient affordance, leading to different solutions for the same problem in different environments. Note that these are probabilistic, not deterministic outcomes – so that in any given sample, few may still choose the perceptually less salient options.

It is therefore conceivable that linguistic conventions arising during interaction are motivated by salient affordances in the environment where these interactions are situated.

2.6 Summary

In this chapter, I have reviewed research that emphasizes the interactive facet of language, with face-to-face conversation being its most typical and common mode. Communicating with a partner is viewed as goal-directed coordinative action embedded in a joint activity. Linguistic communication becomes a tool to make this engagement successful. We saw how during such coordinative interactions, priming mechanisms lead interlocutors to automatically align their language and situational models. This gives rise to linguistic routines that seem to be also partly motivated by egocentric choices and modulated with the addressee in mind (e.g., in socially asymmetric conversation). At the same time, synergies and complementary structures at the level of the conversation also help to fit the conversation to the task and coordinate from the myriad of possibilities in conversations those behaviors that are necessary for the joint activity at hand. When we solve a crossword puzzle together, we will start taking turns, where as soon as one of us knows a word, will say the number and direction. If we have different ideas, we will coordinate our confidence, for which we will likely converge on a scale over time that we can rely on, e.g., *I guess this could be..., I'm sure this could be..., I know this is...* (similar to the scale in Fusaroli *et al.*, 2012).

Normally, the linguistic conventions that emerge in such episodes are short-lived, because they are only needed momentarily to fit the task at hand, but if many members of a community face the same problems over and over again, they can rely on these successful conventions and institutionalize them. When a whole group starts to converge on one specific linguistic routine, it becomes a cultural convention. We can view the diverse structures of the languages of the world as sets of such conventions that emerged as the outcome of many local interactions, where members of a community aligned their lexicon and grammar to refer to the same situations and objects in the same way, thus creating a common cultural ground that made it easier to take part in collective activities. Languages in this sense are culturally negotiated complex adaptive systems that keep changing as new situations require new routines, or simply because a more efficient convention is found to address the same collective

problem. That way, languages gain new vocabulary (like the word *smartphone*, to describe a new technology) or lose cases that are not needed to efficiently describe states of the world (e.g., Old English versus modern English).

It seems that these conventions are driven by the pressure to address problems in a maximally efficient and economical way. This suggests they might not be fully arbitrary, but actually respond to external demands. Considering Gibson's theory of affordances, I hypothesize that the very same collective problems afford for different solutions when they are situated in different contexts. The functions of the conventions that arise in these settings might be the very same, but different environments afford – or even just profile – slightly different solutions. Some simply possess very salient properties that can easily be utilized, while others might require more abstract representations to deal with the same problem. To test whether this is true for linguistic routines that emerge spontaneously during dialogue, an experiment was conducted that will be discussed in the following chapter.

Chapter 3: Case study

3.1 How environment affects linguistic coordination in the maze game

In order to address the research question – if and how environmental affordances can influence the structure of linguistic conventions that emerge as routines in dialogue – a situation had to be created, where subjects engage in a “joint activity” (Clark, 1996) that varies only in terms of the environment in which it takes place. If environmental factors are indeed at play when linguistic coordination takes place, routines occurring in that very same joint activity are expected to differ selectively in response to these environments. Hence, an experiment was conducted that builds on previous work and allowed for observations of linguistic alignment in the laboratory. The “maze game” by Garrod and Anderson (1987) was adapted and extended with new environmental conditions. This allowed for participants to be put in situations in which they had to actively coordinate linguistic forms and align their lexical repertoires in order to reach a common goal. At the same time, affordances provided by the environment could be systematically manipulated, while all other factors remained constant across conditions. Relying on this previously established paradigm also enabled a comparison of the present findings with previous studies regarding the specific linguistic description schemes they have identified. I will elaborate on that in the discussion section after having described the experimental design and results in this chapter.

3.1.1 The maze paradigm

Garrod and Anderson’s (1987) original study (hence G&A) was an important step in linguistic interaction research, insofar as their paradigm, for the first time, allowed the experimental exploration of linguistic co-ordination in a constrained dialogical context. The game was specifically designed “to elicit natural dialogues containing spontaneously generated descriptions of locations within a predefined spatial network, where the exact positions described could be independently verified by the experimenter” (G&A, p. 185). These dialogues emerged over the course of successive interactions, where dyads of participants had to guide each other through virtual mazes to solve the task.

The mazes were designed in a way that required a collaborative effort from both players in order for them to reach their (individual) goals. Participants were seated in front of different computers, each being able to see only one of two versions of the same maze made up of box-like rooms that were arranged in a 7×7 grid (see figure 4 below). Each room had between one and four connections to other rooms, through which players could pass one at a time by moving their position marker. The final goal of the game was for both players to move their position marker to their respective finish room (marked with an x). When both players' position markers reached their destination, the game was solved.

A difficulty was posed by the fact that some of the connections were blocked by black lines, referred to as "gates". These gates were distributed differently for both players and could not be passed as long as they remained closed. This is where the game became co-operative: In order to open the gates and reach the finish box, both players had to work together. Gates could only be opened by using switches. On their screen, players could see marked rooms where the switches for their gates were located (see rooms marked with an s in figure 4 below). Counterintuitively, it was their partner who had to enter these rooms to activate the switch. This means that if player A moved onto a field where player B had a switch (which player A could not see on his/her screen), then all of player B's gates that were previously closed would open and all gates that were previously open would close.

Fig. 4. Layout of a maze from the original study. The overall shape was identical for players A and B, while locations for switches (s), finishes (x), gates (black lines crossing connections) and start positions (*) vary (Figure from Garrod & Doherty, 1994: 184).

The game thus required the two players to inform each other about the position of their switches, gates and their own position markers in order to open the path towards the finish box. Since there is no (pre-existing) conventional way of talking about positions in these box-type mazes, successfully solving the task involves negotiating a common referential scheme – that is, commonly recognizing the same position in response to a given linguistic expression. The emerging “language refinement” thus reflects “the response of the conversationalists to the challenge of communicating efficiently in a locally constrained and novel situation”, as Garrod and Doherty (1994: 214, hence G&D) put it.

G&A’s original study used an additional typology controlling for the size and shape of the mazes and in some conditions introduced a “monster” – a third, semi-intelligent artificial agent marked as an M, who automatically moved and killed the player occupying the same field as the monster, in which case the game was lost. Games with a monster seemed to put participants under a stronger pressure to cooperate, but it was left out in subsequent studies. Similarly, Garrod and Anderson also manipulated maze shape and size, but neglected these variations in the analysis. Instead, the study was focused on how participants align linguistically to different conceptual description schemes in order to solve a communicative problem. I will now turn to these findings in more detail.

3.1.2 Findings from the original studies

In both the original experiment (G&A) and later follow-up studies (Garrod & Clark, 1993; G&D), it was found that participant pairs adopted very similar types of descriptions to communicate locations which were associated with different mental models of the maze configuration. Although the pairs in the original G&A study were isolated (that is, every participant only interacted with one constant partner), there were recurrent patterns in the language the different dyads produced that allowed to subsume almost all location descriptions that occurred in the dialogues under four basic schemes:

1. *The PATH scheme*

In this scheme, the verbal description of a location in the maze relies on the description of a path, usually with a start and an end point that the partner is expected to follow.

Here are two examples:

- (1) See the bottom right, go two along and two up. That's where I am. (G&A, p. 189)
- (2) B: Right guide me to a switch. I'm two along from you.
A: Right.
B: Two from the right from you. (G&D, p. 190)

In (1), the start point is the room in the bottom right corner of the maze. The path then follows two rooms to the left (on the horizontal axis, since there is no room that can be entered on the right side in the corner) and then turns up, again passing through one more room and then reaching the destination in the second room. In (2) the same scheme is used, but this time player B's position is used as the starting point of the path.

One thing that is important for the PATH scheme is whether rooms, connections between rooms or both are counted in the description. This had to be negotiated when establishing the scheme, because otherwise it would remain ambiguous (cf. G&D, p. 191).¹²

2. The *MATRIX* scheme

The matrix scheme (originally called CO-ORDINATE by G&A) is based on a conceptual grid/coordinate system that is superimposed onto the maze. The description of a position in the maze is then achieved by referring to an intersection of the horizontal and vertical lines that make up the grid:

- (3) I'm on the third row and fourth column. (G&A, p. 190)
- (4) I'm at C4. (G&A, p.190)

As can be seen in these two examples, participants either named the axes of the coordinate system (e.g., *rows* and *columns* like in (2)) and counted them, or they applied a more abstract code by labelling each line with a symbol (e.g., similar to a chess board as in (3)). In the latter example, alphabetic letters have been assigned to all vertical nodes, and numbers to all horizontal nodes. The position can then be inferred by looking for the room that lies on the vertical line C and horizontal line 4.

3. The *FIGURAL* scheme

Some participants also relied on particular figural shapes they recognized in the maze, such in the following examples:

¹² There was also a special kind of PATH descriptions that G&D refer to as "Goal-related", but these did not matter for the present study discussed below (see G&D for a discussion of this scheme).

- (5) See the rectangle at the bottom right, I'm in the top left-hand corner. (G&A, p. 190)
- (6) A: I'm not in that L shape I'm in the like the second row downwards - column downwards. (G&D, p. 192)

First, a sub-part of the maze is described as a figural configuration, e.g., a *rectangle* in (5) or an *L shape* in (6) and then a position relative to this shape is indicated (e.g., *top left-hand corner* in (5)).

4. The LINE scheme

The LINE scheme again depends on an abstract representation of the maze, this time in terms of several lines that can either be a horizontal, vertical or diagonal array of rooms:

- (7) Third bottom line, third box from the right. (G&A, p. 190)
- (8) Third column from the right, two from the bottom. (G&A, p. 190)
- (9) I'm one up on the diagonal from the bottom left to top right. (G&A, p. 190)

In all of these examples the description presupposes that the maze has been conceptually divided into parallel lines (that can also be referred to as *rows*, *columns*, *levels* etc.) which are then numbered. The difference to the MATRIX system is that these are only defined on one dimension rather than two. To be able to describe a location according to the LINE scheme, it first has to be agreed upon whether lines are counted horizontally (such as in (7)), vertically (8) or diagonally (9). Then the line of interest is specified (e.g., *third bottom line* in (7)) after which a room lying on that line is specified (e.g., *third box from the right* in (7)).

Frequency of description schemes

The frequency of the different description schemes varied a lot across the different studies. In G&A's original experiment, PATH was the most commonly used scheme across all pairs (36.8% of the corpus) followed by MATRIX (23.4%), LINE (22.5%) and FIGURAL (17.3%). There was some variation between the two games played by the pairs, though: In the first game, most speakers used PATH predominantly ($n = 14$), followed by FIGURAL and LINE (both $n = 9$) and only very few speakers used MATRIX ($n = 8$). However, in the second game, this turned around and more speakers relied on LINE or MATRIX (cf. G&A, p. 195). G&A identified general trends to transfer from FIGURAL to PATH and MATRIX as well as from PATH to LINE. They interpret this as speakers initially concentrating on more perceptually salient strategies (PATH & FIGURAL) and later moving on to more abstract ones (LINE & MATRIX).

G&D had slightly different distributions in their experiments. In the first one (isolated pairs vs. community groups), MATRIX was the most predominant scheme among all pairs in the community condition and from game 5 (out of 9) onwards accounted for 100% of description exchanges. G&D attribute this to convergence onto the most popular scheme (as MATRIX accounted already for 58% of descriptions in early games). Isolated pairs behaved more like pairs in G&A's study using both LINE (27%) and MATRIX (40%) steadily across all games. This suggests that there is a lot of variability, after which isolated pairs stick to either of the two more abstract schemes for the rest of the games, whereas communities rapidly converge onto one single convention that becomes established over the course of several games. Since it could not be ruled out that MATRIX was only used so excessively because partners switched every round and MATRIX was somehow more salient or conservative to use in such a situation, G&D conducted a second experiment, in which partners were exchanged but not drawn from the same community. Here, LINE turned out to be the most predominant scheme, accounting for 44% of all descriptions in the first 5 games, supporting the conclusion that MATRIX did indeed only occur so frequently in the community condition because it proved most successful as a community convention.

Across all maze game studies, the authors mostly discuss the results as examples of automatic alignment in task-related dialogue. Participants come up with different semantic descriptions and quickly routinize these to solve the task. Isolated pairs coordinate a convention (that may differ between pairs), whereas communities are initially less coordinated, but eventually converge onto a single convention.

Unfortunately, G&A, who manipulated maze shape, do not discuss any impact of maze size (varying between 4×4 and 6×6 rooms) or shape (square-shaped mazes, horizontally symmetric and horizontally asymmetric ones), suggesting there was none. However, later G&D (p. 185) discuss that "regular mazes" (perfectly symmetrical, similar to, e.g., maze 10A in appendix I) made it more likely for players to choose a MATRIX or LINE scheme, whereas "irregular mazes" (like the one depicted in figure 4 above) led to slightly more FIGURAL or PATH descriptions. G&D call this "local salience", which along with precedence is considered to be one factor determining which description scheme is chosen.

In summary, the maze game studies provided great examples for emerging linguistic conventions in a real-time joint problem solving context and an important

empirical basis for the interactive alignment model (see chapter 2). However, they do not tell us much about the role of external factors that could also affect these conventions.

3.2 Experiment: A variation of the maze game

To test whether environmental affordances could systematically affect the alignment and routinization processes that have been described in previous maze game studies, the paradigm was adapted by adding novel “environmental” conditions. In these conditions, the spatial layout of the mazes was carefully controlled for to test the hypothesis that this environmental factor alone could systematically influence which of the four different description schemes players would adopt to describe locations in the mazes.

Whereas the original maze game studies relied on spoken data that was subsequently transcribed for data analysis (Garrod & Anderson 1987: 187ff.; Garrod & Doherty 1994: 189f.), the present study utilized a graphical user interface (GUI) including a chat tool as means of communication. The interface allowed for direct analysis of the data collected during experimental sessions without the need of transcription.

The Dialogue Experimental Toolkit (DiET, see Mills & Healey, submitted) was used¹³. The tool is specifically designed to investigate the various levels of dialogue and supports fine-grained word and turn-level interventions as well as the possibility to manipulate dialogical content, e.g., introducing artificial clarification requests or other manipulations.

It should be acknowledged that the chat-tool put some constraints on dialogical dynamics and communicative flexibility for the participants. Paralinguistic cues, such as prosody/intonation or iconic vocalizations (such as discourse markers to express confidence or similar mental states which could assist the task) were restricted, in addition to the visual domain (including, e.g., co-speech gestures) that was already inhibited in the original studies, but would otherwise accompany normal dialogical situations. Furthermore, the nature of the chat interface forced participants to take turns in a linear order, as each submitted “utterance” would appear in a linear sequence after the previous one in the form of a turn in the chat protocol. If a partici-

¹³ A Java-based program, released under GPL license that provides a text-based chat-tool platform to carry out dialogical experiments.

pant was sending a message, there was no possibility for her partner to interrupt her. On top of that, participants saw when their partner was typing, which could have further discouraged them from producing language while their partner was doing so. In general, the chat interface caused more structured turn-taking and less flexibility as would be expected to occur in natural dialogue (cf. Cohen, 1985) ¹⁴.

3.2.1 Materials

The experiments took place at the computer laboratory of Aarhus University's Cognition and Behavior Lab. The experiment room was equipped with eight computers and seats separated by wooden walls to prevent participants from seeing their neighbor's screen (see figure 5). All computers were connected to a server located in an adjacent room outside, where the DiET program was launched.

Fig. 5. Laboratory setup. Identical colors indicate participants playing the maze game together via interconnected computers.

Participants saw a chat window on the right side of the screen, where they could type with their keyboard and see messages their partner was sending. Similar to other

¹⁴ Although, as one at least could argue in favour of the paradigm, this form of chat and short messaging has become a very natural form of communication in recent years, and thus participants were most certainly accustomed to its use.

chat interfaces, they could also see when their partner was typing a message, making the conversation more dynamic. The chat window could also be used by the supervising experimenter to communicate with the participants, e.g., if problems or doubts about the game occurred in the chat. On the left side, both players saw the current maze they had to solve and their position marker within the maze which they could move around by holding the Ctrl-key and pressing the arrow keys in the desired direction. Both windows are shown in figure 6 below.

Fig. 6. Left: Maze window, showing the player A's goal (red X), switches (grey rooms) and gates (black lines blocking connections) as well as her position marker (black dot) which can be moved around. On the right, the chat window is shown. The upper black field shows the chat protocol, where new messages from both players appear in chronological order. The lower part shows the edit field, which is used to type and send messages (by hitting the return key). On the grey area in the middle between both fields it is indicated if the other player is currently typing a message.

The maze game mechanics were mostly adapted from the previous experiments by Garrod and colleagues, but also slightly changed:

- There were much less gates and switches compared to the previous studies, often enabling players to solve one maze within a few chat turns.
- If a switch was activated, gates remained open only as long as the player who activated it remained in the room where the switch was located.

- Players had to remain at least 2.5 seconds on the switch before it was activated. This was to prevent a “lazy” solution, where players just walk around randomly and wait for their partner to tell them when his gates would open (since this would undermine the need to communicate linguistically). Effectively, none of the pairs taking the experiment was caught trying this method.
- In order to investigate the influence of environmental affordances, three environmental conditions were added to the experiment.

All other aspects from the original maze game remained identical: For each maze, there was one version for player A and a different one for player B, wherein the general room-layout on the 7×7 grid was the same for both players. Only start, finish, switches and gate locations differed between these two mazes, as is shown below:

Fig. 7. An example maze: To the left is player A’s maze and to the right that of player B. Note that the overall layout is the same, while the initial start position, gates, switch rooms and finish room vary for the two players.

The key difference between the present study and previous maze game studies was the introduction of three novel conditions that featured mazes with systematically differing layouts, thus altering the task environment. All three conditions were designed in a way that was hypothesized to saliently afford for one of the previously identified description schemes (see 3.1.2).

Regular mazes

The regular condition featured 11 mazes that were designed in a very neutral or “traditional” way in the sense that they were most similar to those of previous studies. The rooms were distributed on the 7×7 grid in a mostly square-like fashion. The den-

sity of rooms was rather high and there were not many perceptually salient aspects or specific areas that were profiled in virtue of foregrounded shapes. Neither rows nor lines were especially prominent. The regular condition was hence the closest to a control condition, affording all description schemes to the same extent. It was therefore expected that MATRIX or PATH would be the most dominant description types, as they had been in previous versions of the maze game (cf. Garrod & Doherty, 1994; Garrod & Anderson, 1987). Figure 8 depicts two examples of regular mazes (for a full list of all mazes from all conditions, see Appendix I):

Fig. 8. Two mazes from the regular condition. While maze **a** is a bit smaller, both feature an overall layout that extends to the same degree on the vertical and horizontal axes and does not feature shapes or sub-areas that are especially perceptually salient.

Irregular mazes

By contrast, the irregular condition featured 11 mazes that were highly non-uniform in their design. Here, figural aspects of the mazes were profiled and made perceptually salient, in that every maze had particular irregular areas that stood out. These could be sub-areas consisting of several rooms resembling geometrical or abstract shapes like an L-shape, a cross or square as well as protruding areas that could be conceptualized as extremities sticking out from the main part. Furthermore, most of the irregular mazes had an overall shape that could be interpreted in various ways, e.g., as the shape of an animal or object. These overall shapes could in turn be segmented again into body-parts or functional segments. For instance, the maze in figure 9a below could be interpreted as a *dog* or virtually any other four-legged animal seen in profile. This allows to segment the maze further into body-parts, like two *legs* in the lower part of the maze, comprised of two rooms each: A room that represents the *tail* (on the far right) and three rooms on the upper left side making the *head* that could yet again be segmented into, e.g., a *snout* (room on the far left) and *ears* (room

on the very top). The point is that the specific interpretation does not matter, but rather the general affordance to see figures or shapes in the mazes of this type. The maze could also be interpreted as a *cat* or *horse* or simply as an abstract object characterized by a *main part* (eight central rooms) and various *secondary parts* that stick out from it.

Fig. 9. Two mazes from the irregular condition, the left one resembling a more concrete figure (*animal*), the right one being more abstract, but still highly irregular and figural.

The maze in the second example (Figure 9b) is more abstract. There are two very recognizable segments that appear rather similar. These non-generic shapes could, for example, be interpreted as *towers*, *candles*, *houses* or simply *the big one* (on the left) and *the small one* (on the right). The missing rooms within these two structures could be called *holes* or *windows* and the area connecting them could be referred to as *bridge*.

It should be clear at this point that the rationale behind the irregular condition was not to provide participants with a specific figural interpretation of the maze that they were expected to use, but rather with a rich set of visually prominent affordances that allow for figural representations of the maze that in turn could be used to communicate spatial descriptions by referencing these aspects in the chat. It was hence expected that participants playing the irregular condition would more frequently converge on the FIGURAL scheme than those playing mazes from other conditions.

Stratified mazes

Lastly, the stratified condition provided players with a set of 11 mazes that were characterized by highly salient horizontal displacements. These were achieved by building all mazes from two components: 1) ‘Lines’, i.e. several rooms connected on the horizontal axis. 2) Vertical connections between these lines, either via a single connection or extended over a single room that was serving as a pathway between two lines (figure 10). The latter version increased the saliency of lines by creating a lot of ‘free’ area around them. With some additional variations (e.g., some vertical connections to dead-end rooms), this design was applied to all mazes in the stratified condition.

Fig. 10. Two mazes from the stratified condition with very salient horizontal ‘lines’ that are only sparsely connected on the vertical axis. Maze a) features some additional rooms between lines and a dead-end room in the lower part, maze b) consists merely of lines.

This design was chosen to evoke LINE descriptions among participants playing this condition. It was expected that the saliency of these horizontal arrays would afford their representation as a conceptual whole, so that the maze could first be segmented into lines and then a specific position on that line could be identified. The mazes that feature bridging rooms between lines further give the advantage of clearly distinguishing between the LINE scheme and other description schemes like MATRIX. For example, the switch room in figure 10a could on the vertical axis be identified as *three down* according to the matrix scheme, whereas according to the LINE scheme it would be on the *two down*, since it is located *on the 2nd line from the top*.¹⁵ The row including

¹⁵ Alternatively, if counted from below, it would be the *3rd line*. Different dyads using the LINE scheme counted from different sides (cf. section 3.3.3). The direction didn’t matter for successful communication – only, that both partners were aware from which side they were counting.

the connection room is not counted in this description scheme. From analyzing the linguistic utterance in relation to the actual position in the maze, it becomes thus clear that one and not the other underlying conceptualization of the maze was used.

Prediction

The rationale behind these three conditions was that when participants repeatedly encountered mazes providing the same kind of perceptually salient affordances, an environmental pressure would emerge that would render certain description types more likely to be used in the first place, reused again and finally conventionalized. It would be expected that – out of all different ways to communicate positions in the maze – those description schemes repeatedly proving themselves adaptive in the task-environment would be stronger attractors for the dyad playing this environment. They would become linguistically entrenched over time and manifest themselves as conventionalized solutions the dyad can fall back on as new trials unfold.

In summary, the predictions for all three conditions were as follows:

- **Regular condition:** Due to generic and neutral design of the mazes, PATH and MATRIX strategies that had been most dominant in previous studies are expected to be most attractive and become predominant in this condition as well.
- **Irregular condition:** Due to salient environmental affordances (figural shapes), the FIGURAL strategy is expected to be more attractive and dominant among dyads in this condition.
- **Stratified condition:** Due to salient environmental affordances (horizontal displacements), the LINE strategy is expected to be more attractive and dominant among dyads in this condition.

3.2.2 Participants

For the study, 72 Danish speakers were recruited through the CoBe lab's online system and took part in the experiment in return for 100 DKK they received as compensation. Out of these, 3 dyads were excluded, in two cases because one partner was not a native speaker of Danish, which led to communication problems. In the third case, the dyad was unable to solve all trials in the maximum amount of time (see procedure below). The remaining participants ($n = 66$, 42 female, 24 male) were all higher educated ($M_{\text{Education}} = 15$ years, $SD = 2.2$) young adults ($M_{\text{Age}} = 23$ years, $SD = 3$).

3.2.3 Procedure

Data was collected over the course of three days, with up to eight participants per session. One experimenter remained in the room with the participants (to prevent them from talking or looking at each other's screens), the other experimenter monitored the ongoing games outside on the server and, if problems occurred in the games or chat, could go back and help the participants.

Participants were not told which computers would be connected and therefore could only find out who they played with after the experiment was finished. Before the game started, participants gave their informed consent as well as data relevant for the payment and were then instructed regarding the game mechanics and chat program (see appendix III for the detailed instruction sheet). Each dyad was then allocated to one of the three conditions (regular, irregular or stratified) and subsequently had to solve all 11 mazes from that condition in a random order as well as a 12th, final maze which was the same for all conditions (Figure 11).

Fig. 11. Final maze that appeared after all condition-specific mazes had been solved.

The final maze designed in a way that equally afforded all strategies. It featured figural parts, like the 4 rooms that could be described as a *square* or the overall shape that could be seen as a snake with *head* and *tail*. Furthermore, the density of rooms and long vertical arrays, make it easy to impose a line or matrix system onto the maze. The idea behind this was to check whether participants would still use the predicted strategy of their condition outside of the condition-specific trials.

The maximum time for each maze was 7 minutes; if the goal was not reached by both participants at that time, they automatically proceeded to the next trial. After all mazes were completed, or after the time was up, participants were presented with a questionnaire, where they were asked to enter demographic information as well as given the opportunity to comment on the strategy they used and their idea of what the experiment was about.

Most participants guessed that the experiment had to do with communication, problem-solving or cooperation, but none of them guessed that there were different environmental conditions or that it was about different description strategies. However, one participant who played the irregular condition actually suspected that it was about “how seeing figures in abstract shapes of the mazes helps the communication” (In Danish: “Om man ser figurer i abstrakte mønstre så som labyrinter og om det hjælper på kommunikationen”). This illustrates the salience of the affordances in this condition.

3.3 Results

3.3.1 Coding of chat protocols

After all data was collected, all chat protocols were coded according to a pre-defined coding scheme. The unit of description was “turns”, i.e. full chat lines that had been sent by one of the players. For the analysis it was only necessary to code those turns that contained a position description like the first turn in this extract:¹⁶

- (10) P56: 2 grey fields. row 3 field number 2 from right. And bottom row field no.2 from left
 P55: Now?
 P55: THanks
 P56: Now I am locked
 P56: *2 grå felter. 3 række nr. 2 felt fra højre. Og nederste række nr. 2 felt fra venstre.*
 P55: *Sådan?*
 P55: *TAk*
 P56: *nu er jeg fanget*

In (10) only the first turn would be coded, in this case as belonging to the LINE strategy. The following turns include only comments related to the game, such as participant 55 asking whether her partner’s gates would have opened (*Now?*) or participant 56 mentioning that she would be unable to reach her finish now, since the gates were blocking her now. These latter turns were disregarded for the analysis. Position description turns (like the first in (10)) were coded as belonging to the PATH, MATRIX, FIGURAL or LINE description scheme (see appendix IV for details of the criteria). There was

¹⁶ Note that the experiment took place in Danish, so for all examples I provide English translations as well as the original from the chat log (grey font).

also a 5th category called UNDEFINED for position descriptions that could not be assigned to any of those four basic categories, mostly because they were too vague, like:

- (11) I am circa in the middle
Jeg er ca i midten
- (12) your switch is on the other side of my door now, quite annoying
din switch er lige på den anden side af min dør, ret træls

Descriptions like that certainly refer to a position in the maze, but were too general and vague to be interpreted as belonging to a strategic scheme. Generally, UNDEFINED turns like this occurred only rarely (3% of coded descriptions), demonstrating the validity of the coding scheme developed by Garrod and colleagues. UNDEFINED turns were excluded from later analyses.

The full corpus contained 4841 turns from the 33 dyads that were coded¹⁷ for position descriptions, which resulted in 1260 annotated turns. Additionally, 27% of the corpus (3 dyads from each condition) were coded by a second coder to calculate inter-coder-reliability (cf. Landis & Koch, 1977; Fleiss, 1981), which turned out substantial (Cohen's $\kappa = 0.7$). The full coding scheme according to which the coding was done can be found in appendix IV.

3.3.2 Examples of different description schemes

Besides some turns that contained UNDEFINED descriptions, as mentioned above, it was found that participants did indeed use the same description schemes that G&A and G&D had already identified in their previous studies. Coded examples of each type are shown below:

PATH descriptions

- (13) I have two. one is in the uppermost right corner, 3 to the left, 1 down and 1 to the right
jeg har to. det ene er øverst højre hjørne, 3 til venstre, 1 ned og 1 til højre
- (14) mine is uppermost left, 1 right, 1 down, 1 right
min er øverst venstre, 1 højre, 1, ned, 1 højre
- (15) right corner on top. 1 down
højre hjørne oppe. 1 ned
- (16) if you are in the left corner and go three down
hvis du er i venstre hjørne og går tre ned

¹⁷ Coding was done blindly, without seeing which maze belonged to which chat, in order to prevent biases.

MATRIX descriptions

- (17) When this is a 6x6 grid where it is counted horizontal, vertical – (5, 1)
Hvis det er et 6x6 grid hvor man tæller vandret, lodret - (5, 1)
- (18) I cannot come to the finish, have you been to e1 or f5?
jeg kan ikke komme i mål, har du været på e1 og f5?
- (19) I am there now. Cannot come to 2,3
Der er jeg nu. Kan ikke komme til 2,3
- (20) mine is in 2nd column from left and second row from the top
min er i 2. kollone fra venstre og anden række fra toppen

FIGURAL descriptions

- (21) There are 9 in a square. The uppermost in the middle
Der er 9 i en firkant. Den øverste i midten
- (22) at the bottom to the right in the middle arm
nederst til højre i midterste arm
- (23) at the bottom to the right in the rectangle in the left corner
Nederst til højre i firkanten i venstre hjørne
- (24) on the "bridge" in the middle of the big square
til "broen" i midten af den store firkant.

3.3.3 Examples of alignment and synergy in the chat data

Inspecting the chat protocols confirmed that these linguistic strategies emerged *ad hoc* and became routinized to solve the task efficiently, as would be expected from the results of previous studies.

- (25) P37 Where is your switch?
P38 2nd "row" from the bottom
P37 Can you come down to it while I stand on my switch ?
P38 Yes
P37 Can you go from your switch to the finish without me standing on mine ?
P38 No
P38 I stand on my switch now
P38 Is your door open
P37 No
[...]
P37 Third row from the top, fully to the right, there is my switch
P37 *Hvor sidder din switch?*
P38 *2. "række" fra bunden*
P37 *Kan du komme ned til den mens, jeg står på min switch ?*
P38 *Ja*

P37 *Kan du gå fra din switch og i mål, uden jeg står på min ?*
P38 *Nej*
P38 *Jeg står på min switch nu*
P38 *Er din dør åbnet?*
P37 *Nej*
[...]
P37 *Tredje række fra toppen helt til højre, der er min switch*

This example is from a pair playing their first maze. Right at the beginning (second turn) P38 establishes the *LINE* scheme and the lexical convention *row* on which it is based. After a few lines of discourse P37 picks up this convention and adapts it to refer to another position in the maze. From this point on the *LINE* strategy becomes fully routinized for the total duration of the interaction and is used almost exclusively by the pair, whenever they need to describe a position to each other (accounting for 98% of their descriptions). Just like in the original studies, participants repeatedly picked up each other's lexical items and reproduced them, supporting Pickering and Garrod's (2004) claim that interactive alignment took place while they tried to solve the task.

There was also evidence for the interactive repair mechanism described in the Interactive Alignment Model (Pickering & Garrod, 2004: 180):

- (26) P01 one in the left corner, 1 up, 1 left
P02 I am in the left corner. so I go one up, but I can't go to the left? the maze ends there
P02 if you understand. is it fully down in the corner?
P01 shit, I also meant the right corner
P01 *et i venstre hjørne, 1 op, 1 venstre*
P02 *jeg er i venstre hjørne. så går jeg en op, men jeg kan ikke gå til venstre? mazen stopper der*
P02 *hvis du forstår. er det helt nede i hjørnet`?*
P01 *shit, jeg mente også højre hjørne*

In this example, both players have already routinized *corner* as an expression to refer to the outermost points of the mazes which are used for their *PATH* strategy as a starting point. However, P01 causes a misalignment of situational models by referring to the corner on the right as the *left corner*. P02 struggles to fit this information to his situational model of the task (i.e. where the switch is located). This is due to the fact that in his situational model *left corner* refers to a position where it is impossible to go further left. P02 therefore requests a more specific description, which makes P01 realize his mistake and repair the communication by reformulating the position of the

switch. Another example, more in line with what Pickering and Garrod call “repair using implicit common ground” is (27):

- (27) P55 3rd row, fully on the left
P55 otherwise I can’t come up and open for you
P56 alright
P55 No.. Are you on row 3, the one with 5 fields, fully to the left
P56 I didn’t think you also counted the 1 field rows. I go up ;-)
P55 *3 række med, længest mod venstre*
P55 *Ellers kan jeg ikke komme op og åbne for dig*
P56 *sådan*
P55 *Nej.. Er du på 3 række, den med 5 felter længst mod venstre*
P56 *troede ikke du talte de der 1 felters rækker med. jeg går op ;-)*

Here, the repair sequence is a result of miscommunication. The communication fails when P56 confirms that she has reached the position that P55 described in the first turn. P55 gives feedback that the doors did not open and specifies the same location, with a more detailed reformulation (by adding *the one with 5 fields*). P56, who was so far under the impression that she was at that position, becomes aware of the misalignment and repairs by stating that she conceptualized the description differently by counting only the rows that contained more than one room (1 *field*) and that she will go to the correct position now. She is able to do this by relying on the implicit common ground, namely the perception of the two different possibilities there are to count the rows (counting the 1-field-rows versus not). She does not explicitly have to model a full common ground (including everything her partner might see, know and have done in the past), but the mere realization that 1) there are two different counting systems available 2) she was using one of them and 3) that one did not work out lead to the conclusion that 4) her partner must have been using the other one.

There was also evidence for increasing structural organization (synergy) at the level of the dialogue as the games proceeded (Table 1).

Maze 6		Maze 7		Maze 12	
P48	bottom left corner and two up	P47	b3	P47	b2
P47	should we make a coordinate system and say that vertical are letters and horizontal are numbers starting from the upper left corner?	P48	b4 and d4	P48	e3
		P48	cannot go to b3	P47	I'm in
		P47	so?		
P48	and bottom right corner, two to the left and one up	P48	yes am in finish		
P48	good idea!	P47	I have d2 and b3		
P48	E1 and F5	P48	is it open?		
P47	does something happen?	P47	I'm in the finish		
P48	where are yours?				
P47	I have one called e5 :)				
P47	okay maybe it doesn't work				
P48	here?				
P47	bottom right corner three up, three to the left				
P47	yes				
P47	I'm in the finish				
P48	I can't come into the finish, have you been to e1 and f5?				
P47	now I am at e1				
P48	so				
Maze 6		Maze 7		Maze 12	
P48	nederste venstre hjørne og to op	P47	b3	P47	b2
P47	skal vi lave et koordinatsystem og sige vertikal er bogstaver og horisontal er tal startende fra øverste venstre hjørne?	P48	b4 og d4	P48	e3
		P48	kan ikke gå til b3	P47	jeg er inde
P48	og nederste højre hjørne, to til venstre og en op	P47	så?		
P48	god ide!	P48	yes er i mål		
P48	E1 og F5	P47	jeg har d2 og b3		
P47	sker der noget?	P48	er der åbent?		
P48	hvor er dine?	P47	så er jeg i mål		
P47	jeg har en der hedder e5 :)				
P47	okay det fungerer måske ikke				
P48	her?				
P47	nederste højre hjørne tre op, tre til venstre				
P47	yes				
P47	jeg er i mål				
P48	jeg kan ikke komme i mål, har du været på e1 og f5?				
P47	nu er jeg på e1				
P48	så				

Tab. 1. Chat protocols for a pair playing two regular mazes and the final maze (including Danish original).

Until trial 6, this pair relied mostly on PATH descriptions, as is still evident from P48's contributions at the beginning. P47 then introduced the MATRIX system by explicitly negotiating how they should count. At first, there are some problems in succeeding with this new system, leading P47 to fall back on PATH (*bottom right corner [...]*). P48 insists on using MATRIX (*have you been to e1 and f5?*) and finally they solve the maze with this strategy (P47: *now I am at e1*). From then on their exchanges become more and more elliptical as they routinize turn-taking relying on the MATRIX scheme. Each turn of a sequence becomes routinized to have a specific function/meaning: players state the locations of their switch (requesting their partner to go there), ask whether their partner's gates opened or confirm that they reached the finish (Maze 7 in table 1). This allows them to solve the last maze with two very short descriptions (*b2* and *e3*) in the final maze simply by following this sequence.

Fig. 12. A graph showing the overall number of turns per trial produced by a pair and both players individually.

Figure 12 is a visualization for this pair showing how over time the number of produced turns decreased. The peak in the middle at game 6 results from the negotiation of the new strategy, which becomes routinized, subsequently leading to even shorter turns (see Mills & Gregoromichelaki, 2010 for a more elaborate discussion on sequential routines). Having confirmed that the same conversational dynamics and routinization processes occurred in this new maze game, we can now ask, whether the environmental condition had any effect on these.

3.3.4 Impact of environment

Analysis of predominant description schemes

Examining the coded chat protocols revealed that the four description schemes were distributed quite differently among the four conditions. Table 2 displays the frequencies, i.e. how often a description belonging to each scheme occurred in each condition (summed up over all pairs):

Condition	PATH	MATRIX	FIGURAL	LINE	UNDEFINED
Regular	108 (29%)	107 (28%)	19 (5%)	130 (35%)	10 (3%)
Irregular	44 (11%)	54 (14%)	191 (49%)	93 (24%)	10 (3%)
Stratified	0 (0%)	71 (19%)	11 (3%)	283 (76%)	9 (2%)

Tab. 2. Frequencies of description schemes per condition. Numbers correspond to counts of turns coded for the description type and percentage (per condition) is shown in parentheses

For the regular condition, the distribution of schemes appears varied, with all schemes corresponding to roughly a third of the occurrences, except for FIGURAL which occurred only rarely (accounting for 5%). This is contrasted by the irregular condition, where FIGURAL accounts for about half of the position descriptions that participants produced, while the other schemes occurred much less frequently. Similarly, in the STRATIFIED condition, the LINE scheme occurred much more often than all other schemes, in this case even accounting for three quarters of all descriptions produced. This already suggests that there might have been an impact of condition on the frequency of the different description schemes: The schemes that were predicted to be influenced by the salient affordances of the irregular and stratified conditions (FIGURAL and LINE) did indeed occur more frequently in these two conditions. Only in the regular condition, neither the predicted PATH nor MATRIX scheme did occur more frequently than other description types.

However, this result could be due to confounding factors, since the overall frequencies for each condition may obscure what was actually going on among each pair. Therefore, frequencies were also calculated for individual pairs, which allowed deriving the predominant strategy for each pair (i.e. the description that occurred most frequently in the chat protocol of a pair; see figure 13).

Fig. 13. Number of pairs using a given strategy as predominant description type (33 pairs across all trials).

This pair-wise analysis reveals a similar pattern. In the irregular condition, six out of ten pairs used the FIGURAL scheme predominantly, two relied on MATRIX, two on LINE. None of the pairs in this condition used PATH predominantly. By contrast, in the stratified condition nine out of eleven pairs relied mostly on the LINE scheme, two on the MATRIX scheme and none on PATH or FIGURAL. The regular condition gives a more varied picture, with five out of twelve pairs predominantly using PATH, four pairs relying on MATRIX and three pairs relying on LINE. Similar to the overall frequencies (Table 2), FIGURAL seems to not have occurred much among pairs in the stratified or regular conditions, where no pair used this strategy predominantly, but it became the most popular scheme in the irregular condition, where more than half of the pairs used it for the majority of their descriptions. Similarly, LINE, again, was much more widely used in the stratified condition than in the others. This supports the implications by the overall distributions, suggesting that the irregular and stratified conditions did indeed evoke more FIGURAL and LINE descriptions, respectively. Also, like in the overall frequencies, there is no indication that PATH or MATRIX would have been more popular in the regular condition. Instead, all description schemes except for FIGURAL were used dominantly by pairs in this condition.

Final maze analysis

As described in section 3.2.3 the final maze was kept constant across conditions and was designed in a way to afford no strategy more than any other. Thus, it could be tested whether participants would preserve the convention they established in a specific environment, even when changing to a one that allowed them to use other strategies. If so, this would lead to systematic differences between the three conditions for the last maze, although there was no difference in the task design at all.

One problem that was not anticipated was that at this late stage, participants were usually highly entrained and well-coordinated and needed only very few lines of chat to solve the task. Many pairs therefore only needed two description turns to solve the last maze (one referring to each switch). If both descriptions relied on different schemes, it was not possible to assess a “dominant” description scheme for this trial. Thus, four pairs had to be excluded from this part of the analysis. Nevertheless, the results indicate differences between conditions.

Fig. 14. Number of pairs using a strategy as predominant description type during the last maze.

The irregular condition looks more varied in this graph. Here, FIGURAL is not the most common dominant strategy; 7 pairs relied on different strategies to solve the last trial. In the stratified condition, however, most pairs relied on the LINE scheme and in the regular conditions we see examples of all strategies except for FIGURAL, which was the same in the analysis of all trials. This could suggest that FIGURAL was generally less stable than the other strategies, or that the maze did not afford for this strategy as

strongly as for the others, after all. From the qualitative inspections above we cannot draw definite conclusions yet, so further assessment was required.

Attraction analysis of description schemes per condition

The above analysis already suggest an interaction between condition and chosen description scheme, but the small sample size (10-12 pairs per condition) did not allow for meaningful statistical comparison. To be able to test for statistically significant differences and get a more detailed picture of the dynamics underlying transitions from one strategy to another, it was decided to consider the attraction strength, i.e. how likely it is for a pair to switch to a certain scheme and stay with it.

For each pair, all turns that contain a location description can be seen as a discrete time Markov chain (Spedicato, Kang & Yalamanchi, 2015) that is a sequence of fixed possible states where the distribution of the forthcoming state X_{n+1} depends only on the current state X_n . The transition probability p_{ij} to move from one possible state s_i to s_j is $p_{ij} = Pr(X_1 = s_j | X_0 = s_i)$.

In the present case, each point of the sequence corresponds to a description that follows either the PATH, MATRIX, FIGURAL or LINE scheme or is UNDEFINED. Cleaning the chat protocols from all turns containing only UNDEFINED descriptions that comprised only a small percentage of the data (see above), then allowed to describe each chat protocol as a Markov chain with a state space containing the four main description types.

The states were coded as 1=PATH, 2=MATRIX, 3=FIGURAL, 4=LINE AND 5=UNDEFINED. An example would be this sequence of a pair playing the irregular condition: 35313553315133334343333333333133333333. If all UNDEFINED turns are excluded, the sequence will be slightly shorter: 3313331133334343333333333133333333. This sequence describes all subsequent location descriptions this pair produced that correspond to one of the four description schemes. From this we can calculate the transition probabilities between these description types/states and represent them in a transition matrix.¹⁸

$$P = \begin{bmatrix} 0.25 & 0.75 & 0 \\ 0.11 & 0.81 & 0.07 \\ 0 & 1 & 0 \end{bmatrix}$$

¹⁸ The Markov analysis was done in R with the markovchain package (Spedicato & Kang, 2015).

In this case, the matrix contains only three states, since MATRIX did not occur among the descriptions of this pair. Some transitions (e.g., from LINE to PATH) did also not occur, resulting in zero probabilities. The strongest probabilities for this pair were to move from PATH to FIGURAL (0.75) and to keep using it (modelled as ‘transition’ from FIGURAL to FIGURAL) in the next turn (0.81) rather than switching. This provides an accurate description characterizing their predominant use of FIGURAL descriptions.

Using this method we can calculate a transition matrix for each pair showing the transition probabilities between all states (e.g., the probability to transfer to PATH after using LINE). By averaging the matrices of all pairs per condition, we get three average transition matrices. To get a better estimate, this was done with bootstrapped samples ($n = 100$) for each condition, resulting in the three matrices:

Regular Condition				
	PATH	MATRIX	FIGURAL	LINE
PATH	0.88	0.03	0.05	0.03
MATRIX	0.02	0.97	0.01	0
FIGURAL	0.54	0	0.05	0.41
LINE	0.04	0.02	0.06	0.89

Irregular Condition				
	PATH	MATRIX	FIGURAL	LINE
PATH	0.33	0.07	0.42	0.18
MATRIX	0.02	0.92	0.02	0.04
FIGURAL	0.1	0.01	0.69	0.2
LINE	0.06	0.02	0.4	0.51

Stratified Condition				
	PATH	MATRIX	FIGURAL	LINE
PATH	0	0	0	0
MATRIX	0	0.82	0.01	0.17
FIGURAL	0	0.09	0.12	0.79
LINE	0	0.04	0.02	0.94

Tab. 3. Bootstrapped transition matrices containing the average transition probabilities between description schemes for all three conditions.

The tables reveal differences between conditions in the probabilities of pairs switching between particular description schemes in each condition. These differences are visualized in figure 15-17.

Fig. 15. Transition graph for the **regular** condition. Arrows indicate transitions between states with line thickness as a function of probability.

Figure 15 shows that in the regular condition, MATRIX was the strongest attractor with a 97% chance of pairs reusing it in a following turn rather than switching to another scheme. LINE and PATH are also very stable at almost 90% each. FIGURAL, by contrast, was a very low attractor in this condition, with only a 5% probability that two FIGURAL descriptions would occur right after one another. Instead, there was a 51% probability of switching to PATH and a 41% chance of switching to LINE after producing a FIGURAL description. All other transitions were rather improbable (< 6%), suggesting that in this condition participants usually stuck with the description scheme they first converged on and hardly switched to others. The only exception was the FIGURAL scheme which was highly unstable. FIGURAL descriptions had a very high probability of being followed by LINE or PATH which then remained almost certainly stable.

Fig. 16. Transition graph for the **stratified** condition. Arrows indicate transitions between states with line thickness as a function of probability.

The stratified condition (figure 16) gives a different picture. PATH was no attractive strategy here – in fact, it did not occur at all. Therefore the graph does not show any transitions to and from PATH. MATRIX was again a very attractive strategy, with 82% attraction to itself. However, unlike in the regular condition, it was not the most attractive scheme, since LINE had an even higher attraction to itself (94%) and was also attractive to switch to from less attractive schemes. FIGURAL was again rather unattractive with only a 12% chance that participants would stick to it in following turns, once they were using it. Instead there was a 79% chance that a pair would switch to LINE after a FIGURAL description. Remarkably, there was 17% chance for participants to switch from the otherwise very attractive MATRIX scheme to LINE, further supporting

the idea that LINE was a much stronger attractor in the stratified condition than in the regular.

Fig. 17. Transition graph for the **irregular** condition. Arrows indicate transitions between states with line thickness as a function of probability.

In the irregular condition (figure 17), MATRIX was again the strongest attractor, with a 92% chance of pairs reusing it in the next turn. Surprisingly, FIGURAL was the second strongest attractor (69%) and had very low probabilities of pairs switching away from it. Instead, there was a relatively high chance of pairs transitioning from PATH or LINE to FIGURAL (42% and 40%). This suggests that FIGURAL was much more stable and attractive in the irregular condition, in comparison to the others, where participants frequently switched to other schemes after using it.

To compare the different description schemes across conditions and test whether they differed significantly in how strongly they “attracted” participants to

keep using them subsequently, the attraction strength of each scheme was calculated for all 33 pairs. This can be done by summing up all transitions to a certain description scheme and then dividing by the overall number of descriptions. This represents how “absorbing” each state of the Markov chain was for a specific pair (Table 4).

Pair	Condition	PATH	MATRIX	FIGURAL	LINE
1	regular	1	0	0	0
2	regular	0.92	0	0.04	0.04
3	regular	0	0.78	0.02	0.20
4	regular	0.10	0	0	0.90
5	stratified	0	0	0.04	0.96
6	stratified	0	0	0.04	0.96
7	stratified	0	0	0	1
8	irregular	0.04	0.71	0.18	0.07
9	irregular	0.12	0	0.82	0.06
10	irregular	0.14	0	0.67	0.19
11	irregular	0.07	0.07	0.59	0.26
12	irregular	0	0	0.04	0.96
13	irregular	0.15	0	0.41	0.44
14	irregular	0.12	0	0.72	0.16
15	stratified	0	0	0.04	0.96
16	stratified	0	0.40	0	0.60
17	stratified	0	0	0.02	0.98
18	stratified	0	0.57	0.09	0.34
19	stratified	0	0.89	0	0.11
20	regular	0.87	0	0.04	0.09
21	regular	0.97	0	0.03	0
22	regular	0.45	0.55	0	0
23	regular	0.62	0.26	0.13	0
24	regular	0.08	0	0.08	0.84
25	regular	0	0	0.04	0.96
26	stratified	0	0	0.07	0.93
27	irregular	0.13	0.76	0	0.11
28	irregular	0.14	0	0.76	0.10
29	stratified	0	0	0	1
30	stratified	0	0	0	1
31	regular	0	1	0	0
32	regular	0.14	0.75	0.11	0
33	irregular	0.16	0	0.51	0.33

Tab. 4. Attraction strengths of all description schemes for each pair showing how likely it was for a pair to keep using each scheme.

A linear model was used to test for relationships between condition and the attraction strengths of a scheme (with respect to the predictions for each condition). Ac-

cording to my original prediction, the MATRIX and PATH schemes were expected to be stronger attractors in the regular condition. This is not the case for MATRIX (*adjusted* $R^2 = -0.04$; stratified condition: $\beta = -0.11$, $SE = 0.14$, $p = 0.45$; irregular condition: $\beta = -0.12$, $SE = 0.14$, $p = 0.40$). However, the regular condition had a somewhat stronger attraction on the PATH strategy (*adjusted* $R^2 = 0.35$; stratified condition: $\beta = 0.55$, $SE = 0.14$, $p = 0.0004$; irregular condition: $\beta = 0.02$, $SE = 0.14$, $p = 0.91$). For the stratified condition, I predicted a stronger attraction on the LINE strategy than in the other conditions. This was found to be the case to some extent (*adjusted* $R^2 = 0.33$; irregular condition: $\beta = 0.11$, $SE = 0.14$, $p = 0.35$; regular condition: $\beta = 0.43$, $SE = 0.11$, $p = 0.0004$). Lastly, for the irregular condition I predicted a higher attraction on FIGURAL. This is indeed found (*adjusted* $R^2 = 0.58$; stratified condition: $\beta = -0.44$, $SE = 0.07$, $p = 0.000001$; regular condition: $\beta = -0.42$, $SE = 0.11$, $p = 0.000002$). In summary, no condition had a significantly stronger attraction on MATRIX. PATH was a stronger attractor in the regular condition, especially compared to the stratified condition. LINE was more attractive in the stratified condition, especially compared to the regular condition. Finally, FIGURAL was a stronger attractor in the irregular condition compared to other conditions.

3.4 Discussion

The experimental results indicate that there are two general pressures influencing the choice of description schemes: 1) An environmental bias, which resulted in the selection of strategies that are more adapted to specific affordances present in the stratified and irregular conditions, 2) a general tendency to use the MATRIX strategy, similar to what G&D observed in their community condition. The fact that every scheme could be relatively stable suggests that re-use is generally likely, as predicted by the interactive alignment model. Thus, the LINE scheme was the strongest attractor in the stratified condition, while the FIGURAL scheme was the second strongest attractor after MATRIX in the irregular condition, and significantly more attractive in this condition than in any other. In accordance with previous results, FIGURAL did not out-compete MATRIX as an attractor, even in the irregular condition, because there seems to be a general tendency to move from FIGURAL to MATRIX or from more concrete to abstract in general (G&A, p. 195 f.). It is furthermore remarkable that besides being a slightly weaker attractor than MATRIX, the FIGURAL scheme was still the dominant

scheme in the irregular condition, whereas it was extremely rare in the other conditions as well as in earlier maze game studies (G&A; G&D). In my interpretation, the results suggest that FIGURAL is a very primitive (and also highly context- and speaker-dependent¹⁹) strategy that becomes easily overridden, once participants discover another scheme, presumably because it is not very effective. However, in an environment with a more salient affordance for FIGURAL descriptions, it might be more effective and remain in use longer, as there is no need to switch to another strategy as long as success is achieved. Similarly, the results suggest that the use of LINE was influenced by the salient horizontal displacements affording it in the stratified condition, making it the most attractive and popular strategy.

G&A describe how PATH is a very popular strategy at the beginning, but pairs tend to switch away from it in favor of a more abstract scheme after a while. My results indicate that PATH is indeed a rather conservative strategy, but it seemed to work fine in the regular condition, where it was dominantly used among a proportion of pairs and remained stable. In the other conditions, it became much less used. In the irregular condition, there was a strong tendency to switch to FIGURAL instead and in the stratified condition, it did not occur even once. My interpretation is that the stability of the PATH scheme depends highly on the salience of affordances for other schemes. Although G&A found a tendency for participants to switch from FIGURAL to PATH, there was a 42% probability for pairs to switch the other way around in my irregular condition. G&A also found a tendency to switch from PATH to LINE. This could explain why PATH did not occur in the stratified condition: The affordance and thus attraction of the LINE scheme was so strong that pairs immediately went for it and “skipped” the PATH stage. In the regular condition, other schemes were not strongly afforded, so PATH became stable.

The question whether MATRIX is a universally attractive strategy and would thus be widely used in general by dyads in these experimental situations requires further validation. It could simply be due to the fact that a matrix conceptualization of the mazes is the most obvious – after all, the mazes are nothing but a series of rooms distributed on a 7×7 grid, which makes the affordance for the MATRIX (i.e. “grid”) scheme strong across all conditions. Another interpretation could be that Western

¹⁹ I.e., there has to be a particular shape affording for a FIGURAL solution, which also has to be recognized as such by a speaker, whereas, e.g., the matrix-scheme can be readily applied to any maze without much perceptual and/or imaginative effort.

education, knowledge about (Cartesian) coordinate systems, geographic maps (constructing the same bird's eye perspective as the maze game) chess boards or simply the frequent use of matrix-like tables with columns and rows in everyday life made the affordance IMPOSE AN X-Y-TYPE GRID ON THE MAZE much more salient to players with this background knowledge.²⁰ After all, experiences from many interactions within a cultural space allow a person to generalize typical patterns of behavior – for example, typical ways in which to act upon objects that occur in this space. For instance, a tablespoon per se does not afford to be of assistance for ingestion without any previous experience of spoons or any cultural concept of CUTLERY, i.e. USING OBJECTS TO TAKE UP PIECES OF FOOD AND RAISE THEM TO ONE'S MOUTH. A person from a population that practices hand-to-mouth eating might perceive completely different affordances from the spoon, e.g., its DECORATIVENESS, by virtue of its shiny metal surface or shape or perhaps its affordance for the action of DIGGING. In return, one could see the affordance of a SPOON in an object that we see primarily as a SHOVEL. To support these speculations, cross-cultural experiments with the present paradigm are desirable to investigate the universality of the findings. Either way, the findings of the present study show how even a strong cultural factor, or general tendency to converge on a specific cultural practice, can be outcompeted by alternatives that are more environmentally adaptive.

It should be noted that this outcompeting was only shown for isolated pairs in the present experiment, since participants only interacted with and attuned to one single person. It is yet unclear how stable the environmental influence would remain if a community condition was added, e.g., by replacing one partner with a new participant from the same pool of participants after a certain amount of time. If the difference between maze-type conditions would collapse in this case, it would suggest that environmental biases would only exert temporary selection pressure on description schemes that are coordinated between dyads, but lose their influence as soon as they have to compete in a community setting, where strong cultural biases might simply mask ecological factors. However, it could also lead to fragmentation, where within one community, different sub-groups differentiate themselves from the rest by stick-

²⁰ Garrod and Clark's study (1993) supports this hypothesis indirectly. They had schoolchildren of different ages play the maze game and found that those from the youngest group (7-8 years) used FIGURAL more than the other three strategies, while children from the oldest group (11-12 years) predominantly used the MATRIX or LINE description scheme. The middle group (9-10 years) lay somewhere in-between, suggesting that more abstract schemes could be influenced by educational factors.

ing to an environmentally adaptive scheme rather than a generally popular one. This could also vary between the two manipulated environments (stratified and irregular). G&D (p. 197) speculate that the reason for the predominance of the MATRIX scheme in their community condition was due to alignment promoting what is initially most commonly used across all participants in the group. That way MATRIX, being the most successful description type, has a very strong chance from the beginning to spread in the community. This suggests the possibility that, if another scheme would be more commonly used in different environmental conditions, these predominant schemes could be preserved and result in communities relying on different schemes in response to their environment. This remains to be tested by placing communities in the three conditions of the experiment. If group conventions were affected by environmental factors in similar vein as the isolated descriptions occurring in the present study (e.g. LINE being most popular in the stratified condition), it would be a strong argument for environmental affordances affecting conventions that groups converge on.

Chapter 4: General discussion

4.1 Taking the interactive alignment model out into the wild

The results of the experiment described in chapter 3 show how a simple manipulation of environmental conditions can influence linguistic descriptions that emerge in task-oriented dialogue. This suggests that linguistic interactions and routines do not depend exclusively on automatic priming mechanisms as suggested in the Interactive Alignment Model (IAM; Pickering & Garrod 2004). There is more to dialogue than two interlocutors – it is a social and cultural endeavor that never takes place in a vacuum, but is always situated in an environment that provides certain affordances. Linguistic adaptations between interlocutors are highly sensitive to these factors of the shared task environment.

Importantly, the environment can influence emerging linguistic conventions independently of the task itself. In my experiment, the task and therefore the communicative goal was exactly the same across conditions. It was also not necessary to use different strategies across conditions to be successful. Every strategy would have worked in every environment. Then why did participants systematically differ in their choice? I suggest that there are two mechanisms at play: 1) Recognition through salience and 2) selection through adaptation.

Recognition through salience means that it is more likely for an agent to perceive an affordance if it appears salient. An example would be the shapes in the irregular condition that afford to be described with a FIGURAL scheme more strongly than the mazes in the regular condition. FIGURAL can be used here, too, but is much less obvious. Selection through adaptation means that if an action is chosen in response to a salient affordance, it will be selected over time based on its adaptiveness (for example, its efficiency) in the task. If we start out with the FIGURAL scheme and solve the first maze, but would then only play regular mazes, it is very likely that we would abandon the scheme again and use another. Conventions are always options chosen from a set of alternatives. Since they are supposed to provide a routinized solution for a recurrent problem, a strategy only becomes a convention when it was indeed repeatedly helpful to achieve a particular goal.

There is an example from cultural psychology illustrating this mechanism: Several studies have suggested that Westerners tend to attend to the environment

analytically and focus on context-independent details, whereas East Asians attend to the environment holistically and engage in context-dependent cognitive processes (see Nisbett *et al.*, 2001 for a review). In a more recent study, Miyamoto, Nisbett and Masuda (2006) showed that these “cognitive styles” could be correlated to culturally characteristic environments. In their experiment, Japanese and Americans were shown scenes from either Japanese or American environments and had to solve a change-blindness task. It turned out that while Japanese did detect a larger number of contextual changes overall, both cultural groups detected a larger number of contextual changes after being primed with Japanese scenes. This suggests that attention can be shaped by the perceptual environment. Since Japanese scenes tended to be more complex (including more objects/buildings) they seem to have afforded for a more holistic rather than analytic perception.

Similarly, in the current study, different ways to perceive the mazes were chosen based on the environment and became expressed as linguistic descriptions that the participants then routinized via alignment. Pickering and Garrod (2004: 172) refer to the priming processes underlying interactive alignment as “primitive” and “resource-free”. This may be true for the environmental priming by perceptual affordances as well. Interlocutors can be understood to automatically “align” to their shared environment, which affords certain situation models (SM) more than others. This, however, means that interactive alignment is not context-free and the framework has to be broadened to include non-linguistic information, as for example Schegloff (2004) pointed out in his response to the original article (although he did not name the environment explicitly).

I propose an “ecological” interactive alignment model (EIAM, figure 18) that takes perceptual affordances into account. Pickering and Garrod (2006: 211) describe how “the specific contents of a person’s situation model [...] depend on the way that he interprets the situation”. If perceptual saliency leads to specific conceptualization of a scene, the SMs of both interlocutors do not only depend on linguistic knowledge (phonological, semantic and syntactic) and general knowledge, but also on recognized affordances. If we play the regular condition, I might not see at all that I could describe figural parts of the maze, but I see the affordance for a MATRIX system. As this becomes a linguistic routine, our SMs become aligned and based on a coordinate system representation of the maze. In one of the other conditions of the experiment, this

would have been less likely, simply because the probability for our SMs to start out with a MATRIX representation would have been lower.

Fig. 18. Schematic representation of the ecological interactive alignment model (based on Pickering & Garrod, 2006: 215). The arrows from “environment” represent pressure exerted by perceptually salient affordances.

Adding this to the IAM is important, because face-to-face communication is always situated in an environment. Garrod & Doherty (1994) rightly identified local salience alongside precedence as a factor influencing which routine is chosen, that is the salience of strategies in a momentary context during the joint activity. In the EIAM, there is also a more global kind of salience: that of the affordances presented by the general environmental context where the joint activity is situated. This general context influences both the salience and precedence of strategies, because it constraints what is initially put into the SM. If an affordance is not recognized, it will not be represented in the SM and subsequently not be expressed in utterances that are directed towards the joint goal. Only strategies based on affordances that are recognized can become

locally salient and have precedence effects, because only they are options to “choose” from²¹.

In a famous paradigm, Tanenhaus and colleagues (1995) showed how language constraints what we attend. When we hear an utterance, we start to pay attention to the most likely referent in our environment from the very beginning. Gaze-patterns related to processing ambiguous sentences show how we rapidly adjust attention even within incomplete words. In other words, language constraints which parts of our environment we attend, or perceive. Likewise, I have shown here how the ecological context itself constrains what we attend and subsequently express in language. It is an interplay of these processes that determines attention, production and comprehension in conversation.

In fact, this appears helpful, since adapting my strategy to the most salient option increases the probability that my partner is using the same. Keysar (2007: 82) speculates that maybe communication is just successful because of a powerful context. Misunderstanding is usually caused by misaligned SMs. But if we use what is most salient in the context for both of us, we reduce the possibility of arriving at different conceptual representation from the beginning. It then makes sense to stick to this adaptive strategy as long as it is successful or until the affordances change and favor other solutions. However, once something becomes a convention, it is not necessary to adjust to local context anymore. This is why conventions make speakers impervious to local context (Garrod & Doherty, 1994: 199) and persist as long as they are not modified or outcompeted.

As mentioned at the end of chapter 3, it would be interesting to test in which way environmental constraints affect community conventions that develop out of the *ad hoc* routines discussed here and to what extent ecological motivations would be preserved.

4.2 Implications for the study of language evolution

The results have some important implications for the study of language evolution as well. While we were looking at task-specific linguistic conventions of an existing language (Danish or English, respectively), languages as conventional signaling systems as such are subject to the same dynamics. Usage-based theories of grammar regard

²¹ This selection process may happen unconsciously, which is why “choose” was put in quotation marks here.

language as a product of experience, based on constructions that arose through repetition across many speech events (e.g., Bybee, 2006). If we see language as a complex adaptive system, where patterns of local use lead to self-organization and selection of global structure (Beckner *et al.*, 2009; Steels, 2011), then it is conceivable that environmental constraints could play a similar role as in joint-problem-solving settings. In section 2.4, I discussed some evidence for acoustic properties and certain grammatical or conceptual choices being shaped by the environment. So far, these studies have been fully theoretical or dependent on correlational evidence. The present experiment is a nice addition, because it tested the hypothesis via manipulation directly in the lab, allowing for causal inferences.

It would therefore be desirable to extend such paradigms to the study of emergent languages (cf. Nölle, 2014). If we consider the case of spatial grammars (cf. section 2.4), they appear quite related to the descriptions in the maze game. Both are essentially conventions of talking about spatial relations. The difference is merely that spatial grammar is used for all kinds of relations (e.g., within a description in the maze game), whereas a specific Danish convention that emerged in the game is only used in this restricted context. But spatial grammar also had to emerge somewhere in the first place and it is highly probable that it was also in a restricted context, where recurring spatial problems were faced that encouraged ready-made solutions. These then resulted in cognitive strategies that became entrenched in grammatical structures. Pickering and Garrod (2006: 225) themselves speculate that alignment in dialogue might also lead to some sort of cognitive alignment, as both interlocutors would focus on the same topics. They propose “a weak ‘Whorfian’ view, where choice of language shapes thought to some extent, and so people who speak like each other are thereby more likely to think like each other”. This is very compatible with the idea that different communities arriving at different situation models, driven by environmental affordances, might over time develop relatively different cognitive routines or reference frames to assess specific situations.

One way to investigate this in the lab could be via “experimental semiotics” (Galantucci & Garrod, 2010). This approach also relies on joint problem-solving experiments where participants have to rely on communication to solve collaborative tasks. The difference to the present paradigm is that they are restricted from using verbal language and instead have to communicate through another medium like ges-

ture or drawing. It can be observed how novel sign systems are bootstrapped and evolve in these settings and possibly which contextual pressures shape their structure. For example, Christensen, Fusaroli and Tylén (2016, in press) show how word order in a gestural communication system is constrained by the event-types participants had to describe. Manipulation events like *someone throwing a paper plane* evoked SOV (subject-object-verb) descriptions, whereas manipulation events (*someone painting a paper plane*) evoked more SVO descriptions. Furthermore, this iconic tendency was modulated by alignment and relative frequency of event types. The experiment thus showed how external social and environmental factors can modify emergent sign systems, similar to what correlational studies concerning languages claim.²²

Based on the insights of the present study, one could think of a similar paradigm, where participants have to communicate spatial relations to solve a collaborative task. Again, environmental affordances could be manipulated in different conditions to test whether novel communication systems emerging in these conditions differ systematically in their underlying spatial strategy. This could inform research in linguistic relativity and support the idea that the environment is one of the pressures that led to the diversity we observe among the world's languages in this domain. One could further test how such different systems spread or adapt in communities put into different environments and see how social factors interplay with environmental constraints.

Theories of enactivism, embodied and situated cognition or ecological psychology have stressed that our cognition stems from interacting with an environment in which our bodies are situated (Varela, Thompson & Rosch, 1991; Hutchins, 1995; Gibson, 1979). It is time to include such thinking in studies about the origin of the world's languages as well, because language is just as embedded in physical activities and socio-cultural contexts as any other practice or cognitive skill. Experiments trying to disentangle the different mechanisms that work back and forth between these factors could be a first step and complement findings relying on statistical corpus data or agent-based modelling approaches.

²² See section 2.4 for studies on correlations of environment and linguistic features. For correlations with social structure see, e.g., Lupyan & Dale (2010).

Chapter 5: Conclusion

In this thesis, I tried to unravel a driving factor shaping human cultural evolution. Particularly, I focused on linguistic routines, a common type of convention that emerges and changes spontaneously while it is used in social interactions. Previous accounts (e.g., Clark, 1996) have rightly identified conversations as joint activities where language is used to achieve mutual goals. It was discussed how conventions come about automatically and without any central control in dialogue, because interlocutors mechanistically align and routinize linguistic structures, which then spread in the community (Pickering & Garrod, 2004; Garrod & Doherty, 1994). I hypothesized that these conventions are not just arbitrary, but rather motivated in that they provide an ecological adaptation. The joint problem-solving experiment described in chapter 3 was designed to test this idea. Participant dyads had to solve the same collaborative game in different environments. The results showed a clear effect of environmental conditions on the linguistic conventions dyads developed. They aligned their linguistic location descriptions needed to solve the task in response to their perceptual environment.

Based on this finding I proposed to add external factors to Pickering and Garrod's (2004) Interactive Alignment Model. Since every conversation is situated in a specific environment, there is always some kind of constraint in terms of perceptual affordances that can drive interlocutors to choose one strategy over another and routinize different behavior than they would in another setting, even if the conversation there would have the exact same goal (solving a specific coordinative problem). If linguistic routines emerging in dialogue depend on more than automatic interlocutor priming mechanisms, the same is likely true for more general linguistic features that are subject to cultural-evolutionary processes. The findings of the present experiment provide a good starting point to design paradigms that can test this, e.g., in an experimental semiotics framework. Future research in this direction might help explain the extraordinary diversity among the world's languages that we can observe today by uncovering some of the mechanisms contributing to it.

References

- Aijmer K. (1996) *Conversational Routines in English: Convention and Creativity*. London: Longman.
- Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. New York: Simon and Schuster.
- Austin, J. L. (1975). *How to do things with words*. Oxford: Oxford University Press.
- Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, 329, 1081–1085.
- Barr, D. J. & Keysar, B. (2004) Is language processing different in dialogue? Response to Pickering & Garrod. *Behavioral and Brain Sciences*, 27(02), 190–191.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D. & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language learning*, 59(s1), 1–26.
- Behnel, M., Cummins, C., Sichelschmidt, L., and de Ruiter, J. (2013). “Priming and conceptual pacts in overhearers’ adoption of referring expressions,” in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, (Austin, TX: Cognitive Science Society), 1869–1874.
- Bernstein, N. A. (1967). *Coordination and regulation of movement*. New York: Pergamon Press.
- Beuls, K., & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLOS ONE*, 8(3), e58960.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13-B25.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482.

- 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.
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82, 711-733.
- Carston, R. 1999. Review of Using Language by H. H. Clark. *Journal of Linguistics*, 35(1), 167-171.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of personality and social psychology*, 76(6), 893.
- Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, 3(1), 1-15.
- Christensen, P., Fusaroli, R., & Tylén, K. (2016, in press). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, 146, 67-80.
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, 31(05), 489-509.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1-39.
- Cleland, S., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49, 214-230.
- Cohen, P. R. (1984). The pragmatics of referring and the modality of communication. *Computational linguistics*, 10(2), 97-146.
- Croft, W. (2010). Relativity, linguistic variation and language universals. *CogniTextes. Revue de l'Association française de linguistique cognitive*, (Volume 4).
- Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The self-organization of human interaction. *Psychology of Learning and Motivation* 59.
- Davies, I. R. L., Sosenskaja, T. & Corbett, G. G. (1999). Colours in Tsakhur: First account of the basic colour terms of a Nakh-Daghestanian language. *Linguistic Typology*, 3, 179-207.

- de Saussure, F. (1983). *Course in General Linguistics*. Edited by Charles Bally and Albert Sechehaye. Translated by Roy Harris. La Salle, IL: Open Court.
- Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladdottir, R.S., Kendrick, K.H., Levinson, S.C., Manrique, E., Rossi, G. & Enfield, N.J. (2015). Universal principles in the repair of communication problems. *PLOS ONE*, 10(9), e0136100.
- Duncker, Karl (1945). *On Problem Solving*. (Psychological Monographs, 58). American Psychological Association,
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye-movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436.
- Epley, N., Keysar, B., VanBoven, L., & Gilovich, T. (2004). Perspective Taking as Egocentric Anchoring and Adjustment. *Journal of Personality and Social Psychology*, 87, 327–339.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and brain sciences*, 32(05), 429–448.
- Everett, C. (2013). *Linguistic Relativity. Evidence Across Languages and Cognitive Domains*. Berlin, Boston: De Gruyter Mouton.
- Everett, C., Blasi, D. E., & Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences*, 201417413.
- Everett, D., 2005. Cultural Constraints on Grammar and Cognition in Pirahã. *Current Anthropology*, 46(4), pp.621–646.
- Fedden, S., & Boroditsky, L. (2012). Spatialization of time in Mian. *Frontiers in psychology*, 3.
- Ferreira, F. Ferraro, V. & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11–15.
- Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: John Wiley.
- Foltz, A., Gaspers, J., Thiele, K., Stenneken, P., & Cimiano, P. (2015). Lexical alignment in triadic communication. *Frontiers in psychology*, 6.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531–534.

- Fusaroli, R. & Tylén, K. (2012) Carving language for social coordination, *Interaction studies*, 13(1), 103–124.
- Fusaroli, R., & Tylén, K. (2015). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive Science*, 1–27.
- Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C. and Tylén, K. (2012) Coming to Terms: Quantifying the benefits of linguistic coordination, *Psychological Science*, 23(8), 931–939.
- Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147–157.
- Galantucci, B., & Garrod, S. (2010). Experimental semiotics: A new approach for studying the emergence and the evolution of human communication. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, 11(1), 1–13.
- Galantucci, B., & Roberts, G. (2014). Do we notice when communication goes awry? an investigation of people's sensitivity to coherence in spontaneous conversation. *PLOS ONE*, 9(7), e103182.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2), 181–218.
- Garrod, S., & Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes*, 8(1), 101–126.
- Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3), 181–215.
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292–304.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Ginzburg, J. (2001). Fragmenting meaning: Clarification ellipsis and nominal anaphora. In Bunt, H., Muskens, R. & Thijsse, E. (Eds.) *Computing meaning*. Dordrecht: Springer.
- Goodwin, C. 2000: Action and embodiment within human situated interaction. *Journal of Pragmatics*, 32, 1489–1522.

- Grice, H. P. (1991). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, 15(6), 409–414.
- Haspelmath, M. (2007). Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology*, 11(1), 119–132.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve?. *Science*, 298(5598), 1569–1579.
- Healey, P. G., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, 31(2), 285–309.
- Hughes, T. P., Baird, A. H., Bellwood, D. R., Card, M., Connolly, S. R., Folke, C., ... Roughgarden, J. (2003). Climate Change, Human Impacts, and the Resilience of Coral Reefs. *Science*, 301(5635), 929–933.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Jacob, C., McDaniels, T., & Hinch, S. (2010). Indigenous culture and adaptation to climate change: sockeye salmon and the St'át'imc people. *Mitigation and Adaptation Strategies for Global Change*, 15(8), 859–876.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (Cognitive Science Series, No. 6). Cambridge, MA: Harvard University Press.
- Kelso, J. S. (1997). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT press.
- Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, 4, 71–84.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32–38.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686.

- Kirby, Simon, Tom Griffiths & Kenny Smith. 2014. Iterated Learning and the Evolution of Language. *Current Opinion in Neurobiology*, 28. 108–114.
- Kuiper K. (1996) *Smooth Talkers: The Linguistic Performance of Auctioneers and Sportscasters*. Mahwah, NJ: Erlbaum.
- Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". *Biometrics* 33 (1): 159–174.
- Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In Bloom, P., Peterson, M., Nadel, L., Garrett, M. (eds.) *Language and space*, 109–169.
- Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity* (Vol. 5). Cambridge: Cambridge University Press.
- Levinson, S. C., & Wilkins, D. P. (eds.) (2006). *Grammars of space: Explorations in cognitive diversity* (No. 6). Cambridge University Press.
- Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
- Lewis, D.K. (1975). Language and languages. In Gunderson, K. (Ed.), *Language mind and knowledge*. (Minnesota Studies in the Philosophy of Science, vol. 7). Minneapolis: University of Minnesota Press.
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46, 551–556.
- Lindsey, D. T., & Brown, A. M. (2002). Color naming and the phototoxic effects of sunlight on the eye. *Psychological Science*, 13(6), 506–512.
- Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, 36(8), 1404–1426.
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLOS ONE*, 5(1), e8559.
- Maddieson, I. & Coupé, C. (2015). Human spoken language diversity and the acoustic adaptation hypothesis. *Paper presented at the Acoustical Society of America (ASA) Fall 2015 Meeting*, Jacksonville, FL.

- Majid, A., Bowerman, M., Kita, S., Haun, D., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in cognitive sciences*, 8(3), 108–114.
- Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522–523.
- Mills, G. J. & Healey, P. G. T. (submitted). A dialogue experimentation toolkit. URL: homepages.inf.ed.ac.uk/gmills/MillsHealey2013Submission.pdf [accessed October 2015].
- Mills, G. J., & Gregoromichelaki, E. (2010). Establishing coherence in dialogue: sequentiality, intentions and negotiation. *Proceedings of the 14th SemDial, Pozdial*. Poznan: Adam Mickiewicz University, 17–24.
- Miyamoto, Y., Nisbett, R. E., & Masuda, T. (2006). Culture and the physical environment. Holistic versus analytic perceptual affordances. *Psychological Science*, 17(2), 113–119.
- Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., ... & Laland, K. N. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature communications*, 6.
- Nilsenová, M., & Nolting, P. (2010). Linguistic adaptation in semi-natural dialogues: age comparison. In: Sojka, P., Horák, A. Kopeček, I. & Pala, K. (eds.) *Text, speech and dialogue. Proceedings of the 13th International Conference, TSD 2010, Brno, Czech Republic*. Berlin/Heidelberg: Springer, 531–538.
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108, 291–310.
- Nölle, J. (2014). A co-evolved continuum of language, culture and cognition. Prospects of interdisciplinary research. *Studies about Languages*, 25, 5–13.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(02), 169–190.
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4(2-3), 203–228.

- Pilot-Raichoor, C. (2010). The Dravidian zero negative: diachronic context of its morphogenesis and conceptualisation. *Rara and rarissima. Documenting the fringes of linguistic diversity*. Berlin: de Gruyter, 267–303.
- Pirie, M. (2010). *The Role of Social Dominance in Lexical Alignment*. M.Sc. dissertation, Edinburgh: University of Edinburgh.
- Premack, D., & Woodruff, G. (1978) Does the chimpanzee have a ‘theory of mind’? *Behavioral and Brain Sciences*, 4, 515–526
- Puglisi, A., Baronchelli, A., & Loreto, V. (2008). Cultural route to the emergence of linguistic categories. *Proceedings of the National Academy of Sciences*, 105(23), 7936–7940.
- Reitter, D., Moore, J. D., & Keller, F. (2010). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. *Proceedings of the 28th annual conference of the Cognitive Science Society*, 685–690, URL: <http://csjarchive.cogsci.rpi.edu/Proceedings/2006/docs/p685.pdf> [accessed October 2015].
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26, 867–891.
- Richerson, P. J., & Boyd, R. (2008). *Not by genes alone: How culture transformed human evolution*. University of Chicago Press.
- Schegloff, E. A. (2004). Putting the interaction back into dialogue. Response to Pickering & Garrod. *Behavioral and Brain Sciences*, 27(02), 207–208
- Schegloff, E.A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361–382.
- Schmidt, R. C., Richardson, M. J., Arsenault, C., & Galantucci, B. (2007). Visual tracking and entrainment to an environmental rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 860–870.
- Searle, J.R. (1969). *Speech acts*. Cambridge: Cambridge University Press.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–76.
- Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: a framework for the emergence of language. *Artificial life*, 9(4), 371–386.

- Solomon, S., Plattner, G.-K., Knutti, R., & Friedlingstein, P. (2009). Irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Sciences*, 106(6), 1704–1709.
- Spedicato, G. A., Kang, T. S. & Yalamanchi, S.B. (2015). *The markovchain Package: A Package for Easily Handling Discrete Markov Chains in R*. Introduction vignette. URL: https://cran.r-project.org/web/packages/markovchain/vignettes/an_introduction_to_markovchain_package.pdf [accessed, October 2015].
- Spedicato, G.A. & Kang, T.S. (2015). *markovchain: discrete time Markov chains made easy*. R package version 0.3.1, URL: <https://cran.rstudio.com/web/packages/markovchain/index.html> [accessed, October 2015].
- Sperber, D. & Wilson, D. (1986) *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Steels, L. (2011). Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4), 339–356.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT press.
- Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.
- Tylén, K., Fusaroli, R., Bundgaard, P. F., & Østergaard, S. (2013). Making sense together: A dynamical account of linguistic meaning-making. *Semiotica*, 2013(194), 39–62.
- Tylén, K., Weed, E., Wallentin, M., Roepstorff, A., & Frith, C. D. (2010). Language as a tool for interacting minds. *Mind & Language*, 25(1), 3–29.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: cognitive science and human experience*. Cambridge, MA: MIT Press.
- von Uexküll, J. (1921). *Umwelt und Innenwelt der Tiere*. Berlin/Heidelberg: Springer.
- WALS: Haspelmath, M., Dryer, M. S., Gil, D. & Comrie, B.(2005). *The World Atlas of Language Structures*. Oxford: Oxford University Press. URL: <http://WALS.info/> [accessed October 2015].

- Whiten, A., Hinde, R. A., Laland, K. N., & Stringer, C. B. (2011). Culture evolves. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567), 938–948.
- Wimsatt, W. C. (2013). Articulating Babel: An approach to cultural evolution. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4), 563–571.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19), 7780–7785.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.
- Wu, S., & Keysar, B. (2007). The effect of information overlap on communication effectiveness. *Cognitive Science*, 31(1), 169–181.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2), 162–185.

Cover Image photos*

https://commons.wikimedia.org/wiki/File:Jubilee_Maze,_Symonds_Yat.jpg

<https://www.flickr.com/photos/artinstituteofportland/3275557790>

<https://www.flickr.com/photos/randomwire/6055829411>

*All released under Creative Commons license, all rights reserved.

Appendix I – Mazes

Condition 1: Regular mazes

Condition 2: Irregular Mazes

Condition 3: Stratified Mazes

Appendix II – Data

Summary of descriptions for all pairs/conditions

Condition	Pair	PNo	DD	DDL	PATH	MATRIX	FIGURAL	LINE	UNDEFINED
regular	1	01&02	1	1	33	0	0	0	4
regular	2	03&04	1	1	25	0	1	1	3
regular	3	05&06	2	2	0	32	1	9	0
regular	4	07&08	4	4	3	0	0	28	3
stratified	5	09&10	4	4	0	0	2	48	3
stratified	6	13&14	4	4	0	0	1	27	3
stratified	7	15&16	4	4	0	0	1	31	1
irregular	8	17&18	2	2	1	20	6	2	0
irregular	9	19&20	3	3	4	0	28	2	4
irregular	10	23&24	3	1	6	0	24	7	2
irregular	11	25&26	3	2	5	5	41	18	2
irregular	12	27&28	4	4	0	0	2	24	0
irregular	13	29&30	4	4	5	0	12	12	0
irregular	14	31&32	3	N/A	4	0	18	4	0
stratified	15	33&34	4	4	0	0	1	28	0
stratified	16	35&36	4	2	0	16	1	24	0
stratified	17	37&38	4	4	0	0	1	55	0
stratified	18	39&40	2	N/A	0	21	3	12	3
stratified	19	41&42	2	2	0	34	0	4	0
regular	20	43&44	1	1	21	0	1	2	0
regular	21	45&46	1	1	36	0	1	0	0
regular	22	47&48	2	2	16	18	0	0	0
regular	23	49&50	1	2	24	10	6	0	4
regular	24	51&52	4	4	3	0	6	63	3
regular	25	53&54	4	4	0	0	1	28	0
stratified	26	55&56	4	N/A	0	0	2	28	0
irregular	27	57&58	2	2	5	29	1	4	1
irregular	28	59&60	3	3	6	0	33	4	1
stratified	29	61&62	4	4	0	0	1	39	0
stratified	30	63&64	4	N/A	0	0	0	35	2
regular	31	67&68	2	2	0	26	0	0	0
regular	32	69&70	2	2	5	21	3	0	0
irregular	33	71&72	3	4	8	0	26	16	0

Tab. 5. Description counts. PNo = Participant Numbers, DD = dominant description scheme, DDL = dominant description scheme for the last maze (N/A means that there was a tie between two schemes).

All coded chat protocols are available under the following link:

<https://onedrive.live.com/redir?resid=BB568FC9EB3CD9E0!44463&authkey=!AGDFZulkOfaBmNU&ithint=file%2czip>

Appendix III : Instructions (Danish and English)

Labyrint-Eksperimentet

Instruktioner

Opgaven går ud på at samarbejde med en partner om at komme igennem en serie af labyrinter.

Labyrinterne består af rum, der er forbundet med hinanden. Hver spiller har forskellige mål angivet med røde kryds (X). Labyrinten er løst, når begge spillere er nået deres mål.

Vanskeligheden ligger i det faktum, at ikke alle forbindelser kan passeres. Nogle er blokeret af **døre**. Disse døre kan kun åbnes, hvis en spiller går hen på et rum, hvor hans / hendes partner har en **switch** (markeret med grå). Dvs. hvis spiller A går hen på et rum, hvor B har en switch og bliver der i min. **2.5 sek.**, så åbner B's døre, og forbliver åbne så længe spiller A forbliver i dette rum.

Bemærk, at mens begge spillere har samme labyrint-layout, ligger deres switches og mål på forskellige positioner.

Der er to vinduer: Chat-vinduet og labyrint-vinduet. I chatvinduet kan du kommunikere med din partner. I labyrinten vinduet vises din position som en sort prik. Du flytter dig rundt i labyrinten ved at trykke **Ctrl + Piletaster** (du kan holde Ctrl nede og bevæge dig selv om du er i chatvinduet).

Snakker ikke under eksperimentet!
Chatten er alt du behøver.

Labyrinth Eksperiment

Instructions for participants

The idea behind the game is to help each other reach the finish in a series of mazes.

The mazes consist of rooms that are connected to each other. The finish is indicated by a red cross (X). The maze is solved as soon as **both** players reach the finish room in their maze.

The difficulty lies in the fact that not all connections can be passed. Some are blocked by closed **gates**. These gates can be opened if a player goes into a room where his/her partner has a **switch** (marked grey). E.g., if player A goes into a room, where B has a grey switch, then after **2.5 seconds** B's gates open, as long as player A stays in that room.

Note that while both players have the same maze-layout, their switches and finishes are located at different positioner.

There are two windows: The chat window and the maze window. The chat window allows you to communicate with your partner, the maze window shows your position marker (black dot). You can move around the maze by pressing **Ctrl+Arrow keys** (you can hold Ctrl and move even when you are in the chat window).

Please, **do not talk** during the experiment. The chat is all you need.

Appendix IV : Coding Scheme

To code:

Code only such turns (i.e. lines in the chat protocol) that contain **descriptions of position** such as concerning the location of a player's location marker, the finish or a switch/grey field.

What counts as description?

New descriptions i.e., those following several utterances on a different topic, even if incomplete.

Restated rejoinders i.e., restatement by the addressee but in a different format from the original.

Complete restatements by original speaker where different from the original (e.g., includes path descriptions with different origin). (Adapted from Garrod & Anderson, 1987)

What does not count as a description?

Other utterances, including procedural clarifications, but also **repetitions** of all or part of an original description, e.g. in clarification requests ("hvad mener du med 2. sidste række?") or in a repair where only the first turn counts, as in the exchange below (LINE):

- (28) Switch i 3. øverste række nr. 3 fra venstre
- (29) 3 øverst nr 3 fra venstre?
- (30) Nr. 3 fra højre

If a line contains more than one type of description, both get a code in their column.

Sometimes one description might be separated over two or more turns in the chat protocol, in that case only the first part of the description (first line) is coded. This is an example of a PATH description distributed over 3 lines, where the first line would get a PATH code.

- (31) ok. Mit felt er placeret næsten midt i
- (32) hvis du er i venstre hjørne og går tre ned
- (33) og så tre til højre

Descriptions that anaphorically refer to positions (e.g., "I am now at the finish", "I am standing at the switch now") are not coded, as they rely on none of the underlying strategies. Vague descriptions are coded as UNDEFINED (see below).

Description types (including some examples of single turns for each type)

PATH description:

Descriptions where first a prominent point on the maze is identified (e.g., a corner) and then a path which follows the links and boxes ending at the location is described.

- (34) jeg har to. det ene er øverst højre hjørne, 3 til venstre, 1 ned og 1 til højre
- (35) min er øverst venstre, 1 højre, 1, ned, 1 højre
- (36) højre hjørne oppe. 1 ned

To differentiate them from spelled out MATRIX coordinates (see below), descriptions that include two coordinates are counted as PATH if they include directions (e.g., *from*, *towards*, *into*, *inside* etc.):

- (37) helt nederst og tre ind fra venstre side
- (38) 2. nederst fra venstre og en ind

LINE descriptions:

Descriptions, where first a line, i.e. a number of subsequently connected rooms is identified with a word similar to *line/row/column/level* or *ground/top* (for the lowest/highest row) and then a location on this line is specified (i.e. a location relative to the line). Lines can be horizontal as in the first example or vertical as in the second.

- (39) gråt felt, 3. nederste række, yderst til venstre
- (40) Min er i bunden 3 til højre
- (41) Jeg har et gråt felt i den næstøverste række. felt 5 fra venstre
- (42) okay mit felt er i anden række fra venstre aller øverst

Sometimes pairs start with a clear cut LINE system, but tend to omit terms like *row* in single turns. Very clear cases like that should still be coded as LINE, not so clear cases should be coded as UNDEFINED. See for example, where clearly the participant is talking about horizontal arrangements that can be described as "second lowest", but omits the word *row* (*række*) that this pair otherwise used.

- (43) Andennederst og nr. 2 fra venstre

This means that as long as an adjective like *highest/second lowest* etc. is present to refer to a line, along with a position on that line, e.g., *left/right* (referring to an end of said line), it still counts as a line description:

- (44) øverst venstre

Note that descriptions don't count as line, if rows are mentioned together with columns to describe an intersection. In that case it counts as a MATRIX strategy.

- (45) Prøv du første kolonne sidste række

MATRIX descriptions:

Descriptions imposing a structure of vertical and horizontal lines on the maze and then indicating a position as being at the intersection of any of those lines. Such descriptions essentially correspond to a coordinate in a grid-like coordinate system, where empty fields (i.e. left out rooms in the 7x7 grid) do not matter:

- (46) Hvis det er et 6x6 grid hvor man tæller vandret, lodret - (5, 1)
- (47) jeg kan ikke komme i mål, har du været på e1 og f5?
- (48) Der er jeg nu. Kan ikke komme til 2,3

The coordinates can also be spelled out linguistically:

- (49) min er i 2. kollone fra venstre og anden række fra toppen

FIGURAL descriptions:

Descriptions that include particular configuration or pattern of boxes on the maze and then indicate a position relative to this pattern. It can be abstract descriptions, e.g., of arrays of boxes that resemble a square, L etc., or conceptualizing the maze as an object/entity that is segmented and then referring to a specific salient segment:

- (50) Ja altså vandret er det kun tre kasser i modsætning til de andre hvor der er fem på tværs
- (51) nederat til højre i miderste arm
- (52) Nederst til højre i firkanten i venstre hjørne
- (53) til "broen" i midten af den store firkant.

UNDEFINED descriptions:

Descriptions that are so vague that they cannot be subsumed under any of the 4 description types above – usually, because they are incomplete or vague. Also vague positions only relative to a finish or switch are considered undefined (like in the last two examples below):

- (54) det er ude i højre side men jeg kan ikke komme til det grå felt endnu, der er en dør for
- (55) Jeg er ca i midten
- (56) Mit mål er ligesom før <.<
- (57) lige mellem de 3 sidste firkant, hvor man går i mål.-
- (58) min port er lige før målet