Universität Bremen*

Linguistic Representation of Problem Solving Processes in Unaided Object Assembly

Dissertation zur Erlangung der Doktorwürde
durch den Promotionsausschuss Dr. phil.
der Universität Bremen

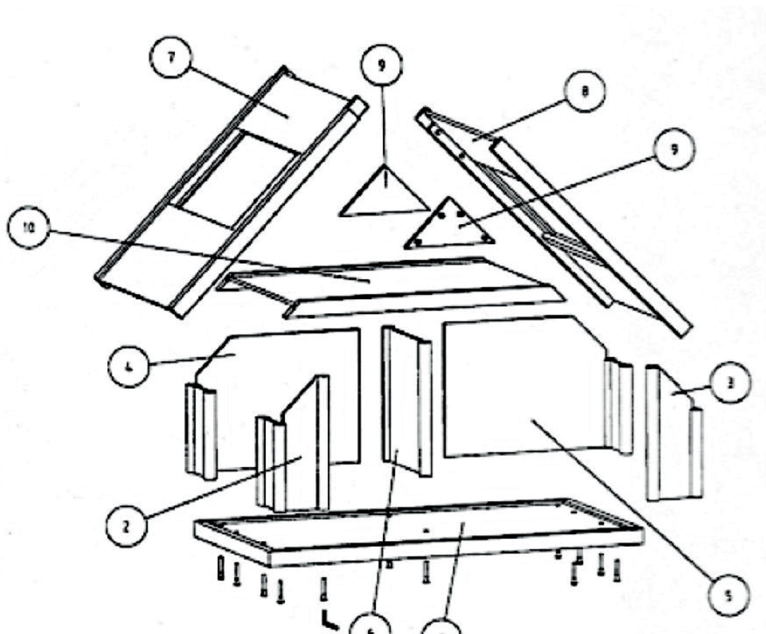
vorgelegt von LINN GRALLA

Hannover, den 31. Juli 2013

Linguistic Representation of
Problem Solving Processes
in Unaided Object Assembly

1. Gutachter: Prof. John Bateman (Universität Bremen)
2. Gutachter: Prof. Ute Schmid (Universität Bamberg)

Verteidigt im Kolloquium am 10.März 2014 an der Universität Bremen mit dem Prädikat 'magna cum laude'.


7
9
8
9
10
4
3
2
6
1
5

Abstract

Cognitive linguists believe that “language reflects patterns of thought” (Evans & Green 2006:5). Based on this framework, this thesis investigates problem solving processes in unaided object assembly and their linguistic representation. This work extends traditional approaches to problem solving that focused on logic-based problems by investigating a real-life problem, namely assembling a two-story dollhouse. Assembly problems are frequently encountered in real-life as indicated by the vast research on manual design and results in a study on assembly experiences (Richardson 2007). In order to ensure that participants need to conceptualize objects and define their possible locations, the assembly was not facilitated by a manual in the experimental studies. Moreover, participants were provided with different amounts of information about the nature of the goal object. Some participants did not know anything about the nature of the object. Some participants were instructed to assemble a two-story dollhouse and some were additionally shown a picture of the goal structure.

Methodologically, problem solving processes were studied by eliciting verbal protocols and performance measures such as time and success. Verbal data comprised protocols of concurrent verbalization during task performance, retrospective reports elicited immediately after the task, verbal instructions, and directed questions. The qualitative analysis of this data combined the well-established methodology of content-based *protocol analysis* with in-depth linguistic analysis in the framework of *Cognitive Discourse Analysis*. The linguistic analysis focused on highlighting the verbalization of experience in the verb phrase, the conceptualization of objects in the noun phrase, and recurring patterns of adjectives, adverbs, and further discourse markers. Based on the content-based description of problem solving processes, the structural patterns of these processes were analyzed. The qualitative analysis was combined by quantitative analyses in terms of descriptive and inferential statistics.

The content-based analysis revealed that participants frequently formulated hypotheses concerning the function of objects, their arrangement and location, or possible manipulations. These hypotheses were most often followed by verbalizations of physical manipulation of objects, such as selecting objects and placing them. The new state was frequently evaluated as confirming to prior expectations or diverging from the expected results. This sequence corresponds to the ‘search and test’ procedure that is generally assumed in the literature on problem solving (e.g. Newell & Simon 1972). Moreover, the analysis revealed that participants described their mental state concerning emotions that are raised or when turning to themselves for answers. In the present context, participants commented on object features; a process that was frequently followed by new hypotheses. This aspect of hypothesis generation was analyzed more closely. The analysis of guidance by object features revealed that participants used specific object features strategically to facilitate their assembly. Mention of these constraints was different between kinds of verbal reports. Participants verbalized different constraints during the assembly than when they were asked to recall their thoughts after the assembly. Interestingly, the comparison between constraints mentioned in those reports with constraints that were verbalized in instructions revealed that the same constraints were used but they served different functions. These insights are specifically interesting because no systematic investigation of constraints was reported prior to this thesis. Therefore, the presented analysis might inspire future research.

The linguistic analysis of the identified problem solving process categories highlighted that those could be distinguished based on the verb type that was predominantly verbalized. Hypotheses, for example, were frequently marked by verbs of 'being and having' (Halliday 1985), such as 'sein' (be). Furthermore, the analysis of verb tense indicated that participants concentrated on the present situation and did not engage in elaborate mental planning activities. The analysis of different aspects of the noun phrase revealed that participants conceptualized half of all provided objects within the goal domain, i.e. 'house', whereas all remaining references were on a general descriptive level, such as 'thing'. The frequency analysis highlighted significant effects of prior information on referential form. First, participants who were provided with an external representation of the goal structure, i.e. a picture, referred to objects by pronouns more frequently than participants without this information. Second, participants who were not given any specific information about the goal structure tended to use structure-based terminology, such as 'story', when explicitly assigning function to objects.

The presented work contributes to research in the field of problem solving by revealing that real-world problem solving involves the same mental processes as described for logic-based problems. However, the findings suggest that their structural patterns might be less well-organized than previously assumed. Methodologically, this thesis applies and refines a methodology that combines content-based protocol analysis with the analysis of systematic recurring linguistic patterns, namely *Cognitive Discourse Analysis*. The results highlight that problem solving processes can be distinguished based on verb type and discourse markers. For researchers who want to adopt this approach in their work, this thesis provides a comprehensive overview on previous findings and a theoretical framework for the analysis of German verb and noun phrases. The theoretical considerations on which the analysis of these aspects is based discuss how findings obtained in experiments in which speech is produced for an addressee can be adopted for the analysis of speech produced for oneself. These theoretical contributions extend previous publications that applied *Cognitive Discourse Analysis*.

Acknowledgments

I want to extend my warmest thanks to my supervisors Dr. Thora Tenbrink and Prof. Dr. John Bateman for their valuable and constructive comments on drafts of this thesis and their support along the long way that this thesis has come.

I also wish to thank my colleagues and friends, specifically Dr. Elena Andonova, Dr. Joana Hois, Sven Brüssow, and Prof. Dr. Alfons Maes, for their encouragement throughout this project and our discussions of new directions and thoughts. I want to specifically thank Evelyn Bergmann and Carsten Gondorf for their patience of proofreading this thesis and commenting on it. I am also grateful to comments and suggestions by anonymous reviewers of contributions to national and international conferences and the inspiring discussions after these talks throughout the years.

This project would not have been possible without the students who dedicated their time and energy to fulfill the tasks that were posed in the experimental sessions. Therefore, thank you to all of them. Furthermore, I would like to thank Katja Abeln who supported me in the collection and transcription of the data at the beginning of this project.

I am grateful to the Zentrale Forschungsförderung Universität Bremen for granting me a stipend for eight months to wrap everything up and write this final version of the thesis. Moreover, I wish to thank them for financial support for the attendance of the annual meeting of the Cognitive Science Society in Sapporo, Japan, in 2012. I wish to thank those organizations that sponsored my trips to national and international conferences by granting me travel stipends, namely the Fachbereich 10 at the University of Bremen, the Gesellschaft für Kognitionswissenschaft, the Cognitive Science Society, and the SFB TR8 Spatial Cognition (Deutsche Forschungsgemeinschaft DFG), I6-[NavTalk].

Least but foremost I want to thank those who supported me along the way by continuously encouraging me to keep going and made this possible by giving me the time to completely dedicate myself to this thesis. Since I cannot mention all family members and friends that supported me, I wish to thank Hendrik Bindewald, my parents, and Malina.

Table of content

1 Introduction

This thesis investigates problem solving processes in unaided object assembly and their linguistic representation. Everyone who has tried to assemble new furniture by herself/himself has probably experienced numerous moments of puzzlement about provided objects, new ideas regarding their function, and the need to disassemble parts to rearrange them. These mental and physical activities are observable sequences of problem solving. Generally, problem solving is conceived of as a transformation process from the currently unsatisfactory state to the desired state either by well-known procedures or by the creation of new procedures. In cases in which people need to devise novel sequences of actions, either mentally or physically, they are involved in a problem solving process. As outlined above, object assembly potentially involves numerous instances in which new procedures need to be devised, such as how objects are connected and fastened as well as by which means.

The aim to investigate the linguistic representation of problem solving processes in unaided object assembly involves two main research issues. First, the hypothesis that unaided object assembly results in problem solving activity needs to be tested by identifying processes of problem solving in the recorded data. Second, the linguistic features, such as verb phrase, noun phrases, and structuring devices, of the identified processes are studied systematically. By studying the verb phrase information can be gained about the type of activity that is performed, i.e. mental vs. physical activities (e.g. Halliday 1985), as well as about the participants' certainty about these actions. Whereas mental activities are expressed by verbs, such as 'think', 'assume', 'believe', physical activities are accompanied by verbs, such as 'take', 'put', 'fasten'. The speaker may take a personal stance on the likelihood of a verbalized action by different grammatical aspects of modality, i.e. modal verbs, modal adverbs, or mood (e.g. Erben 1972). The investigation of noun phrases, on the other hand, allows to study the speaker's conceptualization of objects and object arrangements (e.g. Rosch 1976). Noun phrases can be studied with regard to nominal vs. deictic references and nominal specificity, i.e. if an object is referred to in general or in domain specific terms (e.g. von Stutterheim et al. 1993). Furthermore, the analysis of structuring devices highlights the relations that the speaker assumes between information, such as temporal order, connectedness, or consequential relations (e.g. Halliday 1985; Schifrin 1987).

This thesis explores a number of new directions, i.e. with regard to the task as well as concerning the combination of traditional content-based analysis with in-depth linguistic analysis, thus traditional frameworks need to be adopted and extended. Since this thesis presents empirical work, the refinement of the analysis procedure needs much effort concerning definitions of coding categories, the coding process itself, and decisions about appropriate quantitative measures. The next section introduces the research areas of cognitive linguistics and problem solving as well as general methodological motivations. The empirical studies that are central to this thesis will be briefly introduced in the third section before the structure of the thesis is presented.

1.1 Motivation and research issues

Certainly everyone would agree that he/she knows what language is and how to describe thoughts. However, defining each of these concepts precisely is quite difficult. Starting out by defining the concept of thought, the Oxford English dictionary states that 'thought' is "a person's mind and all the ideas that they have in it when they are thinking" (OED 2000). Central to this definition is the concept of thinking which is defined as the process in which mentally stored information is newly structured and combined (Beller & Bender 2010:15). Three main functions of thinking can be identified: representation, categorization, and combination of information (Brockhaus Enzyklopädie). It is assumed that information is abstracted, generalized, classified, and coordinated in the human mind (Beyer & Gerlach 2011).

Language, on the other hand, is defined as a medium to convey information by words and their respective combinations (Beller & Bender 2010:16) being conceived of as a communication system (Reither 1996:207). Language exhibits the following three characteristics; it is generative since different words can be combined to built an almost infinite number of sentences.[1] It is systematic because it has certain underlying rules of word and sentence formation and it is bound to the situation in which it is uttered (Beller & Bender 2010:16). As choices in language are influenced by the situation and communication is an act of social interaction, language is frequently shaped by communicative intention. Halliday (1985) captures these different functions of language in the distinction between ideational, interpersonal, and textual metafunction of language. The *ideational metafunction* of language describes the process of construing experience, i.e. the classification of objects and people around us to organize them mentally. The *interpersonal metafunction* encompasses the social function of language, i.e. building and maintaining personal relationships with other people. The ability to create fluent, discursively adequate, and coherent speech is subsumed in the *textual metafunction* of language.

The basic question of how these two basic elements of human cognition interact, i.e. the internal process of thinking and the primarily externally directed system of communication, has been debated about for hundreds of years. This question was approached and answered differently by various philosophers and theorists. Based on theoretical considerations as well as on experimental studies and naturalistic observations, many theorists argue that thought and speech are closely connected (e.g. Kleist 1805/2002; Vygotsky 1934; Whorf 1956; Piaget 1959; Dörner 2006). The question of how thought is represented in the human mind is also central to investigations in the field of *cognitive science*. Cognitive scientists are interested in the study of "higher mental processes", namely memory, reasoning, and language.

[1] "Representational schemes that include rules for building complex symbolic structures out of simpler ones are called *combinatorial, generative,* or *productive*. Second, the meaning, or *semantic interpretation*, of a complex symbol is built up from the meanings of the syntactic parts of the symbol." (Stillings et al. 1995:5).

Cognitive science is enriched by the interdisciplinary work of researchers from the fields of psychology, linguistics, computer science, philosophy, neuroscience, and anthropology (Stillings et al. 1996; Thagard 1996). All of these disciplines contribute different research questions and approaches to the methodology of cognitive science.

Cognitive linguists believe that thought and language are closely connected, i.e. "language reflects patterns of thought" (Evans & Green 2006:5). More specifically, researchers in cognitive linguistics believe that the structure of language is influenced by the functions that language serves, e.g. as outlined by Halliday (1985), as well as by environmental, socio-cultural, developmental, and biological factors (e.g. Langacker 1999; Tomasello 2005). Therefore, they study the structure of human language, factors influencing linguistic structure, and developmental questions by drawing upon theoretical perspectives established in related fields, such as philosophy, psychology, linguistics, or anthropology (e.g. McCabe 2011). Moreover, they frequently relate the observed structures to theories and findings about the human mind, thereby "treating language as reflecting and revealing the mind" (Evans & Green 2011:50). By taking this multi-disciplinary approach, findings on structures of thinking, cognitive processes, and social communication are related to each other and distinct representations in language can be interpreted (Fauconnier 1999:96).

It is assumed that each situation can be conceptualized and linguistically expressed in different ways, which systematically differ from all other possible choices (e.g. Tomasello 1999). These differences are studied, for example, in terms of who is speaking (perspective), which information is salient (foregrounding), and how concepts and events are described (framing) (e.g. Talmy 2000; Lee 2001). Conceptualizations are also studied in terms of mental models (e.g. Johnson-Laird 1983; van Dijk & Kintsch 1983; van Dijk 2006), cognitive categories (e.g. Rosch 1976), and image schemas (e.g. Slobin 1996; Gibbs & Colston 2006)[2]. *Mental models* represent different constructs of the human world, such as knowledge in general as well as discourse specific situations (Johnson-Laird 1998). These mental models are constructed based on linguistic as well as non-linguistic input and they are constantly manipulated integrating new information. Thus Johnson-Laird (2004) argued that they are involved in different aspects of human cognition, such as

[2] Fore more detail on cognitive categories and image schemas see chapter 8 (Object reference).

representation of perceptual input, understanding of discourse as well as in thinking and reasoning.

The linguistic analysis that is central to this thesis is based on the outlined theoretical framework of cognitive linguistics. The research area in which this theoretical framework is applied is that of problem solving in which thinking and reasoning are integral processes. The next paragraphs outline the theoretical view on human problem solving that is assumed in this thesis.

As problem solving is a frequent activity in daily life, researchers have been interested in the mental processes that drive this activity and in the structure of these processes for a long time. Among the first proposals for a sequence of mental processes was the *test – operate – test – exit* unit, commonly referred to as TOTE (Miller et al. 1970). This unit contains the basic processes of 'test', 'operate', and 'exit'. It is based on the assumption that humans solve problems by mentally manipulating representations of possible states by different operations. A chess move shall serve as an example of this theoretical view. If chess players decide on the next move, they know the current state, i.e. the positions of all chessmen on the board, and they know which moves are allowed. The theory assumes that they would then think of possible new states by applying numerous possible moves; this process corresponds to the first 'test' process. Then they decide on one move and execute it by physically moving one chessman; this process corresponds to 'operate'. The new state is then evaluated and if it represents the desired goal state, this sequence is ended and a new sequence starts. In the example of a chess player, the execution of the action ends his/her turn. This example highlights that human problem solving behavior is conceptualized as information processing. Moreover, this framework assumes that problem solving is mainly taking place in the problem solvers' mind and information from the environment is only partially integrated. This view is proposed by Newell and Simon (1972) in one of the most influential theories on human problem solving as information processing.

A more recent approach to problem solving proposes that problem solving is an interactive process between the human and the environment. Researchers in the *situated cognition* community (e.g. Hutchins 1995; Kirsh 2009; Steffensen 2013; Cowley & Nash 2013) assume that humans shape the world to facilitate problem solving and this activity is influenced by environmental features. Kirsh (2009), for example, argues that people use objects to think with while working on a problem. He describes that people physically manipulate objects and try different configurations of objects to find a satisfactory solution to a problem. In cases in which the objects cannot be easily moved or consequences of these changes cannot be predicted, people may build real-world models to change arrangements within them. Additionally, it is assumed that people look for clues that provide guidance as to how objects can be used (e.g. Hutchins 1995; Norman 2002). The basic assumption is that this interactive way is essential to thinking because it helps to structure the problem space and it extends the problem solvers' perception and actions.

This thesis adopts the theoretical view that problem solving is a search through a problem space in which new states are continuously created, executed, and evaluated. Traditional theories investigated logic-based problems, such as playing chess that involve a high level of planning, i.e. mental simulation of possible operations. This thesis, on the other hand, investigates if similar problem solving processes can be identified when participants solve a real-world problem. The assembly task resembles a problem in which operations may be performed without much prior planning because states can be redone. Therefore, it is specifically interesting to find out if the same processes are verbalized or if new ones can be identified. Additionally, it will be investigated if these processes are as well-structured as identified for well-defined logic-based problems or if different structures are observed. In line with recent propositions concerning problem solving in the framework of *situated cognition*, the interaction between participants and the objects is studied to highlight how this interaction influences the problem solving process.

Traditionally, human problem solving has been studied, for instance, by means of verbal protocol analysis, i.e. which information is verbalized while participants work on a problem (e.g. de Groot 1969; Dörner 1987) or what they report about it afterwards (e.g. Russo et al. 1989). Newell and Simon (1972), for example, based their influential theory of human problem solving on two sources of evidence. First, they drew on theoretical analogies between computer systems and human thinking and reasoning. Second, they analyzed mental processes involved in problem solving by means of think aloud protocols of one participant for two logic-based well-defined tasks[3]. Based on the practice of using verbal protocols in studies on problem solving, Ericsson and Simon (1984, 1993) outlined the *model of verbalization of thinking* and a theory of verbal protocols as data in research. Adopting the same theoretical concept of the human mind as an information

[3] Newell and Simon (1972) analyzed one participant's protocol working on a crypt arithmetic task and the protocol of one chess player.

processing system, they described how information on these mental processes could be investigated in verbal protocols. This *protocol analysis* is based on the content of verbal protocols. In the proposed application of *protocol analysis* identified traces of mental processes are compared to computer generated process traces in order to evaluate the predictive power of the proposed model.

Some researchers (e.g. Roth 1985; Dörner & Bartl 1998) extended content-based protocol analysis by investigating the linguistic structure of think aloud protocols. These researchers investigated the difference in mental processes, such as forming hypotheses, by participants that succeeded in solving the problem as compared to participants who were unsuccessful. One further research group, headed by Caron-Pargue (e.g. 1991, 2003, 2010), presented accumulating evidence that mental processes in think aloud protocols are systematically represented by linguistic markers, such as verbs, conjunctions, and connectives. However, their approach is restricted in two ways. First, they exclusively analyze think aloud protocols. Second, their analyses are restricted to two logic-based problems, namely the Tower of Hanoi and the Chinese rings.

Similarly to Caron-Pargue and colleagues, Tenbrink (e.g. 2007, 2010) investigates systematic linguistic patterns in verbal data to connect those to the underlying mental processes. Based on Ericsson and Simon's (1993) *model of verbalization of thinking*, Tenbrink analyzed different kinds of verbal reports that were elicited on different tasks. However, contrary to the previously introduced studies, Tenbrink and collaborators (e.g. Wiener & Tenbrink 2008; Hölscher et al. 2011) investigated different kinds of verbal reports on problem solving tasks. They investigated think aloud protocols as well as retrospective reports on related tasks, such as the Traveling Salesman Problem (e.g. Tenbrink 2008) and wayfinding studies in virtual as well as in real-world environments (e.g. Tenbrink et al. 2011; Klippel et al. 2013). Furthermore, Tenbrink and collaborators (e.g. Vorwerg & Tenbrink 2007; Tenbrink et al. 2008; Andonova et al. 2010) investigated the linguistic structure of verbal data highlighting central aspects of mental representations such as descriptions of a visual scene by combining psycholinguistic methods and cognitive linguistic analyses. These analyses revealed information about the speaker's or writer's conceptualization of the described situation or event. Recently, Tenbrink (2010) combined previous findings and experiences regarding procedure of data collection as well as data analysis to propose the methodology of *Cognitive Discourse Analysis*.

The present thesis adopts the methodology of *Cognitive Discourse Analysis* because the elicitation of think aloud protocols allows for the analysis of thoughts that reach the participants' awareness and that are assumed to reflect mental processes. Furthermore, it allows for an observation over time. However, the content-based analysis allows only for the analysis of verbalized thoughts. By combining the analysis with in-depth analyses of linguistic features, it is assumed that the unconscious linguistic choices can be highlighted. The linguistic analysis, in this thesis, focuses on an in-depth analysis of two features, namely verb type and object reference. Whereas verb types encode the process that is verbalized, object references express the conceptualization of objects. Importantly, parts of the linguistic analyses investigate features that have not been studied in the proposed terms before. No verb type classification of problem solving processes was reported so far and object reference has been mainly studied in communicative settings. Therefore, the analyses combine well-established findings on linguistic features obtained in different conditions with theoretical assumptions about the investigated empirical design. As described in previous papers by Tenbrink and colleagues, the analysis investigates the correlation between these linguistic features and evidence from other discourse tasks or performance measures.

In sum, this thesis investigates the hypothesis that assembly performance can be described in terms of problem solving and how these processes are linguistically expressed. Concerning the methodology, this thesis aims at contributing to the establishment and refinement of *Cognitive Discourse Analysis* by providing an overview of previous works on the analysis of linguistic features in think aloud protocols and by applying the methodology to a new field within the area of problem solving.

1.2 Empirical studies

The investigation presented in this thesis is based on verbal data elicited during and after participants solved an assembly task. Participants were instructed to assemble a two-story dollhouse but no manual was provided to them. In order to study the influence of prior information about the goal state, i.e. the goal object in this case, on problem solving aspects, such as conceptualization of objects, participants were tested in three conditions. Prior information about the goal object was verbally provided in the instructions. In one condition participants were told to assemble a sensible object, i.e. no specific information about the goal object was provided. In the second condition, they were told that their task was to build a two-story dollhouse and in the third condition, they were additionally shown a picture of the assembled dollhouse. It was assumed that amount of prior information influences conceptualization of object parts within the mental representation of the goal structure as well as the perception and use of object features that provide information about possible functions, i.e. constraints.

The elicited verbal data corpus contains think aloud protocols of 56 participants that were recorded during task performance. In addition, retrospective reports of all participants were collected immediately after the dollhouse assembly. Both data sets represent unconstrained natural language that is assumed to reflect participants' unaltered thoughts. As it is generally agreed that verbal reports are not complete (e.g. Ericsson & Simon 1993; van Someren et al. 1994), the experimental session was closed by a number of guided questions regarding perception of specific object features.

Moreover, a second study was designed to collect instruction data. In this study 20 participants were instructed to assemble the same dollhouse as in the first study except that they worked silently. They were provided with specific goal state information: they were told that they need to assemble a two-story dollhouse and they were shown the picture. After their assembled dollhouse was assessed as correct, the participants were introduced to an addressee whom they were asked to instruct in a non-interactive setting given a shared workspace. The addressee could see the instructor and his workspace on a computer screen via skype and listen to his instructions. However, no interaction was possible. The elicitation of instructions was motivated by the fact that most research focuses on language use in interactive settings. The comparison between references elicited in a non-communicative setting in assembly for oneself to those collected in a communicatively motivated setting was expected to highlight the difference between speech for oneself and speech for others. This difference is important to the assumption that think aloud protocols reflect participants' unaltered thoughts. Additionally, this design included a silent control group for the third condition in the first study.

In sum, the present approach combines controlled experimental set-ups with the elicitation of unconstrained natural language by recording participants' verbalizations to the instruction of verbalizing everything that comes to their mind. The results of these studies are analyzed qualitatively focusing on the description of the individual phenomenon, e.g. the content of the processes that are identified. Moreover, the content-based qualitative analysis is extended by an in-depth description of linguistic patterns concerning verb phrase form, object references as well as discourse markers that are identified. The qualitative description is supplemented by a quantitative analysis of the frequency of the categories. The collected performance measures, i.e. assembly and instruction time as well as assembly and instruction success, are also quantitatively evaluated.

1.3 Structure of thesis

This thesis focuses on the identification of problem solving processes in unaided object assembly and their description with regard to their nature as well as their linguistic structure. Therefore, chapters 2 and 3 provide the theoretical background on problem solving and verbal data analysis. The remaining chapters 4 to 9 present the analyses of the empirical studies. Each of these chapters focuses on one specific aspect of the analysis.

Chapter 2 introduces the research area of *problem solving*. First, problem solving is defined and two classifications of problem types are presented. Then a detailed description of two influential theories of human problem solving are provided. Newell and Simon's (1972) seminal work on human problem solving as a search process through a problem space is introduced first. In the same line of thought, Dörner (1987) proposed that problem solving is a search through the 'Realitätsbereich' describing the different components of the problem space in much detail. Newell and Simon's (1972) theoretical work is important with regard to the general assumptions of human problem solving especially in combination with the methodology of *protocol analysis* as outlined by Ericsson and Simon (1993). Dörner's (1987) theory is combined with his proposal for a classification of problem types to classify the problem of unaided object assembly with different levels of prior information. Besides introducing these classical theories, chapter 2 reviews a more recent approach to study human problem solving in the framework of *situated cognition*. The chapter concludes by introducing methods that are employed in problem solving studied in the framework of classical theories and proposed by situated cognition theories.

Chapter 3, then, focuses on one method for studying thought through language, namely on verbal protocol analysis. After providing a general introduction to

Ericsson and Simon's (1984, 1993) *model of verbalization of thinking*, the chapter focuses on the methodology of *protocol analysis*. This methodology is the starting point for the analyses in this thesis. To illustrate the frequent use of *protocol analysis* in problem solving research and the scope of insights that have been gained by this analysis, the chapter reviews a selection of studies. In addition to this general introduction to the well-established methodology of *protocol analysis*, the methodology of *Cognitive Discourse Analysis* is introduced. The method extends Ericsson and Simon's (1993) content-based approach to verbal protocol analysis by a systematic analysis of linguistic features. As this thesis provides an exploration and refinement with regard to the methodology, basic theoretical assumptions as outlined by Tenbrink (e.g. 2008; 2010) are reviewed. The application of the methodology is exemplified in a review of reported empirical studies investigating different discourse tasks and problem solving scenarios. These findings serve as a starting point for the analysis pursued in this thesis.

Whereas these chapters provide the theoretical background for the analyses, chapters 4 to 9 focus on the empirical studies that are central to this dissertation project. After summarizing the empirical design in chapter 4, each of the remaining chapters presents one feature of analysis in detail. All of these chapters have the same general structure. First, the theoretical background for the specific analysis is summarized. Second, the research questions for the analysis are formulated and expectations are outlined. Third, the procedure of analysis is described to illustrate how the theoretical considerations are operationalized. Fourth, the results are presented. Fifth, the obtained results are discussed in the light of previous findings and new contributions are highlighted. For reading comfort, the appendix for each chapter is to be found at the end of the respective chapter. After this overview on the general structure, the content of each chapter will be briefly summarized.

Chapter 4 introduces the motivation for the design of the two empirical studies. Besides providing an overview on the number of participants that were tested in the experiments, the procedure of each experiment is described including the exact instructions, a description of the equipment that was used, and the technical problems that were faced. Chapter 4 concludes by providing a general overview on the qualitative and quantitative analyses that are performed in the different chapters of analysis.

Chapter 5 summarizes all performance measures that were collected along with the verbal protocols. These measures were assembly and instruction time as well as assembly and instruction success. Furthermore, the chapter provides a summary on the number of the experimenter's reminders to participants to keep thinking aloud. The video data of instructors assembling the dollhouse for themselves was analyzed regarding spontaneous instances of talk to oneself.

Chapter 6 presents the content-based analysis for identifying problem solving processes in think aloud protocols in line with traditional problem solving research paradigms. Besides focusing on the identification, the processes are described in the qualitative analysis highlighting frequent linguistic patterns with regard to temporal markers, adverbs, and further discourse markers. The quantitative analysis, on the other hand, focuses on the frequency of identified processes within think aloud protocols. Moreover, recurring sequences of problem solving processes are identified across think aloud protocols.

Chapter 7 focuses on one specific linguistic feature of the identified processes, namely on the verb phrase. Taking a functional approach to grammar as proposed by Halliday (1985), verb types in the linguistic representation of problem solving processes are described. A review of literature on modality and verb types as well as studies that investigated aspects of verb phrases in think aloud protocols, highlights that only selected aspects of verb phrases were studied so far, such as tense or mood. To capture the different aspects expressed in verb phrases systematically, verb type, modality, and tense are analyzed in this chapter. The discussion of the findings focuses on the identification of systematic differences in verb phrases between problem solving processes.

The analysis presented in chapter 8 is concerned with the mental conceptualization of the given objects within the goal structure. First, an overview on accounts of reference production in communicative settings is presented since the communicative, interactive scenario is commonly used in studies of discourse analysis. Based on a literature review on reports investigating the influence of the speaker's cognitive state as opposed to the assumed addressee's cognitive state on the produced referential form, a theoretical proposal concerning the different degrees of interactivity and engagement between speaker and addressee is presented. Different types of conceptualization are distinguished in the analysis. Direct assignment of function is expressed in utterances such as "also das hier ist das Dach nehm ich an" (well this one is the roof I suppose). In this example the deictic reference 'das' (this) is connected to the domain specific terminology 'Dach' (roof) by the verb 'ist' (be). Indirect function assignment is analyzed by identifying goal structure specific nominal references, such as 'Haus' (house), 'Wand' (wall), or

'Etage' (story), at different points throughout the assembly process. The discussion focuses on the influence of prior information on referential form in direct as well as indirect function assignment. The comparison between referential form in think aloud protocols and in instructions aims at highlighting the influence of communicative intention on the verbalization of object conceptualization.

In chapter 9 references to physical and cultural constraints are examined. This analysis is in line with recent work on problem solving in the area of *situated cognition* in which it is assumed that humans solve problems by actively engaging with the environment. As this is a not yet well-researched approach to problem solving, the presented analysis is explorative. Based on a review of literature discussing constraints in the context of problem solving in general, assumed functions of constraints and their use are proposed. The qualitative analysis on the nature and function of constraints is supplemented by a quantitative analysis of the frequency of constraints and their distribution within assembly conditions and in instructions. The discussion highlights which constraints are used and at what time during the assembly suggesting possible effects of the recognition of constraints on the problem solving process. Furthermore, the strategic use of constraints in instructions is discussed.

Chapter 10 concludes this thesis by pointing out that unaided object assembly as investigated in the presented study, resulted in the creation of new procedures expressed by *hypotheses*, *actions*, and *evaluations*, for example, instead of representing actions based on known procedures. Moreover, the systematic linguistic analysis revealed that those processes differ with regard to verb type, use of modal adverbs, and referential form. By summarizing the main findings on the linguistic representation of problem solving processes, the contribution of *Cognitive Discourse Analysis* can be critically discussed. After a general reflection on the adopted procedure, the thesis concludes by some remarks on possible future directions.

2 Problem solving – A general introduction

2.1 Problem: A definition and two classifications

Imagine, you are at a friend's house and you want to make yourself a coffee. So you walk into the kitchen and find the coffee in the fridge. Now you hope to find an ordinary coffee machine. However, what you find looks like the object in Figure 2.1. In this case, you are fortunate because you know the goal state and you know how to brew coffee with it, i.e. by placing it on the stove. Now you need to figure out how to disassemble the object into its parts. By this disassembly procedure you already know how to assemble it after you have inserted water and new coffee. But there are still open questions. First, you need to figure out how much coffee and water need to be inserted. Second, you need to know which temperature on the stove is appropriate and at what time the coffee is ready. At that time you have a number of choices, i.e. you may wait for your friend to come home, you can look for information on the internet, or you simply try it. All of these choices represent problem solving strategies.

Figure 2.1: Mokka maker on the stove (H. Bindewald).

This chapter will start by providing a general definition of the term *problem* before two more fine-grained classifications are presented. Then the focus shifts from the problem as such towards the human who solves it. Two influential theories about mental processes involved in human problem solving will be reviewed in detail, i.e. Newell and Simon (1972) and Dörner (1987). Whereas those two theories describe human problem solving in terms of information processing, a more recent approach understands human problem solving as an interactive process between the problem solver and the environment. The approach of *situated cognition* on problem solving will also be introduced as it presents a different perspective on problem solving activities. Whereas traditional problem solving research focused on laboratory settings, researchers in situated cognition investigate problem solving observed outside the laboratory (e.g. Cowley & Nash 2013; Steffensen 2013), i.e. as it occurs in 'the wild' (Hutchins 1995). The chapter will conclude with the presentation of some commonly used problem solving methods as proposed in the literature.

People are faced with a *problem* if the initial state and at least parts of the desired goal state are known but the exact actions that need to be taken to get from one to the other are unknown, as in the introductory example. This gap of knowledge needs to be filled with something new, thus problem solving is often classified as a *creative thinking* process (e.g. Dunker 1960; Dörner 1987). This is different for recurring tasks that can be solved by retrieving sequences of actions from long-term memory[4]. In these cases, *reproductive thinking* (Funke 2003) can be applied to generate an appropriate solution.

Different classifications of problems can be distinguished with regard to the granularity level of the classification and their focus, i.e. information on the states of the problem solving process and appropriate operators. The most general distinction refers to the amount of information about the goal state: well-defined vs. ill-defined problems (McCarthy 1956). A very fine-grained classification, in contrast, takes into consideration the type of problem, its characteristics and content as well as the information processes that are required for solving the problem (Funke 2003).

McCarthy's (1956) broad distinction of well-defined and ill-defined problems is based on the specificity of information available about the goal state. If problems are well-defined, the characteristics of the goal state are clear and thus any current state can be evaluated regarding its potential adequacy as the goal state. An instruction such as "Paint the living room purple." (adapted from Funke 2003) contains information on the action (paint), the location (the living room), and the color (purple). If the painting has been performed with purple paint in the living room, the problem is successfully solved. On the one hand, it could be argued that this instruction does not pose a problem to a professional painter but it can be considered a problem for a novice. The instruction: "Make my flat look nicer", on the other hand, could even pose a problem for a professional painter. Although this instruction contains information on the location (my flat), no explicit information on the actions to be taken is provided. Since the goal state 'look nicer' is very subjective, the painter will probably ask for additional information. The initial instruction represents an ill-defined problem because it can be solved by multiple solutions and success is not guaranteed (also Kitchener 1983:223).

Based on this introduction, the following working definition is proposed: if the current state does not represent the intended goal state, some things need to be changed. If the actions that need to be taken to change the current state into the desired state are not known, the person faces a problem. This problem can be solved by creatively combining known actions and objects or by including previously unknown objects and procedures. After this very general definition of problems, the next two sections introduce two ways of distinguishing different kinds of problems.

[4] A presentation of the memory structure assumed by Newell and Simon (1972) is presented in sub-section 2.2.2. Furthermore, it needs to be pointed out that learning effects are not addressed in this thesis.

Greeno (1978) proposes a classification of problems based on mental skills that are required to solve them. His classification will be presented because this thesis is interested in mental processes involved in problem solving. Since Dörner published most of his works in German only, his contributions to the theory of human problem solving are just recently recognized beyond the German speaking community (e.g. Funke 2003; Betsch et al. 2010; Schott & Ghanbari 2012). Dörner's (1987) theoretical considerations are important because, similarly to Greeno, his classification focuses on the components of the problem space and he specifically considers real-life problems in his classification. Since the analysis of problem solving in this thesis is investigated with regard to a real-life problem task, his theory is sketched as a second approach to classifying problems. In order to illustrate the applicability of the reviewed classifications, some well-researched problem tasks are classified according to them in sub-section 2.1.2.

2.1.1 Two classifications of problems: Greeno (1978) and Dörner (1987)

Greeno (1978) proposes a typology of problems that is based on the cognitive skills required to transform the initial state into the goal state. He distinguishes between *problems of inducing structure*, *problems of transformation*, and *problems of arrangement*. In *problems of transformation*, the initial and the goal state are well-defined and the operators for getting from one to the other are known to the problem solver as well. Greeno (1978) argues that two kinds of understanding are crucial in solving these kinds of problems: initial understanding of the problem and understanding of the solution and the selected operations (Greeno 1978:252). The construction of a *cognitive representation*[5] is important in the first kind of understanding because the generation of new possible states proceeds within this cognitive representation. Understanding the problem description and structuring the given information create an appropriate problem representation. The structuring process is facilitated by personal *problem-solving knowledge* that helps to identify concepts in the problem description. These concepts are identified by matching information specified in the problem description to the problem solver's general conceptual knowledge. This matching determines the meaningfulness of the constructed cognitive representation in which the problem solving process is performed. Potential solution states are generated by a search through the cognitive representation and they are evaluated against the background of

[5] Newell and Simon (1972) coined the term *problem space* to describe this cognitive representation. Their theory is defined in detail in sub-section 2.2.2.

meaningful general concepts. The search process can only be successful if the problem solver understands the underlying "relations between relatively large subcomponents of the solution" (Greeno 1978:255).

In problems in which the goal state can be reached by arranging known elements according to given constraints, the problem solver needs abilities that constitute a successful constructive search (*problems of arrangement*). In the constructive search process partial solutions are generated and those are tested for their consequences. As this search is based on the *generate-and-test procedure*[6], the problem solver needs to generate large amounts of potential solutions. Retrieving and applying known solution patterns facilitates this process. If the constraints are not explicitly stated in the problem description, the problem solver needs to be aware of general principles that confine search, e.g. phonological constraints or frequent combinations of letters in crossword puzzle tasks. In order to select the right general principles, the problem description and the goal need to be well understood. Successful problem solvers are assumed to have stored algorithms for forming arrangements (Greeno 1978:264).

There are also problems in which elements of the problem description need to be restructured to find the solution (*problems of inducing structure*). These problems are based on the understanding of underlying relational structures, such as in problems of analogy. In these problems, the process of understanding encompasses recognition of relations and construction of an integrated cognitive representation (Greeno 1978:243). In analogical problem solving, common relations between elements of the base and the target problem need to be identified and mapped to each other by knowledge transfer thereby generating new solutions. These potential solutions need to be assessed against the characteristics of the goal state.

Since not all problems can be clearly classified as belonging to one of the categories exclusively, Greeno (1978) stresses that the proposed typology does not represent a taxonomy. Most problems will contain elements of all three types. Therefore, understanding the problem state and the goal is crucial for success. More explicitly, Greeno (1978:263) defines understanding as the search for a cognitive representation in which the elements of the problem and the problem goal fit together. This definition of understanding is not limited to the area of problem solving, Greeno (1978:266) rather assumes that "there may be some

[6] A description of this method is presented in the section on frequently observed problem solving methods and heuristics (sub-section 2.3.).

powerful general skill in understanding, planning, and composing rather than a single ability in problem solving". Therefore, he assumes that more detailed knowledge about general problem solving skills may lead to a "satisfactory theory of general intelligence" (Greeno 1978:266).

A different categorization of problems is presented by Dörner (1987). It is based on the assumption that problem solvers encounter different kinds of mental barriers that hinder the transformation from the initial state into the goal state. The basic assumption is that the *problem space*, that is identical to Greeno's (1974) notion of *cognitive representation*, consists of initial state, goal state as well as all intermediate states. Furthermore, it includes all possible operators[7]. As pointed out by Greeno (1978), Dörner also proposes that the structure and especially the components of the *problem space* influence the selection of problem solving strategies. The selection is further constrained by barriers that are faced by the problem solver. Dörner (1987:11-13) defines three barriers that depend on knowledge about the necessary operations and/or the goal state (summarized in Table 2.1):

- *Interpolation barrier*: Initial and goal state are known as well as the operators needed to reach the goal state. The problem solver needs to find a suitable combination or hierarchy of necessary operators.
- *Synthesis barrier*: Initial and goal state are known but the operators are unknown. The problem solver needs to select the necessary operators from a known repertoire of operators.
- *Dialectic barrier*: The initial state and the operators are known but the goal state is unknown. It is defined during the dialectic process that consists of hypothesis, antithesis, and synthesis.

		Level of specificity of the goal state	
		high	**low**
Knowledge about the operators (means)	**high**	Interpolation barrier	Dialectic barrier
	low	Synthesis barrier	Dialectic and synthesis barrier

Table 2.1: Different kinds of barriers that need to be overcome by problem solvers (adapted from Dörner 1987:14).

[7] The components of the problem space were first defined by Newell and Simon in their seminal work "Human Problem Solving" (1972) as will be described in detail in sub-section 2.2.2.1.

Problems in which the problem solver is faced with an interpolation barrier are comparable to problems of transformation as defined by Greeno (1978). These problems call for an arrangement of known operators. Problems that pose a synthesis barrier can only be solved by a creative process in which established ways of thinking are overcome (Dörner 1987:13). But according to Dörner (1987), the majority of problems that people are faced with on a daily basis are problems that pose a dialectic barrier. In these problems, there is awareness for the problem at hand but the goal state is unknown. Dörner (1987:13) provides the example of a historian who needs to fill the gaps in a document that is only partially preserved. This example does not seem very explanatory for several reasons. First, it cannot be classified as a daily problem for many people because it is a rather specific case in one field of science. Second, the goal state is known to the historian in general terms, i.e. recover the document in a way to make it readable for yourself and others.

Dörner's (1987:13) explanation for the selection of the term 'dialectic' may help to get a better picture of the class of problems that he is referring to. He states that solutions to these kinds of problems are often found in a dialectic process in which a potential solution is evaluated against outer and inner objections. Outer objections are defined as objections that do not result from the potential solution itself but are constraints induced from outside. Inner objections, on the other hand, arise from the individual problem components themselves that are combined in a potential solution (Dörner 1987:13). This more detailed description of the class of problems with dialectic barriers highlights that they are comparable to ill-defined problems. Returning to the example introduced to illustrate ill-defined problems, this correspondence gets even clearer. The instruction "Make my flat look nicer." implies a number of constraints regarding potential solutions. Solution candidates need to be checked with the speaker and possibly the landlady (outer objections) and they need to satisfy inner constraints of the flat as a building (e.g. stability, noise etc.). Reither (1996) proposes methods that are specifically suited for solving the different problem types described by Dörner (1987) as summarized in Table 2.2.

Similarly to Greeno (1978), Dörner stresses that problems may pose more than one barrier and they are not independent from the problem solver. The latter is assumed specifically in problems with synthesis barriers that are constituted by established procedures and ways of thinking (Dörner 1987:13).

Problem type (Dörner 1987)	Preferred problem solving strategy
Interpolation barrier	Analytical heurism, i.e. analysis of specific differences between current and goal state and selection of adequate operators for reducing the difference.
Synthesis barrier	Creative thinking to overcome the functional fixedness of objects. Operations are systematically released from its learned functions and rearranged creatively.
Dialectic problem	Systematic reduction of possible action moves. It is characterized as the repeated creation of constraints that are then overcome.[8]

Table 2.2: Preferred problem solving strategies when facing an interpolation or synthesis barrier or a dialectic problem (summarized from Reither 1996:196).

2.1.2 Classification of problems that are commonly investigated in psychological studies

A number of well-researched problems belong to the class of problems of transformation, as defined by Greeno (1978). The problem of the *Tower of Hanoi* represents a transformation problem in which the initial and the goal state are well-defined. The possible moves of elements, i.e. disks, are highly constrained[9]. In this task the problem solver needs to find the correct sequence of moves that adheres to the constraints and represents the described goal state. The problem of *Hobbits and Orcs* (e.g. Greeno 1974), formerly referred to as *Missionaries and Cannibals*[10], is structurally similar but involves different constraints. In *toy-block problems*, such as *Tetris*, a goal arrangement is presented to the participants and they need to arrange the given blocks accordingly by preserving stable structures. A strategy that guarantees high success rates for problems of transformation is means-end-analysis[11] (Greeno 1978:249) in which sub-goals that can be satisfied are

[8] Reither (1996:197) points out that dialectic problems are rarely studied experimentally hence there is not much evidence as to how people go about solving this class of problems.

[9] The following constraints are formulated: "There are three discs of unequal sizes, positioned on the far-left side of three pegs so that the largest disc is at the bottom, the middle-sized disc is in the middle, and the smallest disc is on the top. Your task is to transfer all three discs to the peg on the far-right, using the middle peg as a stationing area as needed. You may move only one disc at a time, and you may never move a larger disc on top of a smaller disc." (Sternberg & Sternberg 2012:452).

[10] For the exact problem description of this problem see for example https://www.courses.psu.edu/psy/psy002_nxt6/hobbits&orcs.html (17.04.2012).

[11] Selected common problem solving methods will be described in detail in subsection 3 of this chapter.

defined and which eventually result in the predefined goal state. Other common problems of transformation and arrangement are *chess problems* in which the initial state as well as possible goal states are known and the moves are highly constrained (e.g. de Groot 1965; Newell & Simon 1972).

There are also problems that represent transformation problems in which the elements that need to be restructured are functions of objects. Some types of insight problems belong to this class of problems that pose a synthesis barrier. A classic example was presented by Duncker and is commonly referred to as *candle problem*. In the problem description Duncker (1974) asked his participants to install three small candles for optical experiments on the wall besides the door. In the laboratory, there was a table on which, among other things, there were pieces of paper, string, pencils, and an ashtray. All of these objects were irrelevant to solving the problem successfully but there were also three small boxes and pins (see Figure 2.2). The participants were told that all objects that were needed for a successful completion of the given task were to be found on the table. Additionally, the instructions stated that everything on the table could be used in any way the participant would see it fit (Duncker 1974:104).

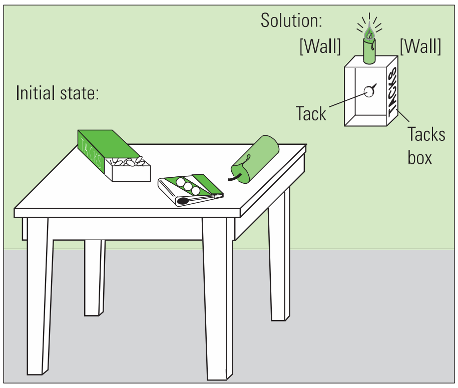


Figure 2.2: Duncker's (1945) candle problem represented by the necessary objects (initial state) and the solution (Reisberg 2010:463).

In this problem description all objects are provided and need to be rearranged in order to install the candles besides the door. This task can only be solved if the problem solver abandons his concept of boxes as containers of pins. By considering the idea of boxes as a candle stand, the objects can be rearranged with regard to their location but more importantly regarding their function. One pin needs to be used to fasten the small box on the wall and then the candle can be placed inside the box (Duncker 1976:103). This insight can be described as a creative process because known objects need to be conceptualized differently by abandoning their known function. A similar problem is the *matchstick problem* in which participants need to rearrange matchsticks, mentally or physically, to represent a defined structure (e.g. Knoblich et al. 1999; Knoblich et al. 2001).

2.2 Theoretical approaches to human problem solving

2.2.1 Theories on human problem solving

So far, two approaches to the categorization of problem types were presented. These categorizations do not make any claims about the mental processes that drive problem solving. The following section will provide a brief overview on the proposed nature of the processes that result in problem solving activities before three theories on human problem solving will be outlined in more detail.

One of the first hypotheses on explaining human behavior on problem solving tasks was proposed by George A. Miller, Eugene Galanter, and Karl H. Pribram in 1960. They postulated that human problem solving behavior can be summarized in a Test–Operate–Test–Exit[12] unit that is executed in a repetitive loop (Miller et al. 1970:26). The authors assume the TOTE pattern to be the underlying structure of human planning. They specify this claim by stating that

> "the test phase of the TOTE involves the specification of whatever knowledge is necessary for the comparison that is to be made, and the operational phase represents what the organism does about it – and what the organism does may often involve overt, observable actions." (Miller et al. 1970:31)

The authors do not go into detail about the processes that are involved in the *operational* or *test phase*. They more generally state that any mental process that determines the appropriateness of the operational phase belongs to the test phase. As Miller et al. (1970) stress, their thoughts are novel because they deviate from the well-established concept of the reflex arc[13]. Therefore, their first proposal cannot be expected to specify all details. Besides their idea of an inherent feedback loop in the TOTE unit, they also stress the hierarchical structure of TOTE units. More

[12] In Miller et al. (1970) as well as in most literature it is referred to by its acronym TOTE unit. This abbreviation will be adopted hereafter as well.

[13] The novelty of the proposed thoughts can also be inferred from remarks such as the following: "The idea that problem-solving can be represented as searching through a large set of possibilities until we find one that solves the problem may seem odd at first, and somewhat novel to a person who has taken his heuristic plans for granted." (Miller et al. 1970:167).

specifically, Miller et al. (1970:32) assume that "the operational components of TOTE units may themselves be TOTE units"[14]. To exemplify their proposal, Miller et al. (1970) compare problem solving to perceptual searches through the environment. When facing a perceptual search problem, such as finding a hammer, the searcher can be observed to scan the room. This behavior can be described in terms of the TOTE schema. In the operational phase the searcher orients his/her attention toward an object. This action is followed by the perceptual test. In the test phase the located object can be evaluated as satisfactory or a new operational phase starts. To limit the duration of the search process, a stop-rule may be executed in the test phase. The processes that are at work in a mental search problem can be directly compared when the objects are assumed to be a set of alternative hypotheses (Miller et al. 1970:161).

De Groot (1969) was among the first scholars to provide a detailed account of the mental processes involved in problem solving. He shows that the *cycle of mental processes* is the basic unit to such diverse human endeavors as creative, analytical, and interpretative thinking. The *cycle of mental processes* contains the following steps: *end–problem–means–freedom–uncertainty–choosing–trying–testing* (deGroot 1969:11). First, the problem solver needs to formulate a problem and thereby the attention is shifted to the means by which the end can be reached (de Groot 1969:7). The problem solver is free in the choice of strategies. This is likely to result in uncertainty about the adequacy of the means that are to be employed. Once a choice has been made, the selected strategy is applied and the resulting state is evaluated regarding the effectiveness of the strategy in a *process of evaluation*. Additionally, de Groot (1969:14) identifies *observation*, *feedback*, and *new input* as important components of experiential processes[15]. Even though observation is not explicitly mentioned in the *cycle of mental processes*, it is the first activity that connects the organism to the present situation in which the organism "perceives certain (selected) aspects of S_1[16]" (de Groot 1969:3). Observation, therefore, constitutes the first process in the *empirical cycle*[17]. Highlighting the influence of

[14] More specifically, Miller at al. (1970:32) state that "the operational phase of a higher-order TOTE might itself consist of a string of other TOTE units, and each of these, in turn, may contain still other strings of TOTEs, and so on."

[15] The term *experiential processes* is based on de Groot's (1969) fundamental belief on which he rests his thinking. He believes that "the organism transform[s] its experiences and observations into *experiences about* the world – which will enable it to function more efficiently than it did before, in a less experienced state." (de Groot 1969:1).

[16] "S1 = the situation as it presents itself to O [= the organism]" (de Groot 1969: 2).

[17] De Groot (1969:2-3) introduces the *empirical cycle* to describe the processes that are involved in gaining experience without reflection. The cycle encompasses the process of

control processes in the cycle, de Groot (1969:13) specifies that feedback, in the sense of evaluation of results, may lead to modifications of the original idea. The importance of feedback is further stressed by the proposal that

> "a major cycle will often comprise a whole series of smaller ones, each dealing with a subproblem of the larger one. This complex, hierarchical structure is made possible by repeated (routine) 'shifts from end to means', which the subject performs repeatedly." (de Groot 1969:9)

This quote states that complex actions such as playing chess (*macro-cycle*) are divided into a sequence of *micro-cycles*. As the outcome of one cycle is "fed back into the input data that start off the next one" (de Groot 1969:12) repetitious cycles form a spiral. Discussing a number of studies and findings, e.g. the "hermeneutic cycle of Verstehen" (de Groot 1969:12), de Groot illustrates that the proposed cycle of mental processes is the basic unit of numerous human mental activities such as creative, analytical, and interpretative thinking including problem solving.

2.2.2 Problem solving as information processing

In the following paragraphs two influential theories will be sketched that also have the search paradigm at the centre of their theories. Proceeding in time, the theories by Allen Newell and Herbert A. Simon as well as by Dietrich Dörner focused more deeply on the nature of the cognitive processes that drive problem solving. However, both approaches are based on the assumption that problem solving is comparable to information processing as found in computers. Gentner et al. (2001:7) consider the "common analogy between human information processing and the processing performed by digital computers" central to the development of cognitive science as this analogy guided work on cognitive processes such as memory, attention, and problem solving. This view is already stated by Newell and Simon (1972), who remark that "an abstract concept of an information processing system has emerged with the development of the digital computer" (Newell & Simon 1972:5).

2.2.2.1 Allen Newell and Herbert A. Simon – human problem solving as the execution of a program of primitive information processes

As introduced above the thought that made Allen Newell and Herbert A. Simon's theory ground breaking was the proposition that human problem solving needs to

---

perception, the organism's reaction to the present situation, the process of acting in 'the world', and the subsequent process of perception of the altered state to evaluate it.

be described and understood in terms of information processes. Along with this revolutionary concept Newell et al. (1958:151) proposed three basic assumptions about their theory. First, they assume that symbolized information is contained in different kinds of memories and that those memories are connected to each other by ordering relations. Second, the authors propose that there are numerous 'primitive information processes'[18] (Newell et al. 1958:151) that operate on the information that is held in the memory structure. Third, these primitive information processes are combined into 'programs of processing' (Newell et al. 1958:151) based on a clearly defined set of rules. By following the assumption that human behavior can be understood as programs of processing, Newell et al. argue that "the vaguenesses that have plagued the theory of higher mental processes and other parts of psychology disappear" (Newell et al. 1958:166). The authors are certain that their theory and their approach to use a program to study the processes at work will yield insights on a fundamental question. The question of how these mental processes "can be compounded out of elementary information processes and hence how they can be carried out by mechanisms" (Newell et al. 1958:152). Importantly, Newell et al. (1958) stress that they do not understand the computer as analog to human behavior, they rather want to illustrate the advantage of understanding problem solving as being based on the execution of programs. They postulate that each mental activity can be described in such precise terms that a machine can execute the same actions. This basic idea is stressed by a quote that highlights the intentions of the first 'symposium on information theory' in 1956 as cited by Lenzen (2002:18). In their 1958 paper, Newell, Shaw and Simon illustrate the adequacy of their theory by referring to de Groot (1946) who described the behavior of chess players in ways that correspond to their idea of identifiable and describable processes.

Their program, called the *Logic Theorist* (abbreviated LT), is based on basic assumptions about the processes involved in human problem solving. Those processes, as described in the previous section, are summarized as "searching for possible solutions, generating these possibilities out of other elements, and evaluating partial solutions and cues" (Newell et al. 1958:152). The basic idea of programming LT was not to investigate problem solving processes but to show that a program based on primitive information process can find correct proofs to mathematical problems. To accomplish this task, the computer needs to generate

[18] Newell et al. (1958:151) do not define primitive information processes exactly but they state that "each primitive process is a perfectly defined operation for which known physical mechanisms exist."

numerous alternative expressions. This generation process is driven by the application of four methods that are stored in its depository, namely *substitution*[19], *detachment*[20], *forward chaining*, and *backward chaining*[21]. As all of these methods are based on comparisons, Newell et al. (1958) identified and programmed two important sub-processes: the *matching process* and the *similarity test*. In the matching process two sub-expressions are made equal. The similarity test determines if two expressions are similar. The elemental idea for programming the LT to generate numerous possible answers was the assumption that "in its simplest aspect, the problem-solving process is a search for a solution in the very large space of possible solutions" (Newell et al. 1958:159).

The analysis of the methods employed by the LT as specified in the program output as well as the sequence in which the methods are applied and the theorems that are employed revealed the following results. The LT does not apply the methods in random order but tries to solve the given problems by using substitution first. If this method does not yield satisfactory results, the program applies the detachment method. If this as well as chaining forward fails, then chaining backward is tried (Newell et al. 1958:159). Newell et al. (1958) call this effect 'method set'[22]. Furthermore, two hierarchies of processes were identified, namely the *component problem hierarchy* and the *sub-problem hierarchy*. The first kind of hierarchy describes the application of the 'method set', i.e. the current problem is solved by the consecutive application of different methods. Hence, the current problem is divided into four problems if all four methods need to be applied to derive a satisfactory solution or to end the search process. The sub-problem hierarchy describes the process in which new theorems are derived, tested, and used as input for new generations of theorems. These sub-problems may yield a satisfactory solution in the end. The *problem–sub-problem hierarchy* was already observed and described by de Groot (1946) (cited in Newell et al. 1958:162). Based on these and other reported findings on human problem solving, Newell et al. (1958) conclude

[19] In substitution new expressions are generated by "the substitution of variable and replacements of connectives." (Newell et al. 1958:158).

[20] In detachment, the validity of a previously proven theorem is used to generate a new and valid expression. (Newell et al. 1958:158).

[21] Both methods of chaining are also based on the rules of syllogism as described for detachment. They differ in the direction of the argument (for examples see Newell et al. 1958:158).

[22] Newell et al. (1958) use the term set as it is defined by Johnson (1955), namely "a readiness to make a specified response to a specified stimulus" (cited in Newell et al. 1985:159).

> "that the processes that we observed in LT are basically the same as the processes that have been observed in human problem solving in other contexts" (Newell et al. 1958:165).

In 1972, Allen Newell and Herbert A. Simon published their groundbreaking book "Human problem solving" that is now a standard for anyone who is interested in human problem solving. In their seminal work, they follow up on the ideas that have been summarized in the previous paragraphs by providing a more thorough description of the memory structures and the mental processes that are involved in problem solving. Their theory is based on the idea that humans can be viewed as processors of information (Newell & Simon 1972:5). This view allows for a reasonable comparison between human problem solvers and computer programs that solve problems of the same kind. But Newell and Simon (1972) also stress that their theory does not represent the most abstract form of an information processing system it rather "introduces a suitable abstract information processing system to describe how man processes task-oriented symbolic information" (Newell & Simon 1972:5). To accomplish this aim, Newell and Simon (1972) restricted their investigations to short tasks, i.e. tasks that do not take longer than 30 minutes to be solved, and symbolic tasks[23] that posed a moderately difficult problem for the problem solver. In their book, they investigate and describe the behavior of particular problem solvers on these problems.

With regard to the scope of their theory, the authors argue that the investigation of the behavior of one particular man when solving one particular task yields "a symbolic model on the basis of which pertinent specific aspects of the man's problem solving behavior can be calculated" (Newell & Simon 1972:5). But the behavior identified in the individual can be assumed to hold for other human individuals as well because the individual's data is abstracted "down to a level where a simple interpreter (such as a digital computer) can turn the description into an effective process of performing the task" (Newell & Simon 1972:11). Furthermore, the approach of comparing computer performance and solution to human behavior allows for very precise descriptions of the input and the processes of output creation. Each new state can be described in relation to all previous states and the operations that have been applied to yield this new state because "a program generally specifies a discrete change in a single component of the state[24] (that is, in just one symbol structure) at a moment in time" (Newell & Simon

[23] Those symbolic tasks were chess, problems in symbolic logic, and algebra like puzzles (Newell & Simon 1972:3).

[24] A *state* is defined as a collection of symbolic structures in a memory (Newell & Simon 1972:11).

1972:11). As a counterpart to the output data of the computer program, Newell and Simon (1972) propose to use think aloud protocols to gain insight on human thought processes.[25]

Newell and Simon (1972) identify two essential components of problem solving, namely the *information processing system* and the *task environment* (Newell & Simon 1972:88). Both components will be briefly described in the following paragraphs. The general structure of the *information processing system* (hereafter abbreviated as IPS) is illustrated in Figure 2.3 (p.31). The memory, more specifically the long-term memory, stores and retains symbol structures[26]. The processes that operate on the symbols[27] and that produce new symbols are called information processes. The *processor* that is connected to the receptors, receiving input, as well as the effectors, contributing output, consists of three basic structures. It encompasses a set of *elementary information processes* and a *memory structure* that has the symbol structures of the elementary information processes as its input and output. Newell and Simon (1972:29) characterize elementary processes as those that "are not further analyzed in the theory into still simpler processes". Furthermore, the processor contains the *interpreter* that determines the sequence of the elementary information processes that is to be executed.

The IPS, as illustrated in Figure 2.3 on page 31, contains a number of features that characterize it as an adaptive system. These features are thought to be independent of the problem solver and the task environment. The IPS varies only marginally between individuals with regard to memory size, rates of reading and writing, and rates of accessing modes within the different memories. Additionally, the information processing as well as the rates of retrieval of elementary information attribute a serial character to the IPS. The global program organization exhibits a production-like and goal-like character (Newell & Simon 1972:792). The memory system contains three memory structures, namely the *short-term memory*, the *long-term memory*, and an *external memory*. The *long-term memory* (abbreviated LTM) contains a potentially infinite vocabulary of systems, has an essentially infinite capacity for symbol structures[28], and is an

[25] For an in-depth description of the methodology including a discussion of its strengths and weaknesses see chapter 3 (Verbal protocol analysis and Cognitive Discourse Analysis).

[26] Symbol structures are defined as "states of instances of symbols that are connected by a set of relations" (Newell & Simon 1972:20).

[27] The authors repeatedly emphasized the notion of symbol tokens as those symbolic elements that can be judged to be equal or different by a computer program (Newell & Simon 1972:23).

[28] Remember that symbols are the smallest unit of information held in LTM. If symbols are connected by relations, they are referred to as symbol structures.

associative memory. Similarly, the *short-term memory* (abbreviated STM) also works in an associative way but it can only hold few symbols. Based on previous research, Newell and Simon (1972:795) assume that the STM can hold five to seven symbols at a time and that there is little variation over tasks. Oberbauer (2012) argues against Newell and Simon's (1972) hypothesis that there is a specific amount of information that can be held in STM. He proposes a cognitive model illustrating that the limited capacity of working memory is due to interference by superposition of distributed representations. According to Oberbauer (2012), this allows the conclusions that the capacity of the STM is variable over tasks.

Information that is stored in STM is immediately available as input to the IPS system and all output of the IPS is first stored in STM before it may be transferred to LTM. In addition to these two internal memory structures, Newell and Simon (1972) propose that the IPS has access to an *external memory* (abbreviated EM) that is defined as the "immediately available visual field" (Newell & Simon 1972:809), for example a paper or a chessboard. Functionally, STM and EM appear to form a single unit in which EM stores information that cannot be held in STM anymore. EM provides fast access to this information in the upcoming processing act, is potentially infinite in capacity, and can be accessed by means of linear scanning.

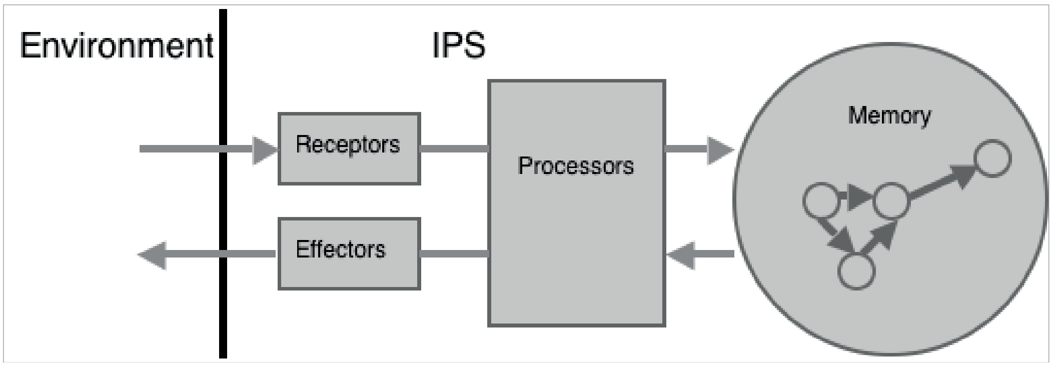


Figure 2.3: General structure of an information processing system (IPS) (Newell & Simon 1972:20).

Besides the information processing system, the *task environment* influences the problem solver's behavior as the task environment determines which behavior is appropriate to reach a certain goal (Newell & Simon 1972:55). These thoughts rest on the assumption that the problem solver always thrives to accomplish the goal. Hence, a theory on problem solving needs to account for the demands of the environment (*task environment*) and for the psychology of the problem solver (*problem space*). The *problem space* is defined as the internal representation of the task environment including actions that the individual considered to perform but which were not executed. The postulate of the existence of a problem space is based on the thought that

> "the subject in an experiment is presented with a set of instructions and a sequence of stimuli. He must encode these problem components – defining goals, rules, and other aspects of the situation – in some kind of space that represents the initial situation presented to him, the desired goal situation, various intermediate states, imagined or experienced, as well as any concepts he uses to describe these situations to himself." (Newell & Simon 1972:59)

Thus, the search space includes entities, referred to as nodes, that either represent realizable or imaginable states, i.e. those states that are internal to the problem solver. The links that connect the different nodes represent the actions that change one state into another one. Since in mental activities there are no limitations to the application of operators, unsatisfactory states can be generated. This results in a large scope of the problem space (Newell & Simon 1972:76). The problem solving process itself is described as a search through the problem space as outlined in Newell et al. (1956). From the description of the *task environment* as well as the *problem space* it becomes apparent that their characteristics influence the structure of the search. In addition to these two components, the program, i.e. strategy, that is chosen by the problem solver influences the problem solving process.

The problem solving process follows a general organization, as displayed in Figure 2.4, in which the internal representation is created first (step 1). Then the problem space is selected (step 2). This mechanism is referred to as *input translation* (Newell & Simon 1972:88). The initial step is of high importance as all search activities take place in the created internal presentation. Then a method is selected from the method store. Newell and Simon (1972:88) define method as "a process that bears some rational relation to attaining a problem situation, as formulated and seen in terms of the internal representation". The application of the method, step 3, controls the internal and external behavior of the problem solver. Nevertheless, the action can be halted at any time. If a method is terminated, three options can be taken by the problem solver in step 4. The problem solver may either apply another method, select a different internal representation[29] and thus

[29] Even though Newell and Simon (1972) do not go into detail regarding the mechanisms of creating a new internal representation of the problem space, such reorientation will probably be based on visual or other perceptual clues. Only by re-evaluating the problem description and all additional information that is given a new representation can be created. The additional information may be symbols or objects. Furthermore, hints might be provided to the problem solvers that help to restructure the problem space.

reformulate the problem, or the problem solver may abandon solving the problem altogether. In some problem solving encounters, sub-goals may be chosen and pursued first to open up new solution possibilities before the actual problem is revisited and solved. Importantly, the problem solver can only execute one program at a time (Newell & Simon 1972:89). This observation stresses the importance of the memory system within the IPS since all untried solution possibilities and methods are stored there.

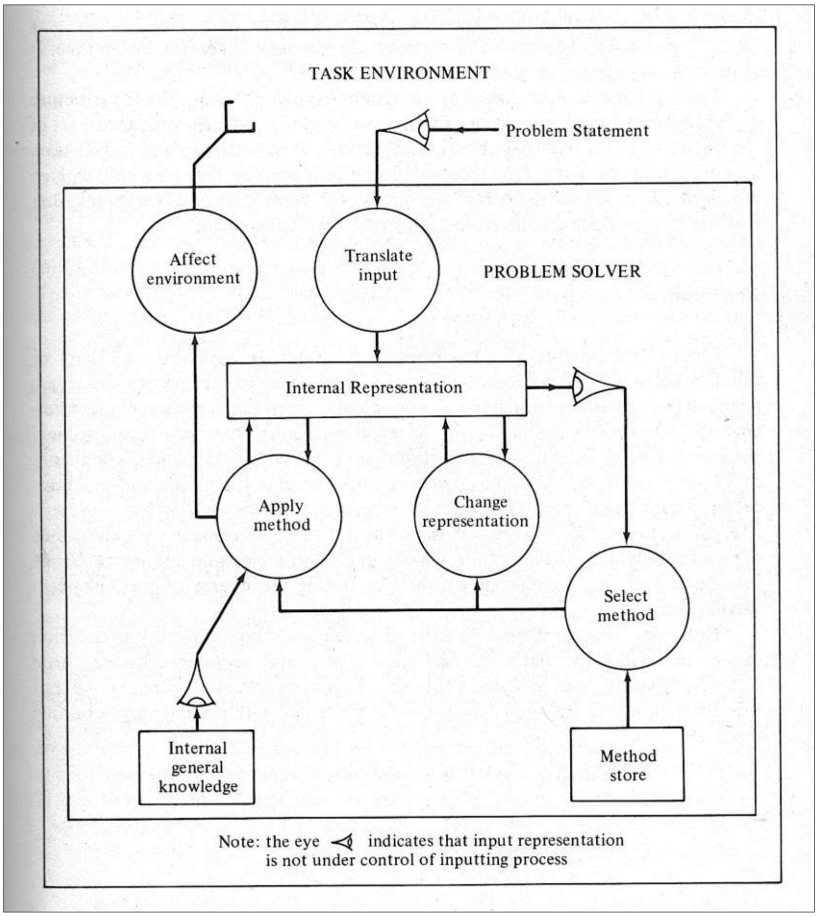


Figure 2.4: General organization of the problem solver (Newell & Simon 1972:89).

The process summarized above can be represented in the basic circuit of "select a goal → select a method → evaluate the results → select a goal again" (Newell & Simon 1972:90). The selection of a method is based on the problem description because the method in itself is not bound to a problem situation. The strategy is selected by the process of interpretation of the information that is encoded in the task environment and the transformations contained in the problem space. Since a problem space is likely to include more information than is necessary for the application of one method, several candidate methods can be selected during the interpretation process. Various methods have been described to be used by human problem solvers. A selection of the most frequently mentioned ones are described in detail in 2.3.

2.2.2.2 Dietrich Dörner – human problem solving as a search through the 'Realitätsbereich'

In his theoretical proposal Dietrich Dörner focuses on the different features of the mental problem representation that effect the selection and appropriateness of the adapted problem solving process. He argues that those selections have a main impact on the success of the whole problem solving process. Dörner's theory is in line with Newell and Simon's (1972) basic concept of problem solving as information processing, i.e. problems are solved in a process of transformation of states. The mental space in which the transformation process takes place is called *'Realitätsbereich'* (space of reality) (Dörner 1987:15). Dörner (1987) specifies that this 'Realitätsbereich' contains *'Operationen'* (operators) and *'Sachverhalte'* (states). This specification shows that the concept of 'Realitätsbereich' is generally equivalent to Newell and Simon's (1972) problem space. But as Dörner (1987) is more specific with regard to the nature of the different states and it is difficult to say if the two theoretical concepts are equivalent on the specific characteristics of states, Dörner's German term 'Realitätsbereich' will be retained.

Contrary to Newell and Simon (1972), Dörner (1987) argues that the nature of the 'Realitätsbereich' as well as the appropriate problem solving processes depend on external as well as internal factors. The appropriateness of the problem solving process is dependent on the nature of the barrier that is being faced by the problem solver and the nature of the 'Realitätsbereich'. The later one, on the other hand, is dependent on the problem solver's goal, such as finding the fastest solution as compared to learning how to solve a specific class of problems. Hence, the features of the states and the operators in the 'Realitätsbereich' constitute its character (Dörner 1987:17). Dörner (1987) identifies five features of states and five features of operators. Those features and their influence on the nature of the 'Realitätsbereich' are summarized in Table 2.3 and Table 2.4.

Feature of states	Influence on the 'Realitätsbereich'
Complexity	A high degree of complexity may call for additional methods to reduce complexity, e.g. abstraction, combination of elements, reduction.

Dynamics	Dynamic situations[30] force the problem solver to act under time pressure and make informed guesses about future events.
Connectedness	If different states are highly connected, the problem solver needs to consider which consequences his actions with regard to one state have on the other states.[31]
Transparency	If few elements constituting a situation can be directly understood or observed, transparency is low and thus effects of actions might not be easily predicted.
Amount of free components for object construction	A high amount of components that can be integrated into the problem solving process fosters the problem solving process.

Table 2.3: Overview on features of states and how they influence the nature of the 'Realitätsbereich' (Dörner 1987:18-21).

Feature of operators	Influence on the 'Realitätsbereich'
Wirkungsbreite (impact[32])	If an operator can change many states, its impact is considered very high.
Reversiblity	If actions are easily reversed, i.e. if the initial state can be easily restored, the problem solver can engage in extensive trying in the world.[33]
Größe des Anwendungsbereichs (scope of application)	If many restrictions are bound to the application of an operator, its scope of application is very low. A low scope of application forces the problem solver to predict the influence of his actions on other states and the problem solver might need to formulate sub-goals to meet prerequisites for the application of a certain operator .
Wirkungssicherheit (certainty of effect)	If the application of an operator under certain circumstances leads to one and only one result, the certainty of effect is highest.
Material and temporal costs	If the energy and the time that is needed to apply an operator is considered too high in relation to the assumed effect, this operator should not be applied.

Table 2.4: Overview on features of operators[34] and how they influence the nature of 'Realitätsbereich' (Dörner 1987:21-23).

[30] Dörner (1987) distinguishes dynamic from static 'Realitätsbereiche'. If a 'Realitätsbereich' changes only by the problem solver's actions it is called static, e.g. a game of chess. But if the 'Realitätsbereich' changes without the problem solver taking action it is classified as dynamic, e.g. politics.

[31] The analysis of resulting effects is called "Nebenwirkungsanalyse" (Dörner 1987:20).

[32] The translations are provided by the author of this thesis since no translations by Dörner himself could be found. They are not literal but content based.

[33] Dörner (1987:22) calls this behavior "spielerisches Probierverhalten" (playful probing).

[34] Dörner (1987) distinguishes between operators and operations. The former term denotes general schemas of actions transforming state A into state B whereas the later denote the actual performance of actions to transform A into B (Dörner 1987:15). Furthermore,

Dörner (1987) describes the different features of states as well as operators in so much detail because the theoretical knowledge about the various influences enables researchers to formulate claims about a general psychology of problem solving (Dörner 1987:25). He argues that if researchers know about the nature of the 'Realitätsbereich' by describing its states and the system of operators therein, the scope of validity of the respective findings can be assessed objectively. This objectivity can be assumed because the problem solver mentally wanders through the 'Realitätsbereich' getting from one state to another one without knowing the way between them. Newell and Simon (1972) associate this mental activity closely to the problem solver's memory structures. Dörner (1987) also mentions working memory and long term memory adding to it the sensory storage capacity[35]. The later one contains all sensory input, i.e. impulses from the nervous system. Its capacity is large but the information is only accessible for a very short time, i.e. about half a second (Dörner 1987:28).

With regard to the mental activities that the problem solver applies when he wanders through the 'Realitätsbereich', Dörner (1987) distinguishes between mental operations and heuristics. *Mental operations* are not important as processes themselves but they rather appear as parts of a cluster of mental processes (Dörner 1987:45). Those mental processes are *processes of change* in which known information is changed into new information or hypotheses, *processes of control*, and *processes of goal explication* (Dörner 1987:38). These categories of mental processes contain the elementary processes of logical deduction, analogical deduction, abstraction, comparing, and classification[36].

The analysis of think aloud protocols showed that the problem solver tends to engage in unguided search right after encountering the problem (Dörner 1987:38). This phase is followed by guided behavior that is introduced by the *process of goal explication*. In the guided search phase, Dörner (1987) observed that *processes of change* are frequently followed by *processes of control*. Thus, Dörner (1987:39) concludes that mental processes do not occur in random order. He assumes that mental processes are ordered, referring to this ordering structure as *heuristics*. *Heuristics* are defined as programs of mental processes that are likely to result in a satisfactory solution to a specific class of problems (Dörner 1987:38). Dörner (1987)

---

operators are often combined into larger elements (macro-operators) which function as an operator system (Dörner 1987:23).

[35] German terminology: "sensorischer Speicher" (sensory storage) (Dörner 1987:28).

[36] Dörner (1987:39) points out that this list is only a selection of possible mental operations. But he does not give reasons for the selection of those specific mental operations. Thus, it is not understandable to the author of the thesis on which criteria the selection is based.

finds the basic ordering unit of *processes of change* and *processes of control* represented in the TOTE unit. Dörner (1987) supports the hypotheses of hierarchical orderings of TOTE units in the operational phase[37] by empirical observations. The data revealed that the test-phase is often succeeded by numerous hierarchically ordered operation phases before the next test phase follows (as illustrated in Figure 2.5).

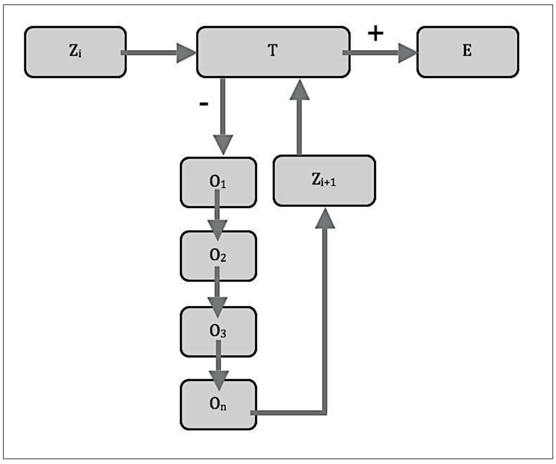


Figure 2.5: Extension of the TOTE-unit as proposed by Dörner (1987:41);
Z_i (state in the 'Realitätsbereich'), T (Test), E (Exit), + (positive evaluation), - (negative evaluation).

Thus, the TOTE unit can be assumed to be one possible ordering scheme of mental operations in a heuristic (Dörner 1987:43). Dörner (1987) explicitly names two heuristics, i.e. the analytical heuristic and the try-and-error heuristic (those will be presented in more detail in section 2.3). Furthermore, he points out that basic schemes of mental operations can be extended by conscious acts of combining mental operations. This active new combination is always necessary if stored heuristics do not yield satisfactory results. Newly derived heuristics that prove to be successful are stored in the so-called *heuristic structure*[38] (Dörner 1987:47). Dörner's (1987) repeated definition of heuristics and his in-depth description of their content expresses the importance of heuristics in the process of problem solving. Following his argumentation, those programs of mental operations are the distinguishing factor between creative and productive thinking as compared to

[37] Remember that Miller et al.'s (1970) proposals were based on theoretical assumptions and research that had been conducted until 1960, which was not as rich as it was when Dörner first published his theory about 16 years later.

[38] Besides the storage for heuristics the *heuristic structure* also contains the capacity for analyzing the characteristics of the 'Realitätsbereich' and of the given task, as well as the control system. The control system determines if the application of a heuristic was successful or unsuccessful (Dörner 1987:47).

reproductive thinking[39]. Only by thinking in new terms, the problem solver will be able to solve problems and overcome barriers that he has not encountered before.

2.2.3 Problem solving with external support – suggestions for a situated cognition approach

The theories on problem solving that have been outlined so far are grounded in the line of thought of *mental empiricism*. De Groot (1969) defines this approach as the "trying out, in the mind, of possible alternatives in hierarchically coordinated, ever-growing cycles" (de Groot 1969:8). He extends this definition to include observations "in a carefully designed and closely controlled experimental set-up" and concludes that "these are obviously important and characteristically human variants" of problem solving (de Groot 1969:15). Researchers in line of situated cognition, in contrast, stress the importance of agent-environment interaction in problem solving. This recent line of research will be briefly outlined in the following paragraphs.

Kirsh (2009) points out that no theory of problem solving with external support exists at this point of writing but he defines the core of a new theory on problem solving in the line of thinking in situated cognition as follows:

> "problem solving is an interactive process in which subjects perceive, change, and create cues, affordances, and large scale structures in the environment, such as diagrams, forms, scaffolds, and artifact ecologies that they work with as they make their way toward a solution." (Kirsh 2009:290)

This quote stresses the importance of an interactive engagement between the problem solver and the environment as the problem space is extended by this interaction. Furthermore, the quote contains the four areas that are central to a situated theory of problem solving, namely *hints*, *affordances*, *thinking-with-things*, and *self-cuing*. Kirsh (2009) argues that classical theories of problem solving, as described in the previous sub-section, fail to explain how the exploratory search process through the problem space proceeds. Furthermore, empirical investigations are based on specific classes of problems, mainly logical reasoning tasks such as chess and math problems. For this class of problems, no special knowledge from outside, i.e. background knowledge, is needed and they are well defined. This makes it easy for researchers to formulate assumptions about the components of the problem space or 'Realitätsbereich' and thus predictions about

[39] The distinction between productive thinking and reproductive thinking is based on Funke (2003). Reproductive thinking denotes the behavior that is needed to cope with known, repetitive tasks. The sequence of actions that is needed is stored in long-term memory and can be applied without further revision.

the problem solving process[40]. In order to explain less restricted problems, Kirsh (2009) argues, that it is necessary to highlight

> "those aspects of problem solving that reveal how much of the machinery of inference, computation, and representation is embedded in the social cultural and material aspects of situations." (Kirsh 2009:265)

Hints are defined as "verbal or nonverbal cue[s] that act like a heuristic bias on search" (Kirsh 2009:291). They are central to the generation of candidates by suggesting possible ways of proceeding to the problem solver. Additionally, they are useful components in the evaluation of possible solutions. This role of hints is also acknowledged in central theories of problem solving. The importance of hints can be seen in the fact that some experiments focus on the nature and the influence of hints that help problem solvers to proceed. As shown in the previous sub-section, it is generally assumed that each problem solving process is driven by some component of generating candidate actions and other components that enable the problem solver to test these actions and evaluate them afterwards. But so far it is unknown, which characteristics make a hint a successful hint and it is not well researched which hints are offered by the environment and to which degree problem solvers make use of them. First attempts to tackle this question have been made in way-finding. In this field, researchers investigate which general strategies problem solvers use when entering for example an unknown building (e.g. Passini 1981; Hölscher et al. 2009) and which environmental hints they take into consideration, such as art work and glass doors (e.g. Frankenstein et al. 2010).

Besides hints that are perceived in the environment there are also cues that are inherent in the object itself, i.e. so-called *affordances*. Norman (2002:9) defines these as "perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used". He stresses the effect that affordances have by suggesting how a thing should be handled. To illustrate this, Norman (2002) gives the example of door handles as there are numerous different ones but we know how to use them by merely looking at them. Kirsh (2009:293) points out that problem solvers will not detect all affordances right away but it depends on the cues that the problem solver encounters during the activity. This leads to the postulate that the environment, which the problem solver engages in, is partly co-constructed by the performed actions. Kirsh (2009:296) claims that the active engagement of the problem solver with the

[40] This high level of transparency was Dörner's (1987) motivation to describe the nature of the 'Realitätsbereich' in much detail.

environment by projecting knowledge[41] or acting to change objects is not included in classical theories.

However, it can be argued that classical accounts of problem solving acknowledge the interaction between problem solver and environment by assuming that the problem space is constantly changing based on performed actions. Although Newell and Simon (1972) as well as Dörner (1987) draw most attention to mental operations, their theories also include changes in the actual world. If the game of chess is considered, the performed chess move will constitute a new state of the world that is defined as the new starting point. Nonetheless, it can be argued that classical theories do not discuss hints or affordances because those are properties of real-life problems rather than of logical problems. However, the later ones are at the center of most problem solving research so far.

It is generally assumed in problem solving theories that internal representations are supplemented by external representations. As Kirsh (2009:297) points out, written sentences, illustrations, numbers, or gestures are listed as external representations. Even though the reduction of cognitive load may be the most prominent reason for externalizing thought, Kirsh (2010) presents further advantages of externalization that foster the process of thinking and reasoning. By converging texts into graphs or other graphical representations, physical constraints can be used to limit the amount of possible options. This process highlights which actions and performances are possible and which ones are prohibited by some rule (either logical, learned, or physical) – these insights might not be gained if thinking and manipulation is performed only mentally (Kirsh 2010:443). Furthermore, externalization of thought makes the state stable and public. It is stable in the sense that one possible state is displayed. Although it can be changed the next moment, it remains unchanged at least at that point of time. Thoughts that are kept in mind are always subject to change when other impressions or thoughts enter the mind. A thought becomes public by being externalized because it can be shared with others and it can be subject to discussion. It stands by itself and is not bound to the thinker anymore, i.e. it becomes intersubjective (Kirsh 2010:449). To summarize, external representations

> "can be operated on in different ways; they can be manually duplicated, and rearranged. They can be shared with other people. Tools can be applied to them." (Kirsh 2010:447)

[41] Kirsh (2009:293) points out that transfer is the result of detecting similar affordances and constraints or invariants.

The interaction with external representations allows for the creation of novel interpretations (Kirsh 1995:64) because even though people can manipulate their internal representations in a way to rearrange objects, internal representations are always private. This means that novel arrangements and thoughts from other minds cannot become part of the thought process. Sterelny (2004) as well as Vallée-Tourangeau et al. (2011) also stress that externalization fosters the perception of structures or relations that are not recognized during mental manipulation. Sterenly (2004:243) describes that this transformation has the general purpose of converting a difficult cognitive problem into a perceptual one that is easier to solve.

However, Kirsh wants to extend this view and argues that people also use objects to think with, e.g. they build real-world models to change arrangements within them. *Thinking-with-things* describes the interaction with physical materials, scaffolds, tools, and structures based on the idea that "thinking is somewhat tied up with the way we encounter and engage the world" (Kirsh 2009:300). This interactive way is essential to thinking because it helps to structure the affordance landscape and it is "a mechanism for extending our perception, action, and regulation" (Kirsh 2009:300). As an example, Kirsh (2009) compares people who encounter an assembly tasks in which the external help, i.e. the manual, is missing to people who are confronted with a tangram task. In both cases the problem solvers will use the different parts as objects to think with. Kirsh (2009) points out that it is necessary to find out which objects are used to think with and to define a thorough theory on thinking-with-things.

So far, two conditions have been described that need to be met. First, all changes that are made can be reversed easily, i.e. the previous stage can be restored or a state can be achieved from which a new path can be taken. Second, all changes that are made and actions that are taken will not have disastrous consequences (Kirsh 2009:299). Assembly tasks are specifically mentioned to meet both of these conditions.

Acting in the real world has numerous advantages as compared to mentally manipulating the current state in order to find a possible solution. One great advantage of real-world manipulation is the precise knowledge about the outcome of the action. In a recent paper, Cowley and Nash (2013) report a study conducted by Nash in which soldiers were faced with a version of the 'Hobbits and Orcs' problem. One experimental group was allowed to use pen and pencil whereas the other experimental group was provided with a papier maché model (Cowley & Nash 2013:191). Unfortunately, no discussion of the effect of an available model on problem solving is reported in this paper. Vallée-Tourangeau et al. (2011) ran a similar experiment investigating matchstick algebra problems. In one group, participants were presented with the problems printed on a sheet of paper whereas the other group of participants was presented with the problems on a magnetic board with magnetized matchsticks (Vallée-Tourangeau et al. 2011:275). The results revealed that participants in the 'interactive group' were more successful in solving these problems highlighting that interactivity fosters insights. Vallée-Tourangeau et al. (2011) suggest that

> "interactivity encourages the rearrangement of the matchsticks which generates configurations revealing new affordances. (…) Manipulation thus leads to opportunities that would otherwise require cognitive effort to identify." (Vallée-Tourangeau et al. 2011:277)

Kirsh (2009) points out another mechanism that needs to be incorporated in a situated theory of problem solving: *self-cuing*. In classical theories, the search through the problem space is driven by the application of operators resulting in the generation of new states. This search takes place in working-memory, thus the search for heuristics, and operators in general, is based on states that are currently active there. In these theories, search processes do not extend to the environment besides those aspects that enter the working-memory through perception. In the account of a situated theory of problem solving the problem solver is assumed to search the environment by intentionally making decisions to look elsewhere. Research in the field of self-cuing is assumed to reveal "new interactive strategies (…) that show unanticipated ways subjects use the environment to shape their problem solving cognition" (Kirsh 2009:302). The central insight that the theory of situated cognition contributes to research on problem solving is summarized to be the observation that "the environment provides organization for cognitive activity, [and] that the world enables and supports such activities" (Kirsh 2009:303).

The objective of researchers working in the framework of situated cognition is to "explain how internal and external control processes work with these supports and organization structures to regulate intelligent activity" (Kirsh 2009:303). Those questions are not well explained by classical theories and they have seldom been the objective of empirical research to date. In order to tackle the question of agent-environment interaction, it is not sufficient to study problem solving in constrained laboratory contexts because as people adapt to the world they live in, their behavior needs to be studied in these contexts as well (Kirsh 2009:302).

Kirsh's list could be extended by Norman's (2002) notion of constraints because as Norman argues constraints are powerful cues for possible actions. Norman

(2002:84-87) distinguishes between physical, semantic, cultural, and logical constraints. Physical constraints are those that limit possible actions. Semantic constraints rely on the problem solver's world knowledge as the situation constraints possible actions. Norman (2002) gives the example of a motorcycle and a rider. The semantic constraint is that the rider needs to sit on the motorcycle facing the front of the motorcycle. Cultural constraints are bound to cultural conventions. In Germany, for example, most roofs have a triangular shape, whereas in other countries roofs are conventionally flat. The category 'logical constraints' subsumes constraints of the general nature, e.g. there is one object piece left and one spot to be filled, hence the piece needs to fit the spot. If the piece does not fit in the spot, the problem solving process needs to restart.

2.3 Methods in human problem solving

So far, the concept *problem* was defined and influential theories about mental processes involved in finding solutions to problems were summarized. As these theories are based on theoretical considerations as well as on empirical observations one focus of investigation was always the identification and description of strategies and heuristics that are used by human problem solvers. The reviewed theories highlight that search and test processes are assumed to be the driving force in problem solving. Dörner (1987) repeatedly stressed that new states are generated out of information provided in the problem description by means of heuristics. Newell and Simon (1972), more generally, state that "ends arise out of the formulation of goals and subgoals" (Newell & Simon 1972:91). In their theory, they further specify that methods by themselves are not bound to problems or certain problem situations but that it is rather the selection of the problem solver that determines which method will be applied first. Thus this concluding sub-section provides an overview on some methods applied to solve a given problem. In empirical studies, human problem solvers have been observed to have a great repertoire of methods at their disposal therefore the presented methods are those that are frequently mentioned in the literature[42].

[42] The selection is based on the methods that are described in the majority of the following works, namely Newell and Simon (1972), Newell et al. (1958; 1960), as well as introductory literature on psychology and problem solving (Funke 2003; Mietzel 2006) and cognitive psychology (Anderson 2005).

The *recognition method* is introduced as one universal method of problem solving by Newell and Simon (1972:95). This method can only be applied when the answer to a problem is stored in long-term memory and its retrieval is evoked when the problem description is interpreted. It can be described as a universal method because it can be applied to all kinds of problems since it is not sensitive to details of the problem description. Newell and Simon (1972) propose that this method is often combined with the *reduction heuristic*. By this combination, general heuristic problems are reduced into a number of sub-problems until one of them can be solved by means of the recognition method.

If there is no solution to a problem stored in long-term memory, a new answer needs to be generated. The *generate-and-test method* is a likely one to yield a successful solution. The information that is available from the problem description is used to generate numerous potential solutions. Each of these is thereafter tested for its adequacy as a member of the set of actual solutions. According to Newell and Simon (1972:95-97), the method has two weaknesses. First, the generation and test process work independently from each other, which implies that partial solutions are not used as new input for the generation process. Second, the method is time consuming depending on the position of a member within the set of actual solutions.

The *generate-and-test* method is the heart of the *heuristic search*. In this heuristic the input elements are not restricted to the problem description, instead all elements contained in the problem space are available as input for the generation process. Generation and test processes are assumed to be combined in the following general cycle: selection of an element form the problem space → selection of an operator → attempt to produce a new element of the problem space → test whether it is a solution to the problem (Newell & Simon 1972:101).

If a generated solution is accepted as a member in the set of actual solutions, it is stored for later use. If it is not accepted, it is stored as an 'untried-problem' and may be tried again with another method later. The cycle ends by either one of the following three choices: the application of operators on a current element is continued, the current element is replaced by a new element, or the solution-path[43] is abandoned altogether in which case attention turns back to an untried sub-problem.

[43] A solution-path is defined as a sequence of expressions belonging to the set of original solutions (Newell & Simon 1972:76).

A number of methods are described that are part of those more general heuristics. These methods are based on the general assumption that sequences of operators can be defined in terms of 'solution-action-sequences' (Newell & Simon 1972:76). If the selection of potential operators is based on information that is encoded in the current state, the problem solver *works forward* in the sequence. In these cases, the selection of operators is combined with evaluative information on the plausibility of the possible operator before it is tested (Newell & Simon 1972:119). The method of *working backwards* starts form the other end of the *solution-action-sequence*; namely with the outcome of the process, i.e. the desired goal state. By applying this method, the problem solver interprets the observable or described goal state and extracts information to generate a list of possible operators. This list includes only operators that are assumed to change the initial state to correspond to the intended goal state. This method involves much planning because a plan of action is mentally created and evaluated before it is executed (Funke 2003:64).

Newell et al. (1960) describe *means-end-analysis* as the most powerful heuristic observed in the LTS. It is based on the *reduction heuristic* in which a goal is divided into a number of sub-goals. These sub-goals (also called *means*) are characterized by two features. First, they are only temporary and they can be reached by the application of a known operator (*end*) (Anderson 2005:256). Second, if a sub-goal is defined that has these features solving, this sub-goal takes precedence over the initial goal. By solving the sub-goals, the problem solver continuously gets closer to the solution of the initial problem. Miller et al. (1970) provide a very good description of this heuristic by summarizing the theoretical questions that the problem solver is faced with when applying *means-end-analysis*. They describe the process of applying *means-end-analysis* as follows:

> "'I want to get from A to B, but I do not know how. What is the difference between what I have and what I want to get? The difference is D. How can I reduce D? Operator T will reduce D, but I do not see how to apply it. Transform A so that T will apply to it. Now apply operator T and get a new object A′. The new problem is to get from A′ to B, but I do not know how. What is the difference?' And so the means end analysis continues." (Miller et al. 1970:189)

Another method that is similar to means-end-analysis is *hill climbing*. However, those two methods represent two distinct perspectives on the overall problem solving process. Problem solvers who select the hill climbing method evaluate the potential new state with regard to its resemblance to the goal state. Only if the new state resemble the goal state more closely than the previous state, the selected operator is applied. In contrast to problem solvers who accept to approach the initial problem by defining intermediate problems and solving those, these problem solvers expect to get closer to the goal state by every step. They do not apply operators that are evaluated to create assumed detours (Anderson 2005:256).

A method that is frequently discussed and that is at the heart of the presented theories on human problem solving is the *planning method*. Newell et al. (1960), for example, observed that it was a very effective method in their Logical Theorist system. Hayes-Roth and Hayes-Roth (1979) proposed a model that specified the mental processes involved in planning. Miller et al. (1970:190) provide a more general description of the method. They assume that by applying this heuristic the problem solver consciously selects certain details of the problem description whereas others are disregarded. By this selection, a simpler version of the problem is defined and a solution is generated for this one. If the simpler problem is solved, the missing details are integrated into the problem representation. The selected strategy is then applied to solve the more complicated problem. As all of these processes happen mentally, planning involves a high cognitive load for the problem solver.

Another method that has been widely studied in experimental settings as well as in real-word scenarios and in scientific reasoning is that of *analogy*. In the process of identifying analogies, elements of a source (also called base) are mapped onto the target problem. If this mapping is successful, the problem solver can assume that the operators that were successful in solving the source problem are also appropriate for solving the current problem (Anderson 2005:251). This description already highlights that analogies can only be applied successfully if the appropriate source problem is selected. Although many people would agree to use analogies on a daily basis, it has been observed that participants rarely notice and apply analogies in experimental settings (e.g. Gick and Holyoak 1980). If participants use this method, they are most often guided by superficial similarities of elements instead of relations between elements. This mismatch between elements often results in erroneous transfer, which does not yield successful solutions (Anderson 2005:253). In a recent project, Taatgen (2012) modeled the solution process on different control tasks, namely task switch, N-Back, and the Troop task to show that all processes are based on small production units. The comparison of the computed solution processes for these different tasks revealed that the solution is derived by either single comparison or single atomic actions. Based on these findings, Taatgen (2012) argues that the generic components of comparison and atomic action are reused by the problem solver when constructing an action plan for a similar problem. Thus, he proposes that it is not the action-sequence that is transferred in analogies but rather the individual components. This finding may imply that analogies are used more often than previous experiments suggested.

In tasks in which the problem description can be represented by actual objects thinking-with-things may be applied. In these cases, the real world is the model in which different configurations are tried and potential solution-paths are created. Vallée-Tourangeau et al. (2011) observed participants who solved matchstick algebra problems by manipulating magnetized matchsticks on a magnetic board (Vallée-Tourangeau et al. 2011:275). Participants were observed to place their fingers on the matchsticks, picking one up and holding it for some time, and moving matchsticks into new positions. Vallée-Tourangeau et al. (2011:277) interpret the last action as instances of "physically testing ideas before placing them back to their original position". Even participants who were shown the equation printed on a piece of paper were observed to

> "move their finger across the printed equation as if to guide or focus thought, often using their finger to represent a matchstick, mimicking rotations and movements to aid visualisation and test spatial configurations." (Vallée-Tourangeau et al. 2011:275)

The observation that participants used their fingers to represent matchsticks supports a claim put forward by Kirsh (2009). He argued that problem solvers create models, i.e. representations, of the elements contained in a problem description if objects in question cannot be used for these manipulations. By manipulating the representations of the elements, they can try different potential methods without too much cognitive cost.

2.4 Conclusion

This chapter provided an introduction to the research field of problem solving by starting out with a definition of problem solving and outlining two influential theoretical frameworks of human problem solving. These traditional approaches understand problem solving as a search through the problem space in which new states are continuously created and checked for their validity as the final goal state. These manipulations proposed to be mainly mental in nature, i.e. executed in the problem solvers' mind. To present different perspectives on human problem solving, a more recent approach was introduced as well. This approach stresses that problem solving is accomplished in an interactive process between the human and the environment. Researchers in situated cognition argue that people actively shape the world to support their problem solving activity either in search for new information concerning a current problem or in the form of reminders for future actions. Both perspectives, traditional and situated, emphasize the search process through a problem space although the scope of the problem space is defined differently. Thus, the theoretical framework in this thesis is based on the assumption that problem solving takes place in a problem space which is defined by old and new states.

In this chapter assumptions have been reviewed regarding the structure of the search process, i.e. the structure of the involved mental processes. Traditionally these structures have been investigated on well-defined, logic-based tasks. This approach is extended in this thesis by investigating the nature of the verbalized mental processes and their structure in an assembly task, i.e. a less logical based but rather physical manipulation task. At different times throughout this chapter, it was outlined that many studies investigated mental processes involved in problem solving by analyzing verbal protocols, such as think aloud recordings. This traditional and well-established approach will be adopted in this thesis. Therefore, the next chapter introduces the methodology of verbal *protocol analysis* and an extension of this content-based analysis by in-depth linguistic analyses as proposed in *Cognitive Discourse Analysis*.

3 Verbal protocol analysis and Cognitive Discourse Analysis

3.1 Introduction

> "The central problem which cognitive scientists face in studying thinking is that *thinking cannot be observed directly by people*." (Ericsson 1999:425)

This introductory quote highlights one of the main obstacles in the endeavor to understand, describe, and model human cognition. Over the past decades, numerous new techniques have been developed to study traces of human mental processes, such as brain activation and eye movements (for an overview see e.g. Funke & Spering 2006). Protocol analysis is among the more traditional methodologies since it has been used from the beginning of the cognitive revolution[44] (e.g. de Groot 1965). While recordings of brain activity constrain the participant's ability to move, recordings of verbalization interfere less with the

[44] The period defined as *cognitive revolution* in the human sciences is characterized by the goal to study the human mind taking contextual and physiological factors into consideration. Its beginning dates back to the mid 1980s (Sinha 2007).

participant's physical space. However, one of the arguments frequently raised against the methodology of recording verbalization during task performance is the potential danger of reactivity, i.e. the mutual effect of verbalization on task performance. Studies that investigate this effect directly reported different results but these results are difficult to compare because those studies investigated performance on different tasks and providing different instructions for verbalization. The first meta-analysis of studies reporting performance measures on think aloud conditions and silent conditions (Fox et al. 2011), on the other hand, provides evidence that the task to think aloud did not tend to affect performance (see section 2.2 for a detailed discussion).

Besides potential interference on primary task performance, one basic question concerning verbal protocol analysis is how conscious mental processes are (e.g. Sternberg & Sternberg 2012:177). Some researchers caution that people have little access to higher order mental processes, such as decision-making, and thus doubt the insights that can be gained by verbal protocols (e.g. Nisbett & Wilson 1977). However, other researchers believe that people can "give verbal expression to those thoughts that spontaneously emerge in attention during the generation of the solution" (Ericsson 2006:228). Ericsson and Simon (1993) introduced a *model of verbalization of thinking* that proposed specific assumptions about which processes can be recorded and by which procedure they need to be elicited. This model will be introduced in the second section of this chapter (content-based protocol analysis); strengths of this approach are highlighted and limitations are discussed.

As think aloud protocols allow for a sequential observation over time, the methodology is widely used in problem solving research (e.g. de Groot 1965; Newell & Simon 1972; Chi et al. 1989; Siegler & Stern 1998; Ericsson et al. 2004; Bilalić et al. 2008). In most studies, verbal protocols are analyzed by applying *protocol analysis*. This methodology, as proposed by Ericsson and Simon (1984; 1993), is described in section 3.2.3 and its application is exemplified by a review of selected studies. The third section of the chapter introduces studies that extend Ericsson and Simon's method by combining the proposed content analysis with the investigating of the linguistic structure of verbal protocols. This section focuses on the introduction of the methodology of *Cognitive Discourse Analysis* that is applied in this thesis.

3.2 Content-based protocol analysis

3.2.1 Ericsson and Simon's model of verbalization of thinking

In 1984, Ericsson and Simon published their seminal work on "verbal protocols as data". In their book, they specified a theory of human information processing and the use of verbal protocol analysis to analyze recorded verbalizations of these processes. Cognitive processes were defined as a "sequence of internal states successively transformed by a series of information processes" (Ericsson & Simon 1993:11). Ericsson and Simon (1993:xxxv) argued that these internal states, i.e. products of perception and retrieval from long-term memory, can be verbalized when they reach attention.

Their proposed *model of verbalization of thinking*[45] was based on four main assumptions. The first assumption addressed the theory of human problem solving behavior. In line with the theory that the human problem solver represents an information processing system, as proposed by Newell and Simon (1972), Ericsson and Simon (1993) assumed that

> "the subject's behavior can be viewed as a search through a problem space, accumulating knowledge (not always correct) about the situation as he goes." (Ericsson & Simon 1993:263)

Second, they assumed that current problem states could be changed by the application of operators. Ericsson and Simon (1993) proposed that all likely operators are held in short-term memory (STM) along with newly created states. Third, the participant's verbalizations correspond to the information that is currently held in STM and recently acquired information. This leads to the fourth assumption in which Ericsson and Simon (1993:264) specified that information recorded in verbal reports

> "consists primarily of knowledge required as inputs to the operators, new knowledge produced by operators, and symbols representing active goals and subgoals that are driving the activity." (Ericsson & Simon 1993:264)

Their assumptions allowed for precise expectations regarding the nature of heeded information, i.e. operators, knowledge states, goals, and subgoals. Their model predicted that different knowledge states that are represented in the problem space are encoded in current-state information, future-state information, and as

---

[45] Different labels are used to refer to the model, e.g. *model of verbalization of thought* (Ericsson 2006:237) or *model of verbalization of thinking* (Fox et al. 2011:337).

probable-state information (Ericsson & Simon 1993:223). These specific assumptions lead to the general proposition that the kind of information which is verbalized depends on the memory system from which the information is retrieved.

In their model of verbalization, Ericsson and Simon (1984 (1993)) distinguished between information that is verbalized at the moment in which it is attended to and information that is retrieved from long-term memory (LTM). Information that is currently attended to is assumed to be held in STM. They proposed that the STM has a limited capacity and is non-permanent[46]. Long-term memory, in contrast, was assumed to have a large capacity for storing information units and to present a permanent storage (Ericsson & Simon 1993:11). This assumed memory structure is important because the model of verbalization proposed by Ericsson and Simon (1993) rests upon these structures. More specifically, they proposed "that any verbalization or verbal report of the cognitive processes would have to be based on a subset of information held in STM and LTM" (Ericsson & Simon 1993:12).

Level 1 and *level 2* verbalizations were assumed to elicit information at the moment of thinking. Ericsson and Simon (1993:16) proposed that some information can be verbalized in the same form in which it is heeded; this is referred to as *level 1* verbalization. They argued that this is the case if information is already verbally encoded. In verbalization at *level 2* the heeded information is not in verbal code and thus needs to be recoded before it is verbalized. Importantly, Ericsson and Simon (1993:18) argued that the sequence of heeded information is not changed in *level 1* and *level 2* verbalizations. Verbalizations at *level 3*, in contrast, involve intermediate scanning or filtering processes in which all retrieved information is evaluated with regard to the appropriateness of being verbalized. The processes of scanning and filtering are activated if selected information needs to be verbalized. Therefore, Ericsson and Simon (1993:18) argued that *level 3* verbalizations represent changed sequences of heeded information.

Based on this distinction, Ericsson and Simon (1993) characterized different kinds of verbal reports in terms of the time of verbalization and the relation between heeded and verbalized information. In their *model of verbalization of thinking* they distinguished between concurrent verbalization, retrospective reports, and probing. *Concurrent verbalization* is defined as "verbalization of task-relevant thoughts generated between the start of a primary task and the completion of the associated task" (Fox et al. 2011:321). Ericsson and Simon's (1980 (1993)) model

[46] For a later discussion on Ericsson's concept of working memory see Ericsson and Kintsch 1995.

predicted that for verbalizations which are recorded while the information is attended to, a one-to-one mapping between heeded and verbalized information can be assumed. The framework predicted that this is the case in *think aloud reports* and recordings of *talk aloud*. Talk aloud represents *level 1* verbalizations, whereas think aloud protocols represent *level 2* verbalizations. Importantly, Ericsson and Simon's (1984 (1993)) model assumed that verbalization may be incomplete because verbalization is secondary to task performance. They stressed that

> "these verbalizations reflect states of heeded information and do not describe the details of information nor why that particular information was heeded. Only the end product of perception and retrieval from long-term memory, those that reach attention are verbalized." (Ericsson & Simon 1993:xxxv)

Ericsson and Simon (1993:252) assumed that concurrent verbalizations are specifically likely to be incomplete in situations in which participants read texts or attempt to understand problem descriptions and in situations in which extensive cognitive ability is needed to perform the primary task.

Besides eliciting verbalization during performance, verbal reports can also be elicited after the task-directed processes are completed, i.e. in *retrospective reports*. If participants are asked to state whatever they remember about their task performance, Ericsson and Simon's model (1980; 1993) predicted a one-to-one mapping between information that is still active in STM and verbalized information. The model predicts that this is only the case if retrospective reports that are elicited immediately after the task is completed and the instructions are phrased as proposed. However, the framework also predicted that some information may need to be heeded from LTM as well because of the limited storage capacity of STM. Ericsson and Simon (1993) cautioned that this retrieval from LTM may cause potential problems of fallibility and separating information. They argued that this problem may especially arise in reports in which participants have completed a number of similar problems. In these cases, participants may have adapted a general procedure across consecutive trials and report this information instead of information on the different trials. Furthermore, participants may not remember which strategy they used in which trial. Ericsson and Simon (1993:19) argued that these limitations could be reduced by instructing participants "to report everything you can remember about your thoughts during the last problem". Nonetheless, researchers need to be aware that the information that is reported may represent a general strategy rather than the activated processes in individual trials.

In sum, if researchers are interested in the participants' mental processes in individual trials, they should record participants' verbalized thoughts during task performance, i.e. in think aloud protocols. If researchers are interested in general procedures that participants adopted across a number of trials, this information can be elicited in retrospective reports. However, observing the application and discussion of their framework over the past 30 years, Ericsson (2006) cloncludes that these verbal reports

> "frequently do not contain sufficient detail about the mediating cognitive processes and the associated knowledge to satisfy many scientists. For example, these reports may not contain the detailed procedure that would allow cognitive scientists to build complete computer models that are capable of regenerating the observed performance on the studied task." (Ericsson 2006:224)

Therefore, in a number of studies, in which researchers were interested in specific aspects of task performance, they asked participants to provide reasons instead of reporting whatever they remember. Different kinds of such directed questioning, also referred to as *probing* (Ericsson & Simon 1993:21), are reported in the literature. Some researchers asked directed questions regarding specific aspects of the task performance (e.g. Nisbett & Wilson 1977; Schooler et al. 1993). Other researchers interrupted participants at specific times during task performance and asked them to provide reasons and explanations (e.g. Bartl & Dörner 1998). Ericsson and Simon's (1993) model predicted that the request for specific information activates intermediate processes, such as filtering. These processes are assumed to change the initially heeded sequence of information (*level 3* verbalizations). Therefore, Ericsson and Simon (1993) repeatedly stressed that the predictions that their framework makes about the relation between verbalized thoughts and mental processes are only valid for the elicitation of undirected thought.

In summary, it is important that researchers are very clear about which information they want to elicit, i.e. information that is currently in attention or general information. Furthermore, they need to be aware of the kind of information that can be elicited by the different kinds of verbal reports and potential errors of reporting. The issue of adequacy of elicited information will be critically discussed in the next subsection. Additionally, the controversial discussion on reactivity of concurrent verbalization and performance on the primary task is addressed.

3.2.2 Critical discussion: adequacy and reactivity of verbal reports

Regarding potential protocol invalidity, two different types can be distinguished, namely reactivity and nonveridicality (Russo et al. 1989). According to Russo et al. (1989)

> "a protocol is *reactive* if verbalization changes the primary process (...) (and) a protocol is *nonveridical* if it does not accurately reflect the underlying primary processes." (Russo et al. 1989:760)

Russo et al. (1989) argued that errors of omission, i.e. forgetting, and errors of commission, i.e. fabricating information, are frequent types of nonveridicality. In order to test the nonveridicality of different kinds of verbal reports, Russo et al. (1989) tested four groups of participants on their performance on different tasks[47]. Three different kinds of retrospective verbal reports were elicited to evaluate which kind of information was reported and how this information related to objective performance measures, i.e. solution accuracy and response times. In 'response cued retrospective reports', on the one hand, participants were allowed to look at the solution. In 'stimulus cued retrospective reports', on the other hand, participants were given the original problem only. In 'prompted retrospective reports' participants saw their eye-movement superimposed on the original problem and were asked to provide reasons for fixating on the specific spot and report what they thought while looking there. The elicited verbal reports were analyzed with regard to forgetting and fabrication of information.

The analysis revealed that forgetting was a commonly observed phenomenon in the analysis of retrospective reports. Participants mainly forgot to report perception of stimuli and information about initially drawn inferences regarding relations between task components (Russo et al. 1989:766). Fabrication of information was specific to algorithmic tasks, i.e. numerical and mental addition tasks. The comparison between the different kinds of verbal reports highlighted that fabrication was specifically frequent in retrospective reports in which participants saw the initial problem. Therefore, Russo et al. (1989:766) caution researchers against eliciting retrospective reports by cuing participants with the original problem.

It needs to be stressed that Russo et al. (1989) did not test their measures of nonveridicality with regard to the elicitation of retrospective reports as proposed

[47] These tasks were verbal tasks, numerical tasks, pictorial tasks, and mental addition of 3-digit numbers.

by Ericsson and Simon's (1980; 1993) framework. Nonetheless, the reported findings confirmed that retrospective reports may miss information on inferences that are made at different times during the process of solving a problem, as pointed out by Ericsson and Simon (1980). A study by Siegler and Stern (1998) suggests an alternative interpretation of these findings. In their study, they were interested in the conceptual change of thinking in children. They tested children on mathematical inversion problems and asked them for explanations on how they solved the problem immediately after the solution was given (Siegler & Stern 1998:382). The analysis of these reports revealed that children did not report using a shortcut strategy although they already applied it efficiently. Siegler and Stern (1998) interpreted these findings as evidence for the assumption that strategy discovery was unconscious and that the strategy could be reported only later. This finding highlights the strength of concurrent verbal reports because at the time at which the strategy discovery reaches consciousness it is likely to be verbalized although it might not be understood as a generalized strategy yet. Furthermore, participants do not need to be interrupted in their performance to provide explanations, an activity that focuses attention on the thought processes and takes it away from the primary task performance.

Different studies suggested that asking participants to provide explanations for their actions and solution strategies during task performance facilitates problem solving and results in better performance (e.g. Chi et al. 1989; Chi 1996; Neuman & Schwarz 1998; Ainsworth & Loizou 2003). The effect of concurrent verbalization, i.e. the task to think aloud, on performance of the primary task is not as clear. As studies revealed different effects, the question of reactivity is still controversially debated. Ericsson and Simon (1993) argued that verbalization, if elicited as proposed in their framework, has no effect on basic performance (Ericsson & Simon 1993:85) or the internal structure of the thought processes (Ericsson & Simon 1993:89).

There are a number of studies that focus specifically on the influence of verbalization on task performance. Most of them addressed the diverging findings from previous studies, i.e. verbalization facilitates performance vs. verbalization hinders successful performance, in their introductory section. Three alternative explanations for those findings are mentioned, namely the different nature of the investigated tasks (McGeorge & Burton 1989; Knoblich & Rhenius 1995; Bartl & Dörner 1998), evaluation of different cognitive processes (Schooler et al. 1993), and different instructions of verbalization (Ericsson & Simon 1993; Bartl & Dörner 1998). Based on the differences in tasks as well as instructions for verbalization, which the different studies report, Wilson (1994) stated that "this area is in sore need of a systematic meta-analysis" (Wilson 1994:251). Only recently, Fox et al. (2011) presented such a meta-analysis of 64 articles that reported 94 studies featuring 3,462 participants. They motivate their analysis by the observation that

> "despite widespread use, concurrent verbalization remains controversial because of concerns that it changes the cognitive processes mediating the performance of tasks in comparison to the traditional silent conditions under which psychological phenomena are typically studied in the laboratory." (Fox et al. 2011:317)

To address the potential reactivity of concurrent verbalization, Fox et al. (2011:322) selected studies that reported objective performance measures, such as accuracy of performance and solution time, and included a think aloud group as well as a silent control group. Furthermore, they included only studies that provided information that was needed to extract an effect size. They operationalized their analysis by categorizing four kinds of procedures, two kinds of task types, and three dependent variables. The reviewed papers included reports on explanatory verbal data, think aloud as proposed by Ericsson and Simon's (1980, 1993) model, and directed questions. In some cases, the instructions for verbalization were only insufficiently reported so they could not be categorized. Fox et al. (2011) distinguished between visual tasks, i.e. those tasks in which "the spatial organization of the stimuli was inherent to accomplish the task" (Fox et al. 2011:324), and nonvisual tasks, such as mental multiplication. With regard to the dependent variables, they classified if a study used time limits and cited the reported accuracy and solution times.

The meta-analysis confirmed their expectation that

> "think-aloud procedures resulted in performance that was indistinguishable from that of silent groups, whereas explanatory procedures were associated with higher performance in verbalization conditions." (Fox et al. 2011:329)

No effects of verbalization were revealed with regard to task type (Fox et al. 2011:330). The expected effect of verbalization on performance time was confirmed; if participants verbalized their thought during task performance, they needed more time to complete the task (Fox et al. 2011:331). Fox et al. (2011) summarize their findings by stating that the analysis revealed the effects predicted by Ericsson and Simon's (1980; 1993) model. They conclude their meta-analysis by observing that

> "although the think-aloud procedure has limits and does not assure a complete record of participants' thoughts, it is at present the only

> nonreactive method of collecting the verbalized contents of thoughts while participants focus on completing challenging tasks.” (Fox et al. 2011:338)

In sum, Fox et al.’s (2011) meta-analysis supported the predictions made by Ericsson and Simon’s (1980; 1984 (1993)) model of verbalization of thinking. Thus it can be assumed that the task to verbalize all thoughts that pass through the participant’s mind does not have an effect on performance except for longer times to complete the task.

Although Schooler (2011) acknowledged the approach by Fox et al. (2011), he criticized their study and their findings for three reasons. First, Schooler (2011:347) criticized the selection of the analyzed studies. He argued that the selected studies were not specifically designed for the purpose of testing reactivity of verbalization on task performance. Second, Schooler (2011:347) pointed out that findings which revealed a significant effect of concurrent verbalization on performance were not sufficiently assessed to contribute adequately to the results of the meta-analysis. Third, Schooler (2011:346) argued that Fox et al. (2011) did not discuss the limitations of concurrent verbalization “to capture the ineffable mental processes that span between one thought and the next”.

In a reply to Schooler’s (2011) comment, Ericsson and Fox (2011) stressed the difference between introspections, as discussed by Schooler, and think aloud reports, as proposed by Ericsson and Simon (1993). They point out that the analysis of think aloud protocols “does not involve attempts to arrest transient phenomenological experiences or decompose thoughts into their smallest elements” (Ericsson and Fox 2011:352). Furthermore, they emphasize that Ericsson and Simon’s (1984 (1993)) framework assumed that verbal reports may be incomplete (for more detail see 3.2.1). Ericsson and Fox (2011) also addressed Schooler’s (2011) critique concerning their selection of studies with regard to studies that suggest reactivity of concurrent verbalization and task performance. They argued that the studies referred to by Schooler (2011) were conducted in the framework of ‘verbal overshadowing’. In reference to a meta-analysis of studies conducted in this framework that did not reveal any significant reactivity with regard to concurrent verbal reports, Ericsson and Fox (2011) did not find their results to be contradicted by these studies.

Addressing Schooler’s (2011) critique on the selection of studies, it needs to be noted that only few studies focus on the investigation of reactivity of concurrent verbal reports. A meta-analysis of this restricted sample would not be sound because the sample size would be much too small to draw valid conclusions. Furthermore, the findings of these studies can only be compared if they used the same instructions for verbalization and investigated and reported the same performance measures on comparable tasks. The comparison of some studies that were specifically designed to investigate reactivity of concurrent verbalization highlights that this prerequisite is seldom met. These studies investigated different tasks, such as the Tower of Hanoi[48] (Ahlum-Heath & Di Vesta 1986), dynamic decision tasks (McGeorg & Burton 1989; Knoblich & Rhenius 1995; Dickson et al. 2000), analogical problems (Short et al. 1991), insight problems (Schooler et al. 1993; Gilhooly et al. 2010), analytical tasks (Deffner 1989), or various different tasks (Merz 1969; Franzen & Merz 1976; Russo et al. 1989) (for more detail see Table 3.2 in the appendix). This comparison, that is restricted to only one of the measures that were compared by Fox et al. (2011), highlights the difficulties to compare the reported findings to resolve the controversial discussion. Only one of these studies was specifically designed to test the findings reported in a previous one, i.e. the study designed by Gilhooly et al. (2010) tested findings that were reported by Schooler et al. (1993).

In their study, Schooler et al. (1993) investigated the effect of verbalization on performance in insight and non-insight problems. *Insight problems* were defined as "problem situations that occur unexpectedly following an impasse" (Schooler et al. 1993:166) whereas *non-insight problems* "can be solved in a logical, incremental fashion and do not require any nonobvious approaches to reach the correct solution" (Schooler et al. 1993:174).

Schooler et al. (1993) reported two experiments that investigated the influence of verbalization on insight and non-insight problems. Think aloud was the only kind of verbalization that was tested in these experiments. The results highlighted that participants thinking aloud were less successful on insight problems but equally successful on non-insight problems. There was no difference with regard to performance time. Schooler et al. (1993) concluded that their findings "support our hypothesis that verbalization disrupts critical nonreportable processes" (Schooler et al. 1993:175). Schooler et al. (1993) assumed that spreading activation was among these nonreportable processes and they argued that this process was important in solving insight problems.

Addressing Schooler at al.'s (1993) report, Gilhooly et al. (2010) provided a critical review of their procedure. They pointed out that the difference between the investigated tasks was not only on the nature of insight or non-insight, as defined above, but also between verbal and spatial form; all non-insight problems were

[48] For more detail on the task of the Tower of Hanoi see chapter 2.

verbal in nature whereas at least two out of three insight problems were spatial in nature. Gilhooly et al (2010) proposed that performance on spatial problems might possibly show a stronger effect because of the necessary re-coding from spatial to verbal code. Additionally, Gilhooly et al. (2010:83) pointed out that the wording of the instruction for verbalization might have encouraged reading the instructions more frequently possibly resulting in a more persistent interpretation. Furthermore, participants trained to think aloud were asked to solve a non-insight problem. Gilhooly et al. (2010) speculated that this encounter might have biased those participants towards assuming all other problems to be of the same kind.

In order to address all of these potential confounding factors, Gilhooly et al. (2010) balanced their design by including insight and non-insight problems that were spatial as well as verbal in nature. The test group was instructed to think aloud while solving the task[49] whereas the control group was instructed to solve the task silently. The results revealed a significant effect of verbalization on performance on spatial tasks. Participants who thought aloud performed worse on spatial tasks than participants who worked silently. Contrary to Schooler et al. (1993), Gilhooly et al. (2010:88) did not find an effect of verbalization on success on insight or non-insight problems. Therefore, Gilhooly et al. (2010:92) conclude that "insight problem solving does not depend on processes that are disrupted by verbalization". They argue that their findings rather suggest that the nature of the task results in observable differences between participants thinking aloud and those who solve the problem silently. This effect of verbalization on task performance is in line with predictions made by Ericsson and Simon's model of verbalization of thinking (1993:88).

To summarize, Fox et al. (2011) presented the first approach to assess reactivity of concurrent verbalization on the basis of a large data set and across a variety of tasks. Their analysis highlighted that instructions for verbalization as well as task type resulted in different findings concerning the instruction to think aloud. The conclusions that they draw from their findings, support assumptions made in Ericsson and Simon's model of verbalization of thinking (1980; 1993). In line with these findings, Gilhooly et al. (2010) presented evidence that effects of reactivity, as reported by Schooler et al. (1993), can be attributed to the nature of the task rather than to reactivity. These findings stress the importance of careful wording in task instructions and suggest that a silent condition should be included in the analysis.

[49] Examples for the training phase were similar to those proposed by Ericsson and Simon (1993) and did not represent either insight or non-insight problems (for exact wording see Gilhooly et al. 2010:86).

Based on these considerations, a silent condition is included in the experimental design in this thesis.

The presented findings on potential invalidity of verbal protocols can be summarized by assessing the possibility of reactivity and nonveridicality. Techniques that elicit verbal reports after task completion are generally assumed to be nonreactive, i.e. introspection and retrospective reports. Asking guided questions, i.e. probing, is predicted to be reactive whereas concurrent verbalization is assumed to be nonreactive if the instructions are based on Ericsson and Simon's (1984 (1993)) model. Whereas introspection and retrospective reports may contain memory errors and fabrication, probing and concurrent verbalization are assumed to be nonveridical (see Table 3.1).

Assuming that both, think aloud protocols and retrospective reports, are non reactive, both kinds of verbal protocols were elicited in the experimental studies reported in this thesis. The combination of think aloud protocols and retrospective reports is adopted in this thesis for two reasons. First, think aloud protocols allow for a sequential observation over time whereas changes in conceptualization or strategy may not be completely remembered and reported in retrospective reports. Second, retrospective reports are likely to include generalizations based on the individual processes during task performance (Ericsson & Simon 1993). More generally, a second verbal report is likely to contain aspects that are not mentioned in the other kind of report.

data elicitation method	reactivity	nonveridicality
introspection	no	yes
probing	yes	no
retrospective report	no	yes
concurrent verbalization (talk aloud and think aloud)	no	no

Table 3.1: Overview on measures of adequacy regarding verbal reports in relation to the presented elicitation techniques.

3.2.3 Protocol analysis

Whereas the previous section introduced the general framework proposed by Ericsson and Simon (1984 (1993)) this section describes the methodology of *protocol analysis* which they suggested for the analysis of verbal data collected in their framework. In a later paper, Fox et al. (2011) explicitly state that "the protocol analysis approach should be seen as an accessory of the Ericsson-Simon model of verbalization of thinking" (Fox et al. 2011:337). The scope of the application of the proposed theoretical framework will be illustrated by examples of how the analysis of verbal protocols contributed novel insights on problem solving processes and strategies (see section 3.2.3.3).

3.2.3.1 Instructions for eliciting think aloud protocols

Ericsson and Simon (1993) repeatedly stress that the phrasing of instructions is crucial to ensure the elicitation of level 1 and level 2 verbalizations, i.e. information about the sequences of mental processes that is not changed by intermediate processes (see 3.2.1). Previous studies revealed that the instructions for verbalization determine if unaltered information is verbalized, i.e. representing the trace of thoughts, or if the information is selected and thereby altered (see 3.2.2). Ericsson and Simon's (1993:18) *model of verbalization of thinking* predicted that the instruction to say out loud everything that passes through the participants' mind does not change the sequence of heeded information. However, if participants are asked to provide explanations regarding what kind of information is being retrieved, additional processing time is required. The model predicts that in order to provide explanations, participants link information that is currently in attention to information and thoughts that were in attention previously (Ericsson & Simon 1993:79). This retrieval of selected information is expected to result in the activation of additional interpretation processes. The framework predicts that this procedure changes the structure of the cognitive processes in two ways. First, it involves additional interpretation processes that may not be activated otherwise. Second, the reflective behavior might direct the participant's attention directly on his/her performance possibly highlighting aspects that would not be noticed otherwise. Therefore, Ericsson and Simon (1993) proposed the following instruction to elicit *level 2* verbalizations:

> "In this experiment we are interested in what you think about when you find answers to some questions that I am going to ask you to answer. In order to do this I am going to ask you to THINK ALOUD as you work on the problem given. What I mean by think aloud is that I want you to tell me EVERYTHING you are thinking from the time you first see the question until you give an answer." (Ericsson & Simon 1993:378)

After providing this general instruction to the task of thinking aloud, Ericsson and Simon (1993:378) strongly suggested practicing thinking aloud with participants prior to the experimental session. They argue that practice tasks help to point out the difference between concurrent verbalization and retrospective or explanatory reports. The practice task that is widely used is that of multiplication. Participants are asked "What is the result of multiplying 24×36?". If participants need more than one task to understand the difference between explanations and think aloud, Ericsson and Simon (1993) provide two additional tasks. These tasks ask participants to provide information held in LTM because they are also recommended for training participants to provide retrospective reports (see Ericsson & Simon 1993:378). If the experimenter is content that the participant verbalizes his/her thoughts during the practice task instead of providing explanations, the experimental session and the recordings start.

3.2.3.2 Coding verbal reports

The recorded speech needs to be transcribed to make it durable and subject to further coding on which analyses can be performed, such as content analysis as well as statistical analysis. Ericsson and Simon (1993:278) defined the following desired characteristics of transcripts: a) usability by investigators with different theories and research foci, b) no reliance on theoretical assumptions that are controversial, and c) all information that is necessary for investigating different theories need to be retained. These characteristics illustrate that Ericsson and Simon (1993) propose that the transcription process should not be theory-driven.

However, it stands to reason that transcriptions are necessarily subjective because of the level of detail that is transcribed, e.g. phonetic vs. orthographic transcriptions. Ochs (1979:44) already stated that "transcription is a selective process reflecting theoretical goals and definitions." Contrary to Ericsson and Simon's (1993) view, Ochs (1979) proposed that transcripts should not contain too much information. She rather argued that "transcripts should reflect the particular interests – the hypotheses to be examined – of the researcher" (Ochs 1979:44). Ochs (1979) repeatedly stressed that researchers need to be aware of the filtering process from which transcripts emerge as these transcriptions are the basis for their analysis and thus influence generalizations that are based on them. These considerations stress that the researcher's theory necessarily has an influence on the transcription process starting with the selection of detail that is transcribed. In this thesis, the orthographic level is chosen in the transcriptions including hesitation markers, pauses, as well as elliptic sentences or words to capture the stream of thought as verbalized by participants.

Information that is transcribed in protocols is coded with regard to specific categories. These categories are defined based on the theoretical framework that is adopted in the investigation (Ericsson & Simon 1993:286). These categories are either defined prior to the analysis based on the theoretical framework that is investigated or they are defined post-hoc during the coding procedure. Ericsson and Simon's (1984 (1993)) primary motivation for describing *protocol analysis* was the matching of the recorded verbal behavior to the output of a computer program that performed the same task. Through this matching the assumptions that were made about the mental processes and defined in theoretical terms in the computer program could be evaluated. Similarly, van Someren et al. (1994) proposed that the coding scheme is derived from the proposed model and a theory of verbalization, i.e. "our knowledge of the way in which cognitive processes will be verbalized" (van Someren et al. 1994:123). The coding scheme specifies the processes that are proposed in the model and possible representations of them (van Someren et al. 1994:121). Additionally, van Someren et al (1994) proposed that there are some processes that are not specified in the model but that need to be anticipated in think aloud protocols. These processes are *talking about non-task related issues*, *evaluation of the task or task-situation at a meta-level*, *comment on self*, *silent periods*, and *actions*[50] (van Someren et al. 1994:122). Van Someren et al. (1994) remark that these categories "might be an indication of the level of difficulty of a sub-task or the cognitive load of the subject" in some cases (van Someren et al. 1994:122).

For assigning the coding categories, the transcribed protocol is divided into segments. These segments may correspond to a statement each (Ericsson & Simon 1993:266) or are well defined coding units of a different type (Krippendorff 2004). In most cases, a category is assigned based on the information that is contained in one unit. If this is not possible, Ericsson and Simon (1993:291) suggested that the previous and subsequent segments may also be consulted. However, they stressed that the amount of information that needs to be inferred from previous to subsequent segments influences the subjectivity of the coding; more inferences lead to increased subjectivity. Ericsson and Simon (1993) argued that the greatest possible objectivity is ensured when the coding can be done automatically by a computer because in these cases all underlying assumptions and inference rules need to be explicitly defined in the computer script. Furthermore, the computerized analysis ensures that the defined rules are consistently applied (Ericsson & Simon 1993:270).

If no automatic coding is possible, the training of coders is essential. Ericsson and Simon (1993:291) as well as the other researchers discussing intercoder reliability assessment, point out that coders should be uninformed with regard to the experimental design and the hypotheses that are tested by the analysis. Moreover,

[50] Actions may be a category that is part of the initial model if an action based task is modeled. Possibly the category 'action' was not part of the models tested by van Someren et al. (1994) as they were specifically interested in knowledge acquisition.

the coding scheme should be defined in a way that does not call for many inferences in cases of ambiguity because the extent of necessary inferences influences intercoder agreement (Ericsson & Simon 1993:294). Concerning the importance of training coders, Krippendorff (2004) points out that during training sessions

> "they[51] refined categories, altered instructions, and revised data sheets until the coder felt comfortable with what was expected of them and the analysts were convinced they were getting the data they needed." (Krippendorff 2004:129)

The importance of a close interaction between the researcher and the coders is specifically important in the light of Och's (1979) observation that transcripts and specifically coded transcripts are the basis for later interpretations and generalizations. Therefore, coding validity needs to be assessed in a separate step of the procedure if coding is done manually. Coding validity is assessed by tests of reliability in which the agreement between two or more persons coding parts of the segmented protocols independently from each other is assessed. This assessment may either be based on percent agreement (Ericsson & Simon 1993; van Someren et al. 1994) or other inferential statistics measures such as Cohen's Kappa (van Someren et al. 1994) or Krippendorff's Alpha (Hayes & Krippendorff 2007).

Van Someren et al. (1994:132) showed that percent agreement between coders represents "a more optimistic estimate" than Cohen's Kappa. Lombard et al. (2002:590) agreed that percent agreement is a "simple, intuitive, and easy to calculate" estimate of intercoder reliability but they stressed that it does not account for chance agreement. This is a fact that weakens the informative value of this measure. Cohen's Kappa accounts for chance agreement but it is limited to nominal data. Nonetheless, Lombard et al. (2002:592) observed that it is a widely reported coefficient in psychological science. However, Hayes and Krippendorff (2007) argued that only Krippendorff's Alpha is appropriate to calculate reliability for numerous observers that are freely permutable and in which only data is considered that is generated by all observers. Additionally, they pointed out that Krippendorff's Alpha can be used to calculate reliability for nominal, ordinal, interval, and ratio data (Hayes & Krippendorff 2007:82). Since there is no agreement on one standard index of intercoder reliability, Lombard et al al. (2002:600) proposed to report more than one index.

If one or more appropriate indexes were selected and calculated, the acceptable level of reliability needs to be defined. Regarding the interpretation of Cohen's

[51] Here Krippendorf refers to content analysts and coders.

Kappa values, van Someren et al. (1994:133) stated that "we would, generally speaking, say a Kappa should be above 0.70 in order to have an intercoder reliability that is acceptable". Based on a literature review of reported intercoder agreement measures, Neuendorf (2002:143), in contrast, concluded more generally that "reliability coefficients of .90 or greater would be acceptable to all, .80 or greater would be acceptable in most situations, and below that there exists great disagreement". Lombard et al. (2002:593) extend this statement by observing that "the criterion of .70 is often used for explorative research".

In sum, the presented overview stresses that the coding process needs to be very transparent, i.e. researchers need to be aware that their theoretical assumptions influence choices at the transcription level as well as on the coding level. Furthermore, researchers need to define these assumptions precisely at the beginning of the coding procedure to ensure that they are consistently applied throughout the analysis. If assumptions change or new information needs to be incorporated, e.g. by defining new coding categories, these changes need to be re-integrated into the coding of the entire data sample (e.g. Taylor & Gibbs 2010).

By applying the coding scheme, a coded protocol is generated. In traditional *protocol analysis* as described by Ericsson and Simon (1993) and van Someren et al (1994), the content and structure of the coded protocol is then compared to the proposed model. The a priori proposed model simulates the assumed cognitive processes to perform the task and the simulated traces of mental processes are recorded. This record is then compared to the traces of mental processes in human problem solvers as specified in the coded protocol. Ericsson and Simon (1993) pointed out that much of their work focused on a small number of protocols, sometimes being single case studies, with the aim to "demonstrate that the information in the protocol could be regenerated, approximately, by the simulation model operating on the same task" (Ericsson & Simon 1993:262).

Similarly to Ericsson and Simon (1993), van Someren et al. (1994:117) apply protocol analysis with the aim to construct a matching between the elicited protocols and the proposed model. Concerning the evaluation of the matching process, van Someren et al. (1994:134) specified three possible mismatches. First, the coded protocol contains processes that are not specified in the model and that are not part of the processes mentioned above. Second, the model may specify processes that cannot be identified in the protocols. And third, in cases of procedural models the observed structure of the coded protocol does not correspond to those specified in the model. All of these deviations have certain implications for the model as described in more detail in van Someren et al. (1994:135-139).

Although Ericsson and Simon (1984, 1993) described *protocol analysis* for a specific kind of task, the methodology has not only been used to compare human problem solving behavior to the behavior simulated by a computer script. Rather the elicitation of think aloud protocols and their analysis has been applied in studies conducted in various disciplines, such as psychology (e.g. Chi et al. 1989; Knoblich & Rhenius 1995; Siegler & Stern 1998; Bise 2008; deSouza et al. 2008), linguistics (e.g. Farrington-Flint & Wood 2007), document design (e.g. Krahmer & Umelen 2004), or pedagogy (e.g. Kucan & Beck 1997; Green 1998; Leow 2005). In these studies, protocol analysis has either been used for testing hypotheses and establishing evidence for models and theories of mental processes as described so far or they have been used to generate explorative hypotheses (Crutcher 1994). To illustrate the variety of findings that have been generated by the application of *protocol analysis*, a selection of studies in the field of problem solving is presented in the next section.

3.2.3.3 Application of protocol analysis in problem solving research

Researchers have investigated various aspects of human problem solving, such as the kinds and causes of errors (e.g. Schoenfeld 1985), mechanisms underlying human reasoning, and understanding processes driving human performance on specific tasks (e.g. Ericsson et al. 2004; Bilalić et al. 2008). This thesis aims at identifying and describing the structure of problem solving processes involved in solving an assembly task. There are only few studies investigating problem solving in this kind of real life task. Therefore, the focus in this section is on studies that investigated the structure and nature of problem solving processes in a variety of tasks by means of *protocol analysis*.

In the early years of the application of *protocol analysis*, it was mainly used in the way proposed by Ericsson and Simon (1980; 1984 (1993)), i.e. for informing and testing models of problem solving (e.g. Newell & Simon 1972; Hayes-Roth & Hayes-Roth 1979; Hayes-Roth 1980; Kintsch & Greeno 1985; Dörner 1987; Blackman 1988; Habel 1988). The content analyses provided information about the use of heuristics and strategies, correct and false associations as well as on actions. These insights contributed to a better understanding of human problem solving behavior.

Hayes-Roth & Hayes-Roth (1979), for example, investigated the strategies that were used by participants in a task in which errands needed to be scheduled given temporal and spatial constraints; this task is similar to the 'plan a day' task (e.g. Funke & Spering 2006:681). They distinguished between top-down and bottom-up planning. Top-down planning processes were defined as those in which participants planned on a general level, i.e. which areas need to be covered, before they made considerations at a more detailed level, i.e. the exact sequence of errands. Bottom-up planning was defined when planning at the detailed level preceded planning at a general level or alternated with it (Hayes-Roth & Hayes-Roth 1979:303). The analysis of think aloud reports highlighted that, against their expectations, participants did not engage in bottom-up planning. The *protocol analysis* rather revealed a 'heterarchical plan structure'. In this structure participants made different decisions about their plans, such as how to approach a problem, how to evaluate the derived plans, and which mental resources to contribute to the execution of these plans (Hayes-Roth & Hayes-Roth 1979:305). A later study (Hayes-Roth 1980) confirmed these results and further suggested that problem solvers adapted their behavior to the problem structure. If the instructions posed strict time constraints, participants tended to use a bottom-up planning procedure. But if only minimal time constraints were imposed, participants used top-down and bottom-up approaches to solve the given task.

Kaplan and Simon (1990) used protocol analysis to study the nature of insights. They assumed that insights were gained if the problem solvers searched for a new representation of the problem space. The results supported their assumption that problem difficulty could be attributed to search in the wrong problem space and the inability to notice the need to restructure the problem space (Kaplan & Simon 1990:393). Additionally, the analysis revealed individual differences regarding the number of things that were noticed and the nature of these cues. Furthermore, recent studies employed protocol analysis to study decision-making (e.g. Cokely & Kelley 2009), transfer (e.g. Chrysikou & Weisberg 2005), and generation of new solutions (e.g. Gilhooly et al. 2007) in the area of problem solving.

In a very recent study, Steffensen (2013) took an explorative approach to investigate the strategies employed by people working on a real-life problem within their natural environment, i.e. an office in this specific case study. Steffensen (2013) reported that he used think aloud protocols to explore

> "how human beings engage in collective and socio-culturally enhanced problem-searching, problem-finding and problem-solving activities in a self-organised problem space." (Steffensen 2013:196)

In order to analyze this data systematically, Steffensen (2013) proposed the methodology of *Cognitive Event Analysis* that extends *protocol analysis* by identifying cognitive events and analyzing their structure along an event trajectory. Steffensen (2013:201) defined an event trajectory as "the path taken by the DCS[52] as it moves through an infinite problem space in a way that yields results". Therefore, Steffensen (2013) proposed that the path of mental processes that yield the final observable result can be visualized by means of such a trajectory. The reported results highlighted that the trajectory of the case study involved processes of reframing the problem representation and highlighted the moment of insight. Although Steffensen (2013) did not refer to traditional theories on insight problem solving, he stressed that the moment of insight was the turning point in the search for an acceptable solution. The analysis revealed that the observed problem solvers engaged in a problem solving cycle consisting of problem-finding, problem-solving, and cognitive events, such as reframing the problem representation (Steffensen 2013:204). This finding was similar to Kaplan and Simon's (1990) observation that the process of restructuring the problem representation was an important component for gaining necessary insights to derive the correct solution. However, in contrast to Kaplan and Simon (1990), Steffensen (2013) described a chaotic behavior that was characterized by changing hypotheses about actions that could be taken and potential solutions. The trajectory revealed that the generation of these hypotheses did not follow a predictable structure but occurred at various times along the trajectory. Based on these new observations, Steffensen (2013:218) concluded that "the method (…) allows for detailed scrutiny of *what happens* as cognitive results are brought forth".

These examples were selected to illustrate how content-based *protocol analysis* has contributed to the description of mental processes that result in differences in observable and recordable problem solving behavior. The differences with regard to strategies, as revealed by Hayes-Roth and Hayes-Roth (1979) and Kaplan and Simon (1990), could arguably not have been revealed by performance measures alone. To gain further insights on the nature of the activated and reported mental processes, some researchers have extended content-based analyses by investigating the linguistic structure of recorded problem solving processes.

3.3 Linguistic analysis of verbal protocols

There is no publication to date that summarizes different studies focusing on the analysis of linguistic features in think aloud protocols as expressions of mental processes. Therefore, the first section provides an overview introducing pub-

[52] DCS stands for Distributed Cognitive System and denotes the human problem solver who interacts with the environment.

lications in German by Roth (1985; 1987) and Bartl and Dörner (1998) as well as publications in English by Caron-Pargue and colleagues (e.g. Caron & Caron-Pargue 1987; Bégoin-Augereau and Caron-Pargue 2009). The second section focuses on the presentation of the methodology of *Cognitive Discourse Analysis* that is applied and refined in this thesis.

3.3.1 Linguistic representation of cognitive style and mental processes in think aloud protocols

3.3.1.1 Linguistic representation of cognitive style in think aloud protocols

Dörner et al. (1983) presented a comprehensive analysis of participants' performance and the content of their think aloud protocols elicited while working on a complex decision task. In the 'Lohhausen task' participants needed to adapt the role of the mayor of a small city and their task was to ensure that the city prospered. All information that the participant considered necessary to make decisions, such as the number of people working in industry, was provided by the experimenter who was present during the experimental session (Dörner et al. 1983:107). The analyses revealed systematic differences between successful and unsuccessful participants, for example with regard to the number of decisions and the scope of these decisions, i.e. which factors were manipulated (Dörner et al. 1983:231).

Based on these findings, Roth (1985) argued that problem solving performance was influenced by participants' cognitive styles. The term *cognitive style* refers to the selection of information that is considered and how this information influences decision processes. Furthermore, he assumed that these differences could be observed in the linguistic structure of think aloud protocols (Roth 1985:183). Thus, Roth (1985) used a similar dynamic decision task, namely the 'tailorshop' problem to elicit think aloud protocols. In this task, participants were told that they were running a tailorshop for 15 months. Their task was to maximize the profit of the business and to make sure that the workers were motivated. The participants could change the current state of the system by taking various actions, such as buying raw material, hiring or firing workers, or increasing or cutting costs for advertisement (Roth 1985:183). The participant's success was evaluated based on the number of months in which the business generated profit.

In the explorative study, Roth (1985) based the linguistic analysis of the elicited think aloud protocols on a large number of linguistic markers identified in various previous studies by other scholars (see Table 3.3 in the appendix for a summary). According to Roth (1985), previous studies revealed that these linguistic markers were correlated with abstractness, negation, and egocentrism; characteristics that he assumed to be indicators of cognitive style. Roth (1985) specifically referred to Ertel's (1971)[53] scheme for rating texts based on the dogmatic content[54] that was expressed by their authors because Ertel's findings had been confirmed in a number of studies. Ertel (1971:251-252) outlined that lexical expressions of frequency, amount, scope, and certainty as well as conjunctions and auxiliaries are indicative of a more dogmatic text or suggest an open-minded non-dogmatic text. In addition to markers previously described, Roth (1985:182-189) proposed that indicative mood and conjunctions might be further indicators of cognitive style.

Roth's (1985) analysis revealed that unsuccessful problem solvers used more linguistic features belonging to the category 'dogmatic writing/thinking'. Furthermore, their protocols contained adversative conjunctions, negations, and modal verbs in the subjunctive mood at a statistically significant higher frequency than those by successful participants (Roth 1985:185-186). Contrary to prior expectations, Roth (1985) did not observe statistical differences at the level of nominal abstractness, causative conjunctions, and modal verbs in the indicative mood (Roth 1985:186). In the discussion, Roth (1985) argued that these findings illustrated different strategies of coping with missing information at the beginning of the task. Besides information that was provided in the instructions, additional information could be gained by close observation of the 'tailorshop' system while working on it. Roth's (1985) analysis revealed that unsuccessful problem solvers had the tendency to consider inferred information about the structure of a real tailorshop and possible influencing factors as statements representing unchangeable background for their actions (Roth 1985:188). According to Roth (1985), this tendency to fill in missing information by inferences from the real world and label it as definite was expressed in the high frequency of features of dogmatic thinking. Furthermore, he argued that the repeated use of negations,

[53] Importantly, Ertel (1971:262) points out that dogmatic features may be bound to the text genre that is analyzed. Thus he stresses that the frequency of the analyzed lexical expressions does not allow conclusions about the writer's personality. These far reaching conclusions can only be drawn after analyzing different texts by the same author. If think aloud protocols are analyzed, as proposed by Roth (1985), it may be possible to draw tentative conclusions about the speaker's personality or at least about the verbalized mental problem solving structure because this text represents the speaker's thoughts.

[54] Ertel (1972:242) defines dogmatic to be a closed system of beliefs that is markedly different from a system of disbeliefs. People operating within a non-dogmatic system are open for new and unexpected information and incorporate those into their existing system.

indicating the boundaries between options, illustrated that unsuccessful participants felt the need to decide on one alternative (Roth 1985:189).

Think aloud protocols by successful participants, in contrast, suggested that these participants considered first associations and background knowledge about the structure of a real tailorshop as explorative hypotheses and analogies for the respective problem that they were asked to work on (Roth 1985:188). Roth (1985) argued that the infrequent use of subjunctive mood indicated that successful problem solvers were working in the respective workspace, continuously incorporating new information that was created during interaction with the system. This latter interpretation was supported by the observed low frequency of linguistic features associated with the language style labeled 'dogmatic thinking' by Ertel (1971).

In a second study, Roth (1987) analyzed the linguistic structure of think aloud protocols elicited while participants worked on another complex problem, namely DORI. In this problem, participants were asked to work out a plan for nomads to raise cattle successfully. In order to change the current state of the system, participants could, for example, buy new pasture or cattle. Success was evaluated based on obtained results on the variables of 'number of nomads', 'number of cattle', and 'quantity of fodder' (Roth 1987:209). The quantitative analysis of linguistic markers, as described in Roth (1985), confirmed the results obtained in the previous study. By combining a content-based analysis with a linguistic analysis, Roth (1987:213) suggested that successful problem solvers considered data provided by the system and continuously integrated this feedback. According to Roth (1987), this tendency was reflected in the frequent use of adverbs expressing limitations, such as 'some' and 'frequently', which was constant throughout the trials. Unsuccessful participants tended to use adverbs indicating generalizations, such as 'absolutely', 'all', or 'always'. At the same time, the use of negation increased in think aloud protocols of unsuccessful participants (Roth 1987:212). According to Roth (1987), these findings suggested that unsuccessful participants were increasingly puzzled by the system and thus they repeatedly rejected new information (Roth 1987:214). Roth (1987) argued that these findings, along with those from his previous study, reflected different cognitive styles, i.e. integration of new information and formulation of hypotheses about them as opposed to focusing on initial inferences and generalizing them along with rejection of new information.

Adopting a similar approach, Bartl and Dörner (1998) investigated the linguistic structure of think aloud protocols recorded while participants solved the 'öko-bug' problem. In this task, participants were shown the insects initial appearance and were instructed to change it into the provided desired appearance. They were told which features could be changed by which actions (Bartl & Dörner 1998:227). Success was evaluated based on the number of problems that were solved and the number of actions, i.e. manipulations, that were needed. Furthermore, the authors assessed participants' memory by asking them to recall the influence of specific operators.

In their linguistic analysis, Bartl and Dörner (1998) compared the think aloud protocol of one successful participant with that of one unsuccessful participant. The successful participant needed fewer manipulations and referred to the provided information about operators less often than the participant who did not solve the problem (Bartl & Dörner 1998:234). Bartl and Dörner (1998:235) focused on the analysis of conjunctions since those highlight the relations that the speaker assumes to hold between the connected statements. In their analysis, they studied the frequency of adversative, consecutive, instrumental, conditional, causal, and temporal conjunctions. The analysis revealed that the successful participant used all of these conjunctions whereas the unsuccessful participant mainly used adversative conjunctions. Bartl and Dörner (1998:235) argued that this difference highlighted the successful participant's tendency to draw different connections between information, such as temporal sequences of operators or causal relations between perceived states and operators. Furthermore, the analysis of verb tense highlighted that the successful problem solver formulated hypotheses that were marked by the use of subjunctive mood and the phrase "I think" (Bartl & Dörner 1998:236). Based on these findings, Bartl and Dörner (1998) conclude that the analysis of conjunctions in think aloud protocols suggested that successful participants were capable of connecting given and generated information by means of different relations and to formulate tentative conclusions, i.e. hypotheses. These findings support those results reported by Roth (1985).

In sum, these studies provide an overview of linguistic markers that were used to identify different ways of integrating information to make decisions. The findings are comparable because all three studies investigated similar dynamic decision tasks. The observation that all of them highlight the same differences between successful and unsuccessful participants with regard to the kind of information that is considered and how it is used to guide the problem solving process, strengthens the individual conclusions. Nonetheless, these conclusions could be further supported by studies that elicit additional behavioral measures, such as eye-tracking data, along with think aloud reports.

3.3.1.2 Linguistic representation of mental processes in think aloud protocols

Whereas Roth (1985; 1987) and Bartl and Dörner (1998) investigated cognitive style in think aloud protocols, another research group focused on the description of the linguistic structure of individual problem solving processes. In a number of publications, Caron-Pargue and her collaborators reported analyses based on think aloud protocols elicited while participants solved the Tower of Hanoi task or the problem of the Chinese rings[55].

In 1987, Caron and Caron-Pargue described their research approach for the first time. They started out by proposing four basic operations of argumentative discourse, namely *lexical selection*, *thematic structure*, *'prise en charge'*[56], and *connectives*. Caron and Caron-Pargue (1987:171) argued that these phenomena had previously been studied in terms of the referential use of language exclusively. In contrast, they wanted to study these basic operations in relation to "the production of speech while the speaker is himself constructing and transforming a representation of a given situation" (Caron & Caron-Pargue 1987:172). Based on the assumption that the linguistic devices used in argumentative discourse and concurrent verbalizations during problem solving were general in nature, they argued that findings on argumentative discourse were a reasonable starting point for their analyses. In this paper, Caron and Caron-Pargue (1987) cited examples from protocols to illustrate their theoretical considerations but they did not provide details about the experiment(s) in which those protocols were elicited. In their concluding statement, they admitted that their paper provided only a "sketchy presentation of our results" (Caron & Caron-Pargue 1987:176). However, they tested their theoretical considerations in a number of studies in subsequent years and refined their theoretical framework (e.g. Caron-Pargue & Caron 1991; Bégoin-Augereau & Caron-Pargue 2003; Bégoin-Augereau & Caron-Pargue 2010). Two of these studies will be described in the following paragraphs to illustrate how Bégoin-Augereau and Caron-Pargue combined performance measures and the presentation of the current state with linguistic analyses to provide evidence for their proposal that specific linguistic features can be interpreted as markers of mental processes.

[55] For more information on the Tower of Hanoi see chapter 2. The Chinese rings task is structurally similar.

[56] 'Prise en charge' refers to the marking of the speaker's attitude towards the propositional content (Caron & Caron-Pargue 1987:171).

Bégoin-Augereau and Caron-Pargue based their theoretical framework on Culioli's (e.g. 1995) enunciative model. In one of their earlier papers, Bégoin-Augereau and Caron-Pargue (2003) proposed that

> "on the basis of linguistic markers, cuts into episodes will lead to a demarcation of the units of cognitive processing; the links and boundaries between utterances will define the elementary actions made by the subjects." (Bégoin-Augereau & Caron-Pargue 2003:81)

They tested their proposal experimentally by eliciting think aloud protocols of 60 participants[57] while they solved the Tower of Hanoi task. They collected verbal reports of four trials for each participant in order to investigate effects of learning. For the analysis, each verbalization in the think aloud protocols was complemented by the description of the scene, i.e. which ring was on each of the three pegs (see table 1a in Bégion-Augereau & Caron-Pargue 2003:83). The analysis of the linguistic features of the elicited protocols revealed, for example, that the frequency of changes in naming or double naming, i.e. repetition, was constant throughout the four trials. Bégoin-Augereau and Caron-Pargue (2003:86) interpreted these findings as evidence that these linguistic features characterized "the simplest and automatic representation constructed through expertise". Based on the changes that were executed by manipulating the rings and the corresponding linguistic description of this action, Bégoin-Augereau and Caron-Pargue (2003:87) proposed that the investigated linguistic markers expressed different levels of 'structurations and reorganizations of the representation'. They argued that *starting terms*[58], *constituent locators*[59], and *connectives* represented main changes in representations. *Interjections* were proposed to express resolutions at critical points during the task. The categories *change of naming* and *double naming* were proposed to signal local reorganization.

In a later study, these findings were extended by an analysis of spaces that were assumed to be involved in problem solving. Bégoin-Augereau and Caron-Pargue (2009) proposed that an *external space* could be distinguished from an *internal space* based on linguistic features in think aloud protocols. They elicited think aloud protocols from 20 children working on the Tower of Hanoi task throughout four consecutive trials. Performance was evaluated with regard to time and moves that were needed. The linguistic analysis focused on the distribution of *starting*

---

[57] These participants were 7, 10, and 14 year-old children; 20 participants in each age group.

[58] 'Starting terms' were defined as "terms around which the predicative relation is organized" (Bégoin-Augereau & Caron-Pargue 2003:82).

[59] 'Constituent locators' were defined to correspond "to the construction of the locator around which the utterance is organized" (Bégoin-Augereau & Caron-Pargue 2003:82).

terms, as defined in Bégoin-Augereau and Caron-Pargue (2003), and *interpropositional locations*. The latter linguistic category was defined, in reference to Culioli's (e.g. 1995) theory, as verbalizations of actions in which the items were individually referred to, i.e. the peg was referred to as such in repeated references instead of being replaced by an anaphora in the consecutive reference (Bégoin-Augereau & Caron-Pargue 2009:103). The distribution of these two linguistic markers was analyzed with regard to their 'ratio to moves'[60] and their 'ratio to words' (Bégoin-Augereau & Caron-Pargue 2009:108). Adopting the approach of combining transcriptions of the think aloud protocols with the scene description (see Bégoin-Augereau & Caron-Pargue 2003), the authors proposed that the verbalization of *starting terms* marked abstract representations in the internal space whereas *interpropositional locations* marked representations in the external space. Furthermore, they proposed that an interaction between these two spaces was expressed by the co-occurrence of both markers in the verbalization of one move (Bégoin-Augereau & Caron-Pargue 2009:109). In a technical process that was not comprehensively described in their paper, they proposed to categorize all possible sub-goals on a graph of states. As a main conclusion from their analyses, Bégoin-Augereau and Caron-Pargue (2009:118) proposed that "two distinct decision processes were characterized from a cognitive interpretation of linguistic markers".

In sum, these studies by Bégoin-Augereau and Caron-Pargue highlighted correlations between physical actions, i.e. beginning of an action, mental processes that were involved in these actions, such as decision-making, and linguistic choices. Furthermore, this brief introduction illustrates that hypotheses about the relation between linguistic expressions and mental processes need to be well grounded in theoretical frameworks, such as Culioli's (e.g. 1995) theory in Bégoin-Augereau and Caron-Pargue's research. As most theoretical frameworks on language use focus on interactive settings in which language is used to communicate, the proposed applicability of these theories for the specific setting of think aloud protocols needs to be theoretically well motivated. Furthermore, the work reported by Caron-Pargue and collaborators, most notably with Bégoin-Augereau, highlights that evidence needs to be accumulated across different studies and a large number of participants.

Although the work by Bégoin-Augereau and Caron-Pargue appears sound with regard to its theoretical assumptions based on Culioli's framework and the general procedure of analyses, it needs to be noted that the technical details are difficult to

[60] This measure was defined as "100×ratio of the total number of moves in the section" (Bégoin-Augereau & Caron-Pargue 2009:108).

grasp at times. Additionally, their approach is limited to the Tower of Hanoi problem and the structurally very similar problem of the Chinese rings. A different approach to the analysis of linguistic features in think aloud protocols and the question of how they are related to mental processes is introduced in section 3.2, namely the methodology of *Cognitive Discourse Analysis* that is adopted in this thesis.

3.3.2 Cognitive Discourse Analysis

The term Cognitive Discourse Analysis is used by three researchers denoting different perspectives. Van Dijk (2000) outlines a general approach to discourse analysis that focuses on discourse properties that are based on cognitive concepts and that need to be interpreted in these terms, e.g. mental models of a text that are expressed by local coherence. Taking up van Dijk's proposition, Kibrik (2001; 2011) introduced Cognitive Discourse Analysis as a discipline building links to neighboring disciplines such as psychology and cognitive science.

Thora Tenbrink introduces *Cognitive Discourse Analysis* (CODA in short) as a methodology that highlights the representation of mental processes in discourse, e.g. how planning activities can be distinguished from travelling activities (Tenbrink & Seifert 2011). In this methodology, a content based linguistic analysis in Ericsson and Simon's (1984 (1993)) tradition is combined with an in-depth linguistic analysis of verbal protocols. Ideally, the linguistic analysis is combined with the analysis of performance measures. This section provides an introduction to the method's theoretical assumptions and the proposed procedure. Additionally, with regard to the focus of this thesis, a review of selected studies highlights the insights that have been gained by analyzing verbal reports on problem solving with this method to date. The studies are selected to illustrate the diverse research questions that have been investigated by applying CODA addressing wayfinding tasks in real-life scenarios as well as in experimental settings, and analogical problem solving.

3.3.2.1 Theoretical assumptions

Tenbrink (2010) argued that analyses by means of CODA combine insights on mental processes as obtained in verbal protocols with experimental designs as traditionally used in psycholinguistics. In psycholinguistics, studies are strictly controlled to investigate processes involved in language production and language comprehension. As a consequence, verbal data is highly constrained with regard to the linguistic options that can be chosen by speakers. Tenbrink calls for a more unconstrained elicitation by allowing participants to state what they think in the way that they would normally express it. Furthermore, CODA is intended to analyze unconstrained verbal production data that is elicited in natural environments as well as in laboratory settings. Tenbrink (2010) argued that this approach to the elicitation of verbal data is similar to research in the field of Discourse Analysis and linguistics. Therefore, the theoretical framework draws on concepts and findings in these areas and refines them to highlight mental processes expressed in spontaneous verbalizations.

The basic assumption in the theory of CODA is that spontaneous verbal reports highlight systematic linguistic patterns that are not consciously selected by the speaker. This proposal is based on Ericsson and Simon's (1984 (1993)) model of verbalization of thinking that predicted that verbalizations of thoughts, that do not involve any intermediate processes, represent traces of mental processes.[61] Therefore, it can be assumed that the linguistic structure of such verbalizations expresses the involved mental processes. However, Tenbrink (2008) points out that not all observed differences in linguistic expressions reflect mental processes because verbalization is also influenced by an assumed addressee and the nature of the discourse task. Tenbrink (2008) argued that different kinds of discourse tasks (e.g. describe the shortest route vs. walk the shortest route while thinking aloud) and communicative goals (e.g. plan for yourself vs. plan for an uninformed partner) result in different conceptual perspectives on that task. These perspectives need to be identified in the recorded verbalizations of the discourse task (Tenbrink 2008). Tenbrink (2010:122) described this process of identification of influencing factors to be a challenge. One possibility to tackle this potential difficulty is the elicitation of verbalizations on different discourse tasks on the same problem by one participant (Tenbrink & Gralla 2009:5).

To highlight the differences of information that can be elicited by different discourse tasks, Tenbrink (2008) provided an overview on the characteristics of different discourse tasks solved by one participant when working on the Traveling Salesperson Problem (abbreviated TSP)[62]. In this paper, she compared information elicited in a think aloud protocol, in a retrospective report on how the task was solved, and in an instruction for an imagined friend. In line with Ericsson and Simon's (1993) model, these tasks were assumed to contain information on

[61] For more detail see distinction between *level 2* verbalization and *level 3* verbalization in section 3.2.1.

[62] In the traveling salesman problem participants were instructed to visit a number of spatial goals before returning to the start point following the shortest possible path (Tenbrink 2008:126).

different mental processes. Whereas think aloud protocols were assumed to contain traces of mental processes when they are attended to, retrospective reports were assumed to contain information on "mental processes associated with the spatial task from memory" (Tenbrink 2008:127). The instructions, in contrast, were expected to include generalizations of the procedure specifying only those actions that were successful. In each of these reports, Tenbrink (2008) analyzed the distribution of linguistic markers indicating topic shifts, conceptual complexity, structure, and coherence. Drawing on Functional Grammar, Tenbrink (2008:129) used Halliday's terminology to identify explicit structural markers.

The analysis revealed that temporal markers, such as 'first', 'now', and 'then', were very frequent in the think aloud protocols. In most cases, those markers were accompanied by a reference to a particular symbol. The mention of specific symbols changed to references to regions over consecutive trials. Furthermore, the data highlighted that the discourse marker 'okay' was used when new trials started but also when a trial was ended. The discourse marker 'so' frequently occurred after one or several objects were visited. Tenbrink (2008:130) argued that this specific pattern suggested that 'so' marked the completion of subtasks. Similarly, the location of the particle 'ja' (well) suggested that it marked immediate, perceptual processes because it was most frequently verbalized when "the presence of a symbol directly close by" (Tenbrink 2008:130) was reported.

Tenbrink (2008:130) observed that most information units exhibited the following structure: temporal marker + action verb + next target symbol/region. The action verb was in the present tense in the majority of cases. Drawing on Ericsson and Simon (1984) she argued that this structure

> "provided evidence that the participant was truly focused on the task and did not produce irrelevant or long-term memory based utterance (Ericsson & Simon 1984)." (Tenbrink 2008:131)

If generalizations were identified in the data, those were marked by the adverb 'again' and were related to the region-based perspective, i.e. referring to "the other side" or "the left" (Tenbrink 2008:131) rather than to specific symbols. This strategy was explicitly stated in instructions to a friend. Tenbrink (2008:132) argued that the discovery and repeated use of this strategy marked high-level mental processes. Additionally, she identified expressions of *search* and *control* processes. In the think aloud protocol, the participant verbalized searching for symbols and repeatedly controlling the "shopping list"[63] (Tenbrink 2008:132). These two pro-

[63] The shopping list contained all symbols that needed to be visited to complete the task and the start symbol which also resembled the goal symbol.

cesses were reported in more general terms, i.e. not related to specific symbols and instances, in both written reports.

In her analysis of the think aloud protocol, Tenbrink (2008) highlighted that the participant did not only use the region-based strategy but the participant also mentioned the "nearest-neighbor" heuristic[64]; both strategies are described as possible strategies to solve the TSP problem in the literature. The comparison between the information verbalized during task performance and the information that was recalled immediately after the task was completed, i.e. in the retrospective report, revealed that the participant did not refer to the "nearest-neighbor" heuristic. This finding suggested that think aloud protocols provided insights that were not recalled by the participant herself after consecutive trials. In sum, these findings revealed that differences in linguistic features correlated well with differences in presumed mental processes motivated by the distinctions proposed by Ericsson and Simon (1993) and other researchers, thus supporting the validity of the approach.

In the concluding remarks, Tenbrink (2008:134) pointed out that these first insights needed "to be quantified and validated across a larger set of participants and different task variations". Only validation across a larger data-base would allow for insights about the relationship between linguistic representations and mental processes and highlight mental processes that are relevant for specific problem solving tasks (Tenbrink 2008:134). Tenbrink and collaborators pursued this aim by eliciting different kinds of verbal reports along with performance measures on different problem solving tasks, such as the TSP (e.g. Tenbrink & Wiener 2009), holiday planning (Tenbrink & Seifert 2010), navigating through real-world and simulated environments (e.g. Tenbrink et al. 2011; Klippel et al. 2013), or analogical problem solving (Gralla et al. 2012). A selection of these studies will be presented to highlight the accumulating evidence for Tenbrink's (2008) theoretical assumptions.

Besides applying CODA to study mental processes in problem solving, Tenbrink and collaborators used the methodology to investigate mental representations expressed in spatial referential identification and location tasks, either with a specific addressee (e.g. Moratz & Tenbrink 2006; Tenbrink & Shi 2007; Andonova et al. 2008; Tenbrink et al. 2008; Andonova et al. 2010) or without (e.g. Tenbrink 2005; Vorwerg & Tenbrink 2007; Tenbrink 2009). Since this thesis focuses on the application of CODA for analyzing verbal reports elicited in a problem solving

[64] In the nearest-neighbor heuristic, participants select routes that connect items that are close to each other.

scenario, only findings in the domain of problem solving will be reviewed in the next sections.

3.3.2.2 Application: Studying problem solving in navigation tasks

Based on theoretical considerations regarding previously identified and described strategies for solving TSP problems as well as the influence of goals, distractors, and the actual setting on performance (for a theoretical discussion see Wiener & Tenbrink 2008), Tenbrink and Wiener (2009) conducted two studies that investigated performance and linguistic structure of verbal reports in a version of the TSP problem involving distractors and feature information. The focus in the experiment was on a task that was shown to highlight usage of three well-described strategies, namely the *cluster strategy*, the *region-based strategy*, and the *nearest-neighbor strategy*. Participants were shown lists of symbols that they needed to collect ('shopping lists') that contained between four to nine target symbols. Performance was evaluated by comparing the length of the chosen path to the length of the optimal path. Furthermore, the number of trials in which the optimal path was found was evaluated. The analysis revealed that participants' performance was better in tasks with fewer target symbols and participants performed better in tasks in which the nearest-neighbor strategy could be adopted (Tenbrink & Wiener 2009:149).

In addition to performance measures, Tenbrink and Wiener (2009) elicited written retrospective reports immediately after all trials were completed and instructed participants to write an instruction for a friend. In the retrospective reports, participants reported having used two strategies, namely a group-clustering strategy and a strategy of focusing on trajectory features. Additionally, the analysis revealed that participants mentioned color much more frequently than shape of symbols. According to Tenbrink and Wiener (2009:153), this finding indicated that color was not only perceived as a feature but that the focus on color rather suggested "a conscious decision to simplify the problem space by focusing on it". This conclusion was supported by findings in the second study in which color facilitated finding the correct path in one condition[65]. The frequency of mention of color increased significantly in the second experiment as compared to the first one (Tenbrink & Wiener 2009:155).

Common trajectory features, that were mentioned in retrospective reports in the first study, were circles or patterns. The analysis of verb types used revealed that

[65] The second experiment was run with participants who did not participate in the first study (Tenbrink & Wiener 2009:153).

participants frequently mentioned searching for patterns and paths and thinking about them (Tenbrink & Wiener 2009:160). These results are similar to those described in Tenbrink (2008). Furthermore, Tenbrink and Wiener (2009) identified the same structuring devices as in Tenbrink's (2008) previous study. Retrospective reports, for example, were frequently structured by temporal markers, such as 'first' and 'then'. Tenbrink and Wiener (2009:151) argued that this highlights how participants imposed a temporal structure on a complex and previously unstructured task. Additionally, the analysis revealed frequent use of markers of extension and matter enhancement, such as 'and', 'also', and 'in addition'. In line with Tenbrink (2008), Tenbrink and Wiener (2009) argued that these connectors reflected the large number of subtasks that needed to be performed. This comparison highlights that the second study conducted on a refined version of the TSP problem and with more participants supported results that were obtained in Tenbrink's (2008) earlier case study.

Tenbrink and Wiener (2009:141) argued that the TSP is structurally similar to navigation tasks that involve finding the route from a start location to a target location. Hölscher et al. (2011), for example, investigated the influence of situated and prospective planning processes on route choice and their linguistic representation in different verbal reports. They reported two experiments that were very similar, i.e. both involved real-world navigation through downtown Freiburg eliciting data on prospective planning and situated planning. In the first experiment, participants were asked to plan how they would walk from a predefined start to a given goal location and provide a verbal route direction (*prospective planning*). Afterwards participants were asked to walk from the same start to the same goal location and think aloud along the way as just described (*situated planning*). Participants were not constrained to walk the same way. Hölscher et al. (2011:229) argued that prospective planning required that a complete route was planned in advance in which "detailed spatial information about streets and intersections needs to be derived from memory". Situated planning, in contrast, "allows for an incremental optimization of the overall plan by adding in local direction information" (Hölscher et al. 2011:230). In the second experiment, participants were instructed to write down a route that they would walk themselves from a pre-defined start to an indicated goal location (*prospective planning for self*). Additionally, they were instructed to write a route direction for an informed addressee (*prospective planning for other*). As a third task, participants were instructed to walk from the start to the goal location while thinking aloud (*situated planning*). In this condition participants needed to walk the planned way and were encouraged to mark corrections on their written descriptions.

Performance was evaluated with regard to efficiency of the selected route in comparison to the optimal route, route complexity, i.e. number of turns and number of streets walked on, and its characteristics, such as side street or main road (Hölscher et al. 2011:238). Furthermore, participants completed spatial ability tests in the second experiment. The findings on performance measures were the same in both experiments. The analysis highlighted a significant difference between routes described in prospective planning and those walked in the situated planning task. The actual routes that participants walked in the situated planning task were more efficient including more turns and streets. Moreover, an influence of planning for oneself and planning for someone else was observed with regard to street characteristics. When participants described a route for an uninformed addressee they choose prominent streets and offered more local guides for orientation (Hölscher et al. 2011:241). No influence of spatial ability on route choice was observed in the second experiment.

The content based analysis of think aloud protocols collected in the second experiment[66] highlighted that participants selected routes based on personal preferences, opportunity, e.g. a green stop light, route length, and orientation towards the intended goal location in situated planning conditions (Hölscher et al. 2011:241). Hölscher et al. (2011:241) argued that the linguistic analysis revealed that intentions, decisions, and current planning processes were frequently expressed by modal verbs (e.g. should, could, might), verbs of thinking (e.g. consider, think, decide), conjunctions signaling elaborations and reasons (e.g. whether, or, if), and adverbs expressing uncertainty (e.g. maybe). One quoted example further highlights that participants used temporal markers, such as 'jetzt' (now) (Hölscher et al. 2011:242). Hölscher et al. (2011) suggested that

> "together these verbalizations indicate that the participants failed to produce a complete mental representation of the spatial array or the planned path prior to navigation, and that the route plan was constantly updated during navigation." (Hölscher et al. 2011:242)

This interpretation was supported by the observation that participants corrected place or street names, details for orientation, such as salient landmarks, and information about turns or segments in their written directions when they actually

[66] The paper reports more findings concerning the linguistic analysis for the second experiment thus the focus is on these findings.

walked the route. Differences between directions for oneself and for an uninformed addressee were observed with regard to number of words.

The linguistic analysis confirmed previous findings reported on the TSP problem by Tenbrink (2008) and Tenbrink and Wiener (2009). In line with findings by Hölscher et al. (2011), Tenbrink and Wiener (2009:151) reported that retrospective reports frequently contained verbs of thinking, conditional enhancement and elaboration, as well as temporal markers. Furthermore, Tenbrink's (2008) analysis of think aloud protocols revealed the frequent use of the temporal marker 'jetzt' (now) that was also observed in think aloud reports elicited by Hölscher et al. (2011). This finding supports Tenbrink's (2008:129) assumption that 'now' accompanies action and signals the participant's focus on the current situation.

Whereas the studies by Hölscher et al. (2011) analyzed route descriptions and verbal reports in a natural outdoor environment, Tenbrink et al. (2011) studied these two kinds of verbalization in an indoor environment. Tenbrink et al. (2011) asked participants who were unfamiliar with the building to describe the shortest route from a pre-defined start to an indicated goal location in a complex building (*novice*). Additionally, they asked employees, who were very familiar with the building, to describe their route choice (*expert*). This condition resembles *prospective planning*. Afterwards, participants were asked to walk from the start location to the destination while thinking aloud, which resembles *situated planning*. Participants were not constrained to walk the route that they had just described. This study was explorative as no previous studies investigated navigation in indoor real-world environments (Tenbrink et al. 2011:1263).

There were two independent variables, namely knowledge about the environment, i.e. expertise, and the pre-defined start location. As in Hölscher et al. (2011) and Tenbrink and Wiener (2009), behavioral performance was evaluated in terms of route efficiency in relation to the shortest route. The linguistic analysis focused on route characteristics, markers of orientation, and uncertainty. Inspired by previous work in outdoor scenarios, route characteristics were analyzed with regard to information on start and end location, segments, change in direction, landmarks, regions, and distance (Tenbrink et al. 2011:1264). As expected, the analysis revealed that experts found more efficient routes than novices. This finding was mirrored in the observation that novices mentioned significantly fewer concrete spatial elements but more markers of uncertainty, such as 'maybe', and more orientation markers, such as references to public orientation aids (Tenbrink et al. 2011:1264). Based on these findings, Tenbrink et al. (2011) argued that

> "linguistic markers such as vagueness particles, hedges, modals, and perception and orientation verbs to some extent indicate the lack of expertise on the part of the speaker, whereas references to concrete spatial entities convey spatial certainty." (Tenbrink et al. 2011:1267)

Comparability of results between these findings and those reported by Tenbrink and colleagues for previous studies is limited for two reasons. First, this study was conducted in an indoor environment whereas previous studies were conducted outdoors (e.g. Tenbrink & Wiener 2009; Hölscher et al. 2011). Second, previous studies elicited verbal reports from participants who were very familiar with the area, i.e. corresponding to the expert status in this study. Therefore, the main findings of this study cannot be compared to previous studies but they suggest trends that should be tested in follow-up studies, specifically with regard to markers of uncertainty. Furthermore, the analysis highlighted that participants' verbal reports could be analyzed by applying the same categories, such as references to landmarks, information on turns and segments, and change of direction.

In contrast to these previous studies, Tenbrink and Seifert (2011) investigated participants' planning behavior in a task that did not involve actual navigation but rather planning on a map. This task involves two layers of conceptualization, i.e. the information that is displayed on the map and their real-world correspondence. They investigated participants' planning activity in written tour plans for holiday trips around Crete (a real island) or Cretopia (a fictional island). Whereas Crete has an oblong shape, Cretopia was designed as a round shaped island. The performance analysis focused on the shape of the selected route, i.e. its trajectory, and the content based analysis investigated the strategies that were chosen to satisfy the task constraints, i.e. either visit 13 predefined cities or to ensure that different activities are included. The linguistic analysis focused on the representation of the two layers of conceptualization, i.e. "planning with respect to the map, versus traveling in the real world" (Tenbrink & Seifert 2011:111).

Similarly to previous studies on the TSP (e.g. Tenbrink & Wiener 2009), Tenbrink and Seifert (2011) analyzed performance based on the trajectory of the route that was drawn on the map and the verbal description of the applied strategies. The linguistic analysis of retrospective reports focused on the representation of the two conceptual layers. To identify these layers, Tenbrink and Seifert (2011:113) coded the semantic content of nouns, verbs, adjectives, adverbs, and temporal markers. They categorized each annotated item as indicator for the conceptual domain of planning (e.g. 'map' (noun), 'helpful' (adjective), and 'plot' (verb)) or travelling activity (e.g. 'trip' (noun), 'boring' (adjective), and '2 days' (temporal markers))

(Tenbrink & Seifert 2011:115). Cases in which this binary categorization was not possible were also indicated.

The analysis of the shape of selected routes revealed that participants preferred round routes for both islands. Furthermore, different strategies, such as *nearest-neighbor*, *regionalization*, and *grouping*, were reported. Out of those, the first two strategies were most frequently mentioned. The linguistic analysis highlighted that participants referred to the travel domain almost as frequently as to the planning domain. The analysis of informational units that contained references to both domains revealed different ways of connecting these domains. Tenbrink and Seifert (2011:116) reported that modal verbs (i.e. 'could' and 'should'), the discourse marker 'in order to' as well as the preposition 'as' frequently occurred when the two domains were mentioned in one informational unit.

The reported analysis suggests that participants who are planning a holiday trip rely on the same strategies that are employed when solving a less naturalistic problem, e.g. the TSP. Furthermore, the comparison between the findings reported by Tenbrink and Wiener (2009) and those reported by Tenbrink and Seifert (2011) highlight that these strategies can be identified based on their linguistic representation in retrospective verbal reports. The study by Tenbrink and Seifert (2011) also highlighted one specific aspect of planning activities that involve maps, namely that of distinguishing between the visually present domain of the map and the conceptual domain of the environment that is displayed on the map. Their analysis highlighted the linguistic repertoire with regard to nouns, verbs, and adjectives by which these domains are referred to. Additionally, the analysis revealed patterns by which these domains are linguistically connected, i.e. conceptually mapped.

In sum, the reviewed studies individually highlighted how linguistic patterns correlated to differences in performance, e.g. in situated planning vs. prospective planning. By comparing the results between studies, individual findings are strengthened as pointed out throughout this section. The reviewed studies explored new areas of application (e.g. Tenbrink et al. 2011) while consistently establishing connections to well-researched paradigms such as navigation and solving the TSP. However, as the approach of *Cognitive Discourse Analysis* is a fairly recent one, comparability is still limited; a fact that encourages further studies investigating similar aspects.

3.3.2.3 Application: Studying analogical problem solving

Based on the finding that CODA was well suited to distinguish between different navigation strategies, it was recently applied to investigate strategies in analogical problem solving. Gralla et al. (2012) studied participants' spontaneous strategies and verbalizations on solving the Eulerian trail problem, a well-known path-finding problem. In previous work, Schelhorn et al. (2007) investigated the influence of saliency of analogous elements on strategy selection based on performance measures. Their theoretical framework was based on Carbonell's (1983; 1986) theory on derivational and transformational analogy. Schelhorn et al.'s (2007) analysis was based on reaction times as well as mapping times and accuracy between source and target entities. Gralla et al. (2012) extended this study by eliciting different kinds of verbal reports during and after performance. One aim was the identification of the applied strategies and their comparison with the theoretically presumed ones. Additionally, the study aimed at identifying systematic linguistic differences characterizing the two expected strategies.

The independent variables were saliency of analogy and kinds of verbal reports. In the 'planning condition', participants were shown an exemplary problem with the solution. Then they were asked to plan how they will solve the presented target problem before actually solving the problem. In the 'retrospective report condition' participants worked their way through the exemplary problem with the solution and provided an answer to the target problem without interruption. Afterwards they were asked to report how they solved the problem and to provide an instruction for a friend describing how this problem should be solved.

In order to compare the observed performance measures to those reported in Schelhorn et al. (2007), solution times, solution correctness, and mapping performance were assessed. Additionally, linguistic categories were identified that were defined based on theoretical considerations relating to the assumed strategies of analogical transfer as outlined in detail by Carbonell (1983; 1986) and addressed in explorative work by Schmid and Carbonell (1999). Further markers were identified iteratively during the analysis. The analysis of performance measures confirmed the results reported by Schelhorn et al. (2007) and no interaction between elicitation of verbal reports and mapping times was observed (Gralla et al. 2012:399).

The content based analysis of retrospective reports revealed that some participants did not work through the example problem as suggested by the instructions but rather solved it on their own. When providing an answer to the target problem, they reused their solution after discovering the analogical nature of the example and target problem. If the initial solution was wrong, the solution to the target problem was also wrong despite the fact that the participant discovered and strategically used the analogous nature of the problems. Considering solution correctness only, the discovery and application of analogy could not be detected in these cases.

The linguistic analysis revealed systematic differences between verbalizations depending on the saliency of the analogy. The analysis further revealed that the theoretically proposed strategies of analogical transfer differed in their linguistic representation (Gralla et al. 2012:400). The analysis emphasized differences in the linguistic repertoire with regard to verbs, nouns, temporal order, and markers of generalization ('again') and similarity ('same'). Additionally, the analysis revealed two strategies that were not previously described highlighting the diversity of strategies applied in analogical problem solving. Based on these findings, Gralla et al. (2012) concluded that

> "the analysis of verbal data contributes to a more detailed understanding of the processes at work during analogical transfer." (Gralla et al. 2012:401)

The results of this study suggested that analyses of performance measures alone might not capture the whole picture of people's problem solving behavior because the same observable performance might result from the use of different strategies. This finding is in line with those reported in the reviewed studies on navigation highlighting the gain of supplementing the analysis of performance measures by an in-depth verbal protocol analysis.

3.4 Conclusion

This chapter outlined one methodology that has a long tradition of being used to study human problem solving, i.e. verbal protocol analysis. In reference to the second chapter that focused on problem solving, it has been outlined that verbal protocol analysis was used long before more recent technologies such as eye-tracking and brain imaging were invented and available to a large group of researchers. The reviewed literature illustrated that it is widely acknowledge that concurrent verbal reports provide rich information on the content of working memory and allows for a sequential observation over time. However, the literature review clearly highlighted the importance of the wording of the instructions that are given to the participants, i.e. the instruction to say whatever comes to mind vs.

the instruction to explain and describe the behavior. A recent meta-analysis of a large number of studies (Fox et al. 2011) that conducted performance measures under conditions of concurrent verbalization as well as in a silent condition supported previous findings that the observable influence of verbalization on performance additionally depends on the task that is solved.

One of the acknowledged and discussed limitations of verbal protocol analysis is the fact that only processes that reach awareness can be verbalized. Therefore, each researcher using verbal protocols needs to be aware that verbal protocols can never be complete. However, it has been argued that further insights about mental processes can be gained by supplementing the content-based analysis with in-depth linguistic analyses highlighting systematic patterns of unconscious linguistic choices. The methodology that aims at revealing these patterns across a range of discourse tasks, i.e. concurrent verbalization vs. description vs. instructions, and problem solving tasks was introduced. *Cognitive Discourse Analysis* will be applied in this thesis and refined because it has not been applied to an assembly task before. Moreover, this thesis combines different kinds of verbal protocols, i.e. think aloud reports, retrospective reports, instructions, and guided questions, to elicit different kinds of information. It is proposed that these different sources of information allow for a more complete picture of the processes involved in unaided object assembly.

3.5 Appendix

Reference (abbreviated)	Investigated tasks	Independent measure: kind of verbalization	Dependent measure	Finding
Merz (1969)[67]	Figure-reasoning task, matrices tests, parts of the 'Intelligenztest'	silent, think aloud (with/ without hearing ones own speech)	1^{st}: number of correct solutions 2^{nd}: time	Significant effect of verbalization on 1^{st} and 2^{nd} measure. Verbalization improved performance but those participants needed more time.
Franzen & Merz (1976)[68]	Parts of the 'Intelligenztest', specifically analogy, sentence completion, math, similarities	silent and think aloud	1^{st}: number of correct solutions 2^{nd}: eye movement: fixation rates 3^{rd}: time	Significant effect of verbalization on 1^{st} and 3^{rd} measure. Verbalization improved performance but those participants needed more time. Fixation rates were different.
Ahlum-Heath & Di Vesta (1986)	Tower of Hanoi	silent and constrained concurrent verbalization[69]	1^{st}: number of moves 2^{nd}: performance time	Significant effect of verbalization on 1^{st} and 2^{nd} dependent measure.
McGeorge & Burton (1989)	"Sugar factory" – dynamic decision task	constrained concurrent verbalization[70] and silent	1^{st}: output de-pendent measure of success 2^{nd}: number of outputs for each task	Better performance by group with verbalization on 2^{nd} dependent measure.

[67] Merz (1969) reports results of numerous studies that he conducted with his students. All of them investigated the performance of school children on selected parts of the 'Intelligenztest' as well as the figure-reasoning-test, and matrices tests. The great majority of experiments showed the same results thus this result is summarized for all experiments in the 5^{th} column of this table.

[68] Franzen and Merz (1976) also report numerous studies thus the finding will be summarized as outlined in the previous footnote.

[69] Ahlum-Heath and Di Vesta (1986) instructed their participants to state where a peg will be placed (location) and why (reason) prior to the move thus this verbalization is highly constrained with regard to the information that is being asked for. To distinguish this type form free concurrent verbalization the label *constrained concurrent verbalization* was chosen.

[70] McGeorge and Burton (1989) refer to their method of elicitation as think aloud but they instructed their participants to describe what they are doing and mention all heuristics that they employ (1989:459). This instruction calls for selective information, i.e. description of action and reflections on the meta-level, i.e. heuristics that are used. Since this is not

Deffner (1989)	n-term series tasks – analytical thinking	silent thinking[71] and thinking aloud	1^{st}: performance time 2^{nd}: strategy	No significant difference on 1^{st} measure. Significant difference with regard to 2^{nd} measure.
Russo, Johnson & Stephens (1989)	Verbal, numerical, pictorial, and mental addition task	silent, think aloud, 3 kinds of retrospective report[72]	1^{st}: accuracy 2^{nd}: response time 3^{rd}: forgetting 4^{th}: fabricating	Significant effect of think aloud on 1^{st} measure with regard to numerical and mental addition tasks. Significant effect of think aloud on 2^{nd} measure. Significant differences for 3^{rd} and 4^{th} measure between kinds of retrospective reports.
Short, Evans & Friebert (1991)	Analogies (verbal and spatial)	silent and think aloud[73]	improvement in success rates from 1^{st} to 2^{nd} trial	Significant effect of verbalization for 5^{th} graders but not for adults.
Schooler, Ohlsson & Brooks (1993)	Insight & non-insight problems	4 experiments with different verbalizations, i.e. interruption for reflection, no interruption, interruption for an unrelated task, and think aloud	1^{st}: time 2^{nd}: number of correct solutions	Significant difference between those providing reasons and those with unrelated interruption. No effect on 1^{st} measure in all 3 experiments. Different performance on insight and non-insight problems.

unconstrained verbalization as proposed by Ericsson and Simon (1993) this type is referred to as *constrained concurrent verbalization*.

[71] Deffner (1989) points out that participants may as well verbalize their thoughts when working silently thus he calls the group in which there is no instruction to verbalize all thoughts 'silent thinking'.

[72] Russo et al. (1989) distinguished between retrospective report in which the solution was available (response cued), retrospective report in which the original problem was available (stimulus cued), and retrospective report in which eye fixations were superimposed on original problem (prompted retrospective report).

[73] Besides the independent variable verbalization, Short et al. (1991) introduced the independent variable 'normally achieving children', 'learning disabled children', 'developmentally disabled children', 'normally achieving adults', and 'learning disabled adults'.

Knoblich & Rhenius (1995)	"Kühlhaus-Problem" (cold store problem) – complex systems task	think aloud and silent	1st: eye movements 2nd: solution accuracy[74] 3rd: variation between input values 4th: response time	No significant effect on 2nd and 4th dependent measure. Significant effect of verbalization on 3rd measure. Significant effect on duration of fixation (1st measure).
Bartl & Dörner (1998)	'Käferproblem' (bug problem) – analytical problem	think aloud, interrupted for self-reflection, thinking without speaking, concurrent repeating of numbers, concurrent repeating of clapping, silent	1st: solution accuracy 2nd: memory	Significant differences with regard to 1st and 2nd dependent measure.
Dickson, McLennan, & Omodei (2000)	"Fire Chief" – complex systems task	procedural verbalization[75], associative verbalization, and silent	1st: percentage of saved landscape 2nd: number of decision actions in each trial	Significant effect on 1st measure but none on 2nd. Silent group better than procedural verbalization group.
Gilhooly, Fioratou, & Henretty (2010)	insight and non-insight problems (spatial and verbal)	think aloud and silent[76]	1st: solution rates 2nd: latency	Significant effect of verbalization on 1st and 2nd measure in spatial task. No effect on 1st measure with regard to insight and non-insight problems.

Table 3.2: Summary of studies that investigated the reactivity of concurrent verbalization on performance, highlighting the investigated tasks, the kind of verbalization, the dependent measures, and the findings (ordered by date of publication).

Grammatical feature	Example	Author from whom Roth (1985) adapted it
dogmatic writing/thinking	all, always, without any doubts ...	Ertel (1972)

[74] In German papers, authors either refer to "Lösungsgüte" or "Güte der Systemsteuerung" (Knoblich & Rhenius 1995). Both of these measures are translated by solution accuracy, abbreviated as *accuracy*.

[75] In the *procedural verbalization* Dickson et al. (2000:222) instructed their participants to put into words the basis of each decision for the use of a fire appliance. In the *associative verbalization* condition, on the other hand, Dickson et al. (2000) instructed the participants to verbalize whatever came to their minds during the decision action.

[76] Gilhooly et al. (2010) recorded the session of the experimental as well as the silent control group on audio tape. This procedure is unique to this study.

non-dogmatic writing/thinking	some, sometimes, maybe, possibly …	Ertel (1972)
abstractness	suffixes: '-heit', '-keit', 'ung' …	Güther & Groeben (1978)
negation	none, no, never, nothing …	Ertel & Bloemer (1975)[77]
egocentrism	I, mine …	Schwibbe (1981)
subjunctive mood	'hätte', 'könnte', 'müsste' …	Schöne (1982)
indicative mood		Roth (1985)
causative conjunction	hence, thus, if …	Roth (1985)
adversative conjunction	but, besides, either … or …	Roth (1985)

Table 3.3: Grammatical features analyzed by Roth (1985) in think aloud protocols with examples[78] and their sources.

[77] In their experiment, Ertel and Bloemer (1975:336) test their hypothesis that "a sentence is basically a unit of cognitive action". Thus construction and understanding of a sentence is basically a constructive cognitive action. They postulate that a sentence is decomposed into two cognitive units that are related to each other. In affirmative sentences the cognitive units are joined and in negative sentences they are separated. Roth (1985) does not give any explanation for choosing this study as a reference for the feature of negation. But it can be assumed that he refers to Ertel and Bloemer (1975) because they are investigating the mental representation of a grammatical structure and Roth (1985) is interested in this combination. Following Ertel and Bloemer's (1975) argument, the use of negation indicates that the speaker separates information (hypotheses, facts etc.) into two distinct cognitive units.

[78] Note that the examples as well as the studies that are referred to are in German. Most examples in the table are translated for the reader.

4 Experimental Studies: Methodology and design

4.1 Introduction

This chapter introduces the methodological background and the design of the two experimental studies that are central to the analysis presented in this thesis. Besides providing an overview on the number of participants that were tested in the experiments (section 3), the procedure of each experiment is described including the exact instructions, a description of the equipment that was used, and the technical problems that were faced (section 4). The chapter concludes by providing a general overview on the qualitative and quantitative analyses that were performed in the different chapters of analysis.

Imagine you want to buy a dollhouse for your child for Christmas. You find one at a virtual auction place and you buy it. But when the package arrives you find the parts to be without an assembly manual. What to do next? If you are lucky, you remember the picture, which was displayed during the auction and if you are even more fortunate, you still have the picture on your pc. The task at hand now is to assemble the parts in a way, which matches the picture in the end. The process of matching physically present objects to their function in the goal structure can be classified as problem solving.

Assembling furniture or toys by yourself is not uncommon in times in which warehouses such as IKEA are very popular. The sustained popularity of stores that sell self-assembly[79] products and a study by Richardson (2007) suggest that people are willing to assemble their furniture on their own. As stated in Richardson's (2007) definition, they are generally aided in their effort by a manual that is supplied by the manufacturer. Therefore, considerable time and effort has been invested in research concerning manual design (e.g. Marcus et al. 1996; Maes and Lenting 1999; Norvick & Morse 2000; Agrawala et al. 2003; Ganier 2004) but few researchers focused on how people actually assemble an object (Rieser 1996; Tversky et al. 2009; Daniel and Tversky 2012). In the area of problem solving the majority of studies investigated the performance on logic-based problems (e.g. de Groot 1969; Newell & Simon 1972) or the specific type of insight problems (e.g. Knoblich & Ohlsson 1999; Knoblich et al. 2001). However, considering the complexity of many objects (e.g. furniture, toys) and the popularity of self-assembly products, it is also necessary to study how people solve complex tasks that they face in their daily lives. The task to assemble an object from scratch given different levels of complexity, i.e. with or without external support, represents one of these cases.

As outlined in the general introduction to this thesis, the aim of this investigation is three-fold. The first aim is the investigation of the representation of problem solving processes involved in an everyday task. As furniture assembly is a common phenomenon, it resembles a recurring activity although people do not assemble furniture on a daily basis, except for those who do it professionally. Importantly, people can be assumed to have general knowledge about furniture assembly but will probably not have a mental plan as how to proceed exactly. Research showed that manual-assisted assembly might result in problem solving activities (Richardson 2007). In order to make problem solving activity very likely in the experiments that are analyzed in this thesis participants are not provided with any external help to assist the assembly. This is motivated by the assumption that the lack of visual or textual aid will foster problem solving activity. The second and

[79] The terminology self-assembly is used in the sense defined by Richardson (2007:307) who states that self-assembly furniture are "furniture that are sold 'ready to assemble'. [More specifically,] the purchaser must follow supplied instructions and assemble the item before it can be used."

third aims concern the influence of two factors on the assembly process, i.e. amount of prior information and communicative intention.

This chapter focuses on the presentation and motivation of the experimental design of the two studies that were conducted. First, a brief summary of problem solving in assembly will be provided to establish the general background; for a description of the theoretical framework see chapter 6. Second, the theoretical motivation for choosing prior information and communicative intention as independent variables is outlined. This general introduction to the topic is provided here to motivate the design of the two experiments. However, each of the areas will be re-addressed in the light of different foci of investigation in the individual chapters on features of analysis. The methodological background of the analysis of verbal reports is not restated here because it has been outlined and discussed in the previous chapter. Therefore, only the selection of the specific kinds of verbal reports will be motivated in 4.2.2.

4.2 Methodology

4.2.1 Theoretical background

4.2.1.1 Problem solving task: object assembly

Intuitively, assembling an object involves numerous sub-tasks. First, a mental representation of the goal structure needs to be constructed. Second, an action plan is derived, i.e. which objects need to be manipulated and in which way. Third, the action plan needs to be executed and resulting states need to be constantly evaluated. The first step in the assembly procedure is based on perception because the assembler combines information from the equipment with prior knowledge. Additionally, information from the manual can assist the creation process if it is provided (Ganier 2004). The second step of deriving an action plan is based on the human ability of decision making (Helander & Willén 2003). Generally, action plans are hierarchically structured (Dixon 1982; Tversky et al. 2009). More specifically, they specify which objects are manipulated, in which sequential order, and in which fashion. In assembly tasks these manipulations are based on elemental tasks such as grasp, reach, move, position part, and insert (Helander & Willén 2003). On a more general level these manual operations can be described by the following activities: select, orientate, position, and fasten components (Richardson et al. 2004). Concerning the activity of monitoring, Ganier (2004) distinguishes between a global and a local level of monitoring and regulation. On the global level, the reference point for monitoring is the mental task representation whereas the currently performed instructions are the point of reference on the local level. In the context of the experimental studies in this thesis, the goal representation, i.e. the dollhouse, corresponds to the global level. Whereas the individual object parts correspond to the local level.

This brief description of the general processes involved in assembly highlights that manuals can facilitate the assembly process at different levels. Manuals can help the creation of a mental representation, provide an appropriate action plan, and serve as a point of reference in monitoring activities. However, research shows that following instructions does not always result in success (e.g. Mani & Johnson-Laird 1982; Gentner & Stevens 1983; Novick & Morse 2000). In cases in which the manual fails to provide support, the assembler is faced with a problem. Problem solving situations are generally characterized by the desire or need to change the current state in order to reach a different goal state. In these situations the procedure to change the current state into the goal state is not known but needs to be constructed. The appropriate procedure can be derived by one of two strategies. On the one hand, known operations can be rearranged and combined in a creative process (e.g. de Groot 1969; Dörner 1987). On the other hand, known procedures and sequences of actions can be applied by means of analogical problem solving (e.g. Greeno 1978; Anderson 2005).

Based on this general description of problem solving situations, it is proposed that object assembly may result in problem solving activities. This is specifically likely if assemblers have no manual and the assembly procedure is not known to them. Then the assembler needs to devise an appropriate sequence of actions to manipulate the given objects in a way that results in the desired goal structure. Although the basic ways of manipulating objects are known to assemblers, the correct sequence of these actions and the appropriate objects need to be selected and combined by the assembler himself/herself. The selection of appropriate objects is based on knowledge about their potential function. This knowledge can to be acquired by the assembler by close inspection of objects for example or by remembering the function of similar objects in other tasks.

4.2.1.2 Influence of prior information on understanding

The description of the basic steps in the assembly highlighted that the assembler needs to conceptualize objects within the given context. This context is normally provided by the knowledge about the goal structure. It is reasonable to assume that different amounts of information about the nature of the goal representation influences the process of conceptualization and thereby the overall assembly performance.

The influence of prior knowledge on performance has been addressed in a number of studies in linguistics. Research in the field of comprehension, for example, highlights the influence of prior knowledge on comprehension as well as recall of information. In experimental studies on text comprehension, Bransford and Johnson (1972:718) showed that “Ss [subjects] create semantic products that are a joint function of input information and prior knowledge”. They observed that participants who were presented with a picture that was related to the meaning expressed in the test passage performed better on comprehension and recall tests than participants who were given only partial or no semantic context. In follow up studies they also showed that

> “prior knowledge of a situation does not guarantee its usefulness for comprehension. In order for prior knowledge to aid comprehension, it must become an activated semantic context (… and) for maximum benefit the appropriate information must be present during the ongoing process of comprehension.” (Bransford & Johnson 1972:724)

This research highlights that activated prior knowledge may help to create contextual structures that facilitate understanding and recall of immediate input. Bransford and Johnson (1972) suggest that their findings are not limited to sentence processing but may also hold for other kinds of “present input events” (Bransford & Johnson 1972:725). Kintsch’s (1988) construction-integration model supports this hypothesis. The model proposes that first associations that are activated by linguistic clues activate associated concepts in turn. These associations are rated and integrated into a consistent model representing the current context. As already pointed out by Bransford and Johnson (1972), the activated mental concepts may facilitate or deteriorate task conceptualization, i.e. the task specific mental representation, and performance.

In a different series of experiments on text comprehension, Dixon investigated the structure of action plans derived from instructions. Dixon (1972; 1987a) proposed that mental plans are hierarchically structured by action information, i.e. which action to perform, at the top level and condition information, i.e. description of the current state, on the lower level. Furthermore, Dixon (1987b) showed that prior knowledge about the current state had a significant effect on reading time. If participants were provided with information about the current state prior to reading the instruction, they were faster in reading than when no such information was provided. Thus he concludes that prior knowledge crucially influences the construction of mental plans. According to Dixon, information about the current state is needed to select the appropriate schema[80] and precise actions are only subsequently incorporated. If condition information is provided before the actions are specified, these actions can be integrated during reading whereas rereading is necessary if the information is provided in the instruction only and needs to be extracted while reading.

The findings in these studies suggest that performance, such as comprehension and recall of information as well as reading time, is influenced by prior knowledge. Prior knowledge can be different in nature; it may be semantic knowledge as described so far, knowledge about procedures, or perceptual information. Procedural knowledge consists of sequences of actions that need to be performed to reach a specific goal. Making coffee will probably be a routine action for most people who enjoy to start their day with a mug of fresh coffee. Dixon et al. (1997) observed that participants tend to apply routine[81] procedures when new tasks show some similarity to previously encountered tasks. This is in line with analogical problem solving.

Perceptual information can also activate prior knowledge in terms of cognitive categories. It is generally assumed that objects are mentally organized in categories of similar objects. Ungerer and Schmid (2006) argue that this categorization is either culturally determined or based on obvious ways of grouping objects based on their apparent relation to each other (e.g. similarity in shape, function, color). It is proposed that these *cognitive concepts*[82] and the different contexts[83] in which they are activated are stored as *cognitive models* in the human mind (Ungerer & Schmid 2006:49). Thereby a network of categories and cognitive models evolves which is constantly activated and extended when new objects are encountered. It is further argued that each interaction between objects in the real world activates the mental concepts of the present objects (Ungerer & Schmid 2006:48). The activation of cognitive concepts and associated features also resembles a knowledge base that can be defined in terms of prior knowledge. This effect,

[80] In his experiments, Dixon investigated reading time and performance on tasks that followed a basic schema. Devices that needed to be manipulated were either on or off in the current situation, i.e. prior to the participant's manipulation.

[81] Routine is defined as the memory of "sequences of steps that accomplish task subprocedures." (Dixon et al. 1997:392)

[82] Ungerer and Schmid (2006:40) argue that categorization as well as producing and understanding language involves cognitive processes thus they refer to the emerging conceptual categories as *cognitive concepts*.

[83] *Context* is defined as "the cognitive representation of the interaction between cognitive categories." (Ungerer & Schmid 2006:58)

commonly referred to as *priming*, is well known and studied in experimental psychology (e.g. Rosch 1975).

The reported findings suggest that prior knowledge influences mental plan construction as well as the general categories that are activated. Plan construction is assumed to be influenced on the temporal, i.e. duration of plan construction, as well as on the structural level, i.e. which actions to perform in which order. The activation of associated cognitive categories may facilitate the conceptualization process because it raises expectations regarding the function of specific parts. However, in the same line of thought, activation of associated concepts may hinder the assembly process if misleading expectations are raised.

4.2.1.3 Influence of communicative intention on verbalization

Besides a manual, other people may also serve as a source of background information. In cases in which two people work on an assembly task collectively, the solution is worked out in a joint effort. Normally, people would use language to exchange ideas about possible functions of objects and they may also manipulate objects collectively. A research group from MIT headed by Kneeper presented a very interesting project at this years' IEEE International Conference on Robotics and Automation. They introduced their KUKA youBots, i.e. two autonomous robots that collectively assembled an IKEA coffee table without being shown the instructions in advance (for more detail see Knepper et al. 2013). In this scenario, the robots were coordinated by software (see Knepper 2013:5) but if two humans work on a problem collectively, they are very likely to use language to communicate. Since verbal as well as physical interaction is a fundamental part of daily life, researchers have been interested in the influence of a partner on verbalization for long. Tomasello (1998) proposeses that "people create symbols in order to use them as instruments in acts of communication – and these communicative acts *always* involve another person as recipient" (Tomasello 1998:230; emphasize added).

An overview on the literature studying the influence of an imagined or physically present addressee on referential form is provided in the analysis chapter on object reference (chapter 8); thus, only main findings will be highlighted in this overview. The influence of communicative intention on discourse structure and content has been extensively researched in studies conducted within the referential communication paradigm (e.g. Bavelas et al. 2000; Clark 2004; Clark 2005). Researchers found evidence that speakers tailor their utterances for their addressee at different levels. Considerations about the addressee's knowledge state are expressed at the syntactic level (e.g. Arnold 2009), the semantic level (e.g. Garrod & Anderson 1987; Gundel et al. 1993; Heller et al. 2012) as well as on the discourse level (e.g. Auer 1984; Clark & Krych 2004; Tenbrink et al. 2008). This phenomenon is commonly referred to as *audience design*. However, based on a literature overview, Schober and Brennan (2003) caution that the results may not be straight forward because

> "people can be shown to adapt under some circumstances and not to adapt under others at virtually every level of language use – from higher discourse-level functions to articulation" (Schober & Brennan 2003: 155).

In order to test how and to which degree people tailor their utterances for an addressee, researchers devised various experimental designs. In some studies researchers investigated this influence in an implicit way by asking participants to imagine an addressee or in a direct way, i.e. in real interaction. In studies with an imagined addressee, participants were either asked to write instructions (e.g. Lovelace et al. 1999; Hölscher et al. 2011) or to speak them out loud while their speech was recorded by a tape recorder (e.g. Levelt 1981; Habel 1988). In studies investigating interaction, participants either talked to a partner who was physically present (e.g. Horton & Gerrig 2002; Clark & Krych 2004) or audible only (e.g. Krauss & Weinheimer 1966; Schober & Clark 1989; Fukumura & van Gompel 2012). Some studies also analyzed real-life interaction and conversations (Schegloff 1979; Steffensen 2012) as opposed to those elicited in experimental settings.

One integral part of interaction while collectively solving a task is feedback about current states, proposed ideas, or future actions. Some studies investigated the influence of different kinds of feedback, such as concurrent feedback vs. no feedback, on language choices. Krauss and Weinheimer (1966), for example, studied the influence of the addressees' feedback on verbalization, more specifically the length of the reference phrase, in a matching task. They tested participants in a non-feedback condition in which there was no interaction between instructor and addressee. This data was compared to a set of data collected in an interactive scenario in which participants were allowed to communicate freely. Their results revealed that concurrent feedback has a significant effect on participants' verbalization, i.e. participants who received concurrent feedback used significantly fewer words than participants who did not receive feedback.

Similarly, Schober (1993) compared which perspective participants adapted when describing the location of an object for a real addressee or an imagined one. Instructions for an imaginary addressee that were recorded on tape were compared to descriptions in an interactive situation. His results highlighted that participants used different perspectives for an imaginary addressee then with an actual partner.

The results revealed the same trend as observed by Krauss and Weinheimer (1966); participants who interacted with a partner used significantly fewer words than participants describing for an imagined addressee. Moreover, the analysis of adapted perspective revealed that participants who interacted with a partner started off by adapting an egocentric perspective, i.e. "on my left" (Schober 1993:1) and negotiated this perspective with their partner.

Although there may be doubts as to the universal character of audience design, numerous studies provided evidence that participants tailor their utterance for their addressee, i.e. addressee-oriented speech. This prominent discourse type will be compared to language produced without the presence of an addressee and no explicit communicative motivation, i.e. self-oriented speech, throughout this thesis. As outlined in the theoretical discussion in chapter 8, it is argued that think aloud protocols, as introduced in chapter 3, represent self-oriented speech. Furthermore, the results reported by Krauss and Weinheimer (1966) and Schober (1993) highlight the influence of an imagined or a present addressee on verbalization.

4.2.2 Data collection: verbal reports

The analyses in this thesis are based on verbal protocols that were elicited at different times during the experimental phase. First, concurrent verbal reports were recorded while participants solved the assembly task (*think aloud protocols*). Second, participants were instructed to recall their thoughts immediately after they indicated that the task was completed (*retrospective reports*). Third, these free language production tasks were supplemented by a number of guided questions regarding specific features of the provided objects (*interview questions*).

Think aloud protocols were recorded by instructing participants to say out loud everything that comes to their mind while working on the task. According to Ericsson and Simon's (1984 (1993)) model of *verbalization of thinking*, participants' thoughts that are at the level of awareness and in focus of attention are recorded in think aloud protocols. Therefore, think aloud protocols are proposed to contain traces of mental processes (for more detail see chapter 3). These records of concurrent verbalization are elicited to study the mental processes that are involved in unaided object assembly and their linguistic representation. Think aloud protocols are specifically suited to study mental processes over time because those are continuously verbalized therein. However, think aloud protocols are not complete because only those processes that reach awareness are verbalized. Therefore, a second type of verbal report is elicited in this study, namely retrospective reports.

Retrospective reports were elicited immediately after participants indicated that they finished the task. Participants were asked to say everything that they remembered and to indicate if they were uncertain in their memories. As discussed in chapter 3, retrospective reports are assumed to contain more meta-level information, such as generalizations about procedures, reasons for specific actions and decisions, or conclusions. The recall of information was expected to be facilitated by providing the assembled goal structure as a visual cue. Supplementing information reflecting processes at the level of current attention with information that is generalized from these processes, was expected to yield a comprehensive picture of processes involved.

However, as it is acknowledged that perception might occasionally be too fast to reach attention and thus be verbalized in think aloud protocols (e.g. Ericsson & Simon 1993), some guided questions were asked at the end of the experimental session. These questions asked if participants noticed specific features that were either salient but not crucial for a successful assembly, i.e. boreholes, or not salient but facilitating for the assembly, i.e. grooves, or non-salient and not crucial for a successful assembly, i.e. the manufacturer's label. The combination of these three kinds of verbal reports was assumed to provide a rich amount of information about mental processes involved in unaided object assembly and it allowed for the analysis of the linguistic representation of these processes across two kinds of discourse tasks.

A third discourse task was included by collecting instructions on the same assembly task. This discourse task is different from the previous ones because it involves a direct addressee. This data was collected in a second experiment with participants who did not participate in the self-assembly experiment in which think aloud protocols and retrospective reports were collected.

In the conducted experiment, the addressee was introduced to the participant who provided an oral instruction. However, instructor and assembler, i.e. addressee, could not communicate with each other; a description of the experimental set-up is provided in section 4.4.2. Instructions that are tailored for an addressee were expected to contain meta-level information concerning the sequential order of actions, descriptions of these actions, and possibly reasons for taking those actions. In this experiment, instructors assembled the goal structure themselves prior to the instructions; thus, they were experts on the assembly. It was assumed that reflections on the personal assembly might also be verbalized in instructions.

The information that was contained in the verbal protocols was evaluated in a qualitative analysis whereas performance measures, such as assembly time and assembly success, were evaluated quantitatively. The qualitative analysis resulted in different categories to describe observed linguistic phenomena. The distribution of these categories was evaluated by descriptive and inferential statistics; an overview on the qualitative and quantitative analyses is provided in section 4.5.2 of this chapter. The following section presents information on the participants that participated in the two experimental studies before the experimental design is described.

4.3 Participants

4.3.1 Experiment 1: Unaided object assembly

56 participants were invited for the experiment. Their age ranged between 19 and 42 years (M = 24.02, SD = 5.06). They were all students at the University of Bremen and received course credit for compensation. 25 of the 56 participants were male and 31 participants were female.

Due to technical problems, two participants' speech was not recorded (one in the *underspecified goal* and one in the *verbal goal* condition). Three participants did not finish the experiment (one participant in the *underspecified goal* and two participants in the *verbal and visual goal* condition). One participant knew the dollhouse from a previous experiment (*verbal goal* condition). This left 50 data sets to be analyzed. Those 50 participants were between 19 and 42 years (M = 24.22, SD = 5.25). There were more female participants (28 women) than male participants (22 men) in the final data set for the analysis. Table 4.1 provides an overview of the distribution of age and sex between the three experimental conditions.

condition	age (mean)	age (SD)	female participants	male participants	sum participants
underspecified goal	24.00	5.70	12	5	17
verbal goal	24.88	5.58	6	10	16
verbal and visual goal	23.81	5.81	10	7	17

Table 4.1: Summary of participants' age and sex in the three experimental conditions.

4.3.2 Experiment 2: Instructions

20 participants were invited for the experiment. Their age ranged between 21 and 28 years (M = 22.7, SD = 1.98). They were all students at the University of Bremen and received course credit or monetary compensation for their attendance. Nine of the 20 participants were male and eleven participants were female. Due to technical problems with three data sets[84] and one participant who did not reach the expert status, thus 16 data sets remained for the analysis. The age of those 16 participants ranged between 21 and 28 years (M = 22.9, SD = 1.91). There were more male (nine men) than female participants (seven women) in the final data set for the analysis.

4.4 Design

This section provides an overview on the experimental design, procedure, and the equipment that was used. Furthermore, problems with the technical equipment will be reported. The section is divided into two sub-sections. The first sub-section provides information on the first experiment on unaided object assembly and the second sub-section presents information about the second experiment on instructions. The presentation of both experimental designs follows the same structure:

1) the design is presented,
2) the instructions are restated,
3) the material is presented,
4) a detailed account of the procedure is provided, and
5) information about equipment and technical challenges is given.

4.4.1 Experiment 1: Unaided object assembly

4.4.1.1 Design

The first experiment was designed to test the hypothesis that problem solving processes are involved in an assembly task. In order to create a problem solving scenario, an object has been selected that can be assumed to be relatively unfamiliar to the participants, i.e. an object with which they do not interact on a daily basis. The initial idea to choose a bookshelf sold by the furniture store IKEA was inspired by research reported by Tversky and colleagues who asked

[84] More detail about the technical problems is provided in 4.4.2.5.

participants to assemble a TV stand (e.g. Tversky et al. 2009; Daniel & Tversky 2012). This idea was not pursued for two reasons. First, each assembly procedure leaves marks at the material that might be recognized by the next participants and influence their assembly decisions. Second, most bookshelves that consist of more than six parts are too big to be assembled by one participant alone. Both of these reasons are crucial because the intention was to create a situation in which each participant needed to solve the problem from scratch under controlled conditions and without a partner.

Based on these considerations, a wooden dollhouse by the German brand 'Selecta Spielzeug' was selected. In order to constrain the number of hints and to treat the objects with care, no screws were provided. Nevertheless, it was possible to arrange the objects in such a way that a stable dollhouse is assembled. In terms of problem solving types, this assembly task posses a *transformation problem* because known procedures need to be rearranged in order to create an appropriate solution path (Greeno 1978) (for more detail see chapter 2).

The influence of prior information on performance was studied by testing participants in three conditions. In the first condition participants were presented with *unspecific goal* information when they were asked to assemble a sensible object. In the second condition, participants were provided with a verbal clue about the nature of the goal object (*verbal goal condition*). In the instructions, they were provided with the basic level term describing the goal object, i.e. 'two-story dollhouse'. This decision was based on Tversky and Hemenway's (1984) finding that "the basic level is the highest level of abstraction for which a generalized outline form can be recognized and the highest level for which an image can be generated" (Tversky & Hemenway 1984:186). Ungerer and Schmid (2006:88) argue that 'house' is a basic level term that can be activated as a super-ordinate term. They argue that this is the case when the term 'house' is decomposed into rooms serving different functions (e.g. living room, bedroom, kitchen). Although the term 'dollhouse' is more specific, it is still on the basic level and it can be decomposed into the same functional structures, i.e. rooms, as any kind of house. In the third condition, participants were additionally shown a black-and-white picture of the assembled dollhouse for 30 seconds (*verbal and visual goal condition*).

The different conditions resembled different kinds of transformation problems and participants needed to overcome different barriers in Dörner's (1987) sense[85]. Problem solvers were faced with a dialectic barrier in situations in which the initial

[85] See Table 4.2 for a summary and chapter 2 for more detail.

state and the operators were known but the goal state was unknown. The goal state needs to be defined during the dialectic process that is characterized by hypotheses, antitheses, and syntheses (Dörner 1987). In the presented assembly problem, the goal state was known but ill defined since no information about the nature of the goal object was provided. If the problem solver knows the initial state, the goal state, and the general operations that need to be applied, he/she faces an interpolation barrier. This barrier needs to be overcome by creating an appropriate combination or hierarchy of necessary operations (Dörner 1987).

condition	Greeno (1978)	Dörner (1987)
underspecified goal	transformation problem with an ill-defined goal state	dialectic problem
verbal goal	transformation problem with a well-defined goal state (based on a verbal cue activating background knowledge)	interpolation barrier
verbal and visual goal	transformation problem with a precisely defined goal state (based on verbal & visual cues)	interpolation barrier

Table 4.2: Kinds of transformation problems and problem barriers that are faced by the participants depending on the experimental condition.

The general procedure was tested in a pilot study. In this study, all objects that were needed to assemble the two-story dollhouse were lying on a table. Four participants were tested in the condition of the *underspecified goal condition*. The pilot study revealed some limitations of the design. First, it was observed that participants needed a large table in order to have enough space to arrange and manipulate the objects. Second, the experimenter observed that participants did not notice all object parts displayed on the table. Participants, rather, showed a tendency to start assembling right away and noticing objects during the assembly or not at all. Both limitations have been overcome in the experimental phase. The space problem has been addressed by choosing a conference room in which a large table could be used and no other distractors, such as bookshelves, were present. In order to facilitate object perception, all small objects were placed in a cardboard box that was placed on top of the table (see Figure 4.1). This solution is supported by Kirsh's (1995) observation that

> "an agent need not register all the actions feasible in a situation, the action set which is perceived as feasible (the perceived action set) is sensitive to the situation, particularly arrangement." (Kirsh 1995:43)

The solution with the box did not only foster perception of objects but it also prevented biasing participants by placing objects into some pattern that might be interpreted as meaningful presorting. The experimenter paid careful attention to put the objects into the box in the same order. The order did not suggest any obvious grouping.

Figure 4.1: Arrangement of experimental material as encountered by participants entering the conference room.

4.4.1.2 Instruction

condition	German original	English translation	additional material
underspeci-fied goal	„**Bauen** Sie alle auf dem Tisch liegenden und die in der Kiste befindlichen Teile **sinnvoll zusammen** ohne zu schrauben. Die Kiste soll nicht benutzt werden.“	“Assemble a **sensible object** using all parts lying on the table and those contained in the box without using screws. Do not use the box.”	none
verbal goal	„Bauen Sie alle auf dem Tisch liegenden und die in der Kiste befindlichen Teile sinnvoll zu einem **zweistöckigen Puppenhaus** zusammen ohne zu schrauben. Die Kiste soll nicht benutzt werden.“	“Assemble a **two-story dollhouse** using all parts lying on the table and those contained in the box without using screws. Do not use the box.”	none
verbal and visual goal	„Bauen Sie alle auf dem Tisch liegenden und die in der Kiste befindlichen Teile sinnvoll zu dem **abgebildeten Puppenhaus** zusammen ohne zu schrauben. Die Kiste soll nicht benutzt werden.“	“Assemble the **pictured dollhouse** using all parts lying on the table and those contained in the box without using screws. Do not use the box.”	picture of the assembled dollhouse (see Figure 4.2)

Table 4.3: Original instruction texts in German and its English translation.

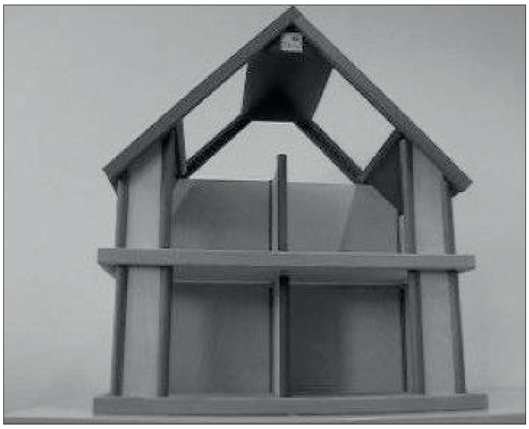


Figure 4.2: Black and white picture of the two-story dollhouse that was presented to participants in the *verbal and visual goal* condition.

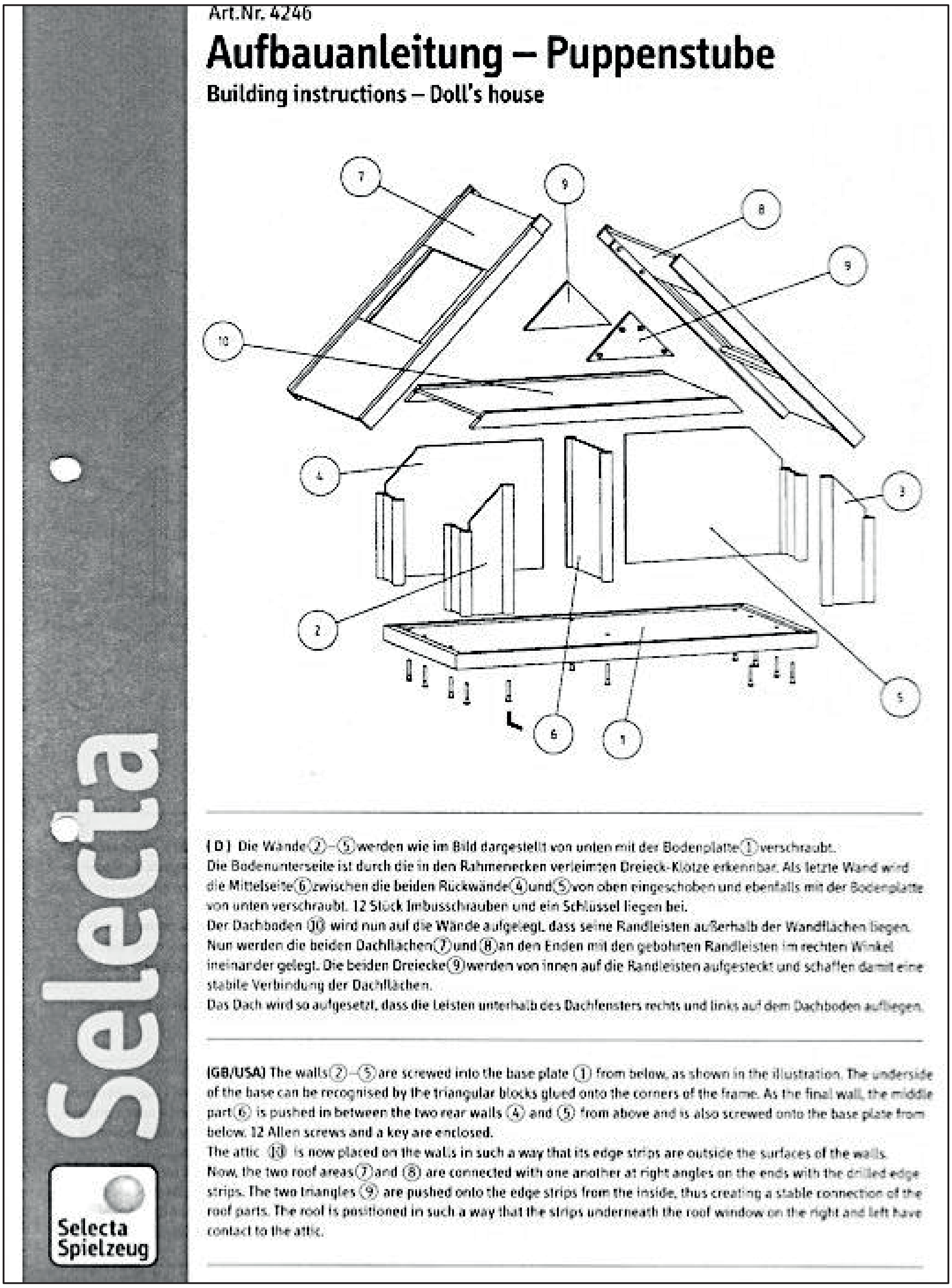
4.4.1.3 Material

The goal object that participants were asked to assemble was a two-story dollhouse. It is a child's toy produced by the German brand "Selecta Spielzeug" (see Figure 4.3 for an advertisement[86]). Normally, the object pieces come with a manual, as displayed in Figure 4.4, describing how to assemble the pieces. However, as motivated above, participants were not given the manual and they did not have screws to fasten objects. The schematic drawing will be used on the following description to facilitate understanding of the general layout of the dollhouse.

Figure 4.3: Advertisements of the dollhouse 4246 by 'Selecta Spielzeug'[87].

[86] 'Selecta Spielzeug' has been contacted in March 2012 asking for permission to use this specific picture. Consent was given by Mrs. Kordula Harting.

[87] (http://www.selecta-spielzeug.de/index.php/selecta_en/produkte/kleine_welt/ puppen-haeuser/puppenhaus; 20.04.12)


Art.Nr. 4246

Aufbauanleitung – Puppenstube

Building instructions – Doll's house

(D) Die Wände ②–⑤ werden wie im Bild dargestellt von unten mit der Bodenplatte ① verschraubt.
Die Bodenunterseite ist durch die in den Rahmenecken verleimten Dreieck-Klötze erkennbar. Als letzte Wand wird die Mittelseite ⑥ zwischen die beiden Rückwände ④ und ⑤ von oben eingeschoben und ebenfalls mit der Bodenplatte von unten verschraubt. 12 Stück Imbusschrauben und ein Schlüssel liegen bei.
Der Dachboden ⑩ wird nun auf die Wände aufgelegt, dass seine Randleisten außerhalb der Wandflächen liegen.
Nun werden die beiden Dachflächen ⑦ und ⑧ an den Enden mit den gebohrten Randleisten im rechten Winkel ineinander gelegt. Die beiden Dreiecke ⑨ werden von innen auf die Randleisten aufgesteckt und schaffen damit eine stabile Verbindung der Dachflächen.
Das Dach wird so aufgesetzt, dass die Leisten unterhalb des Dachfensters rechts und links auf dem Dachboden aufliegen.

(GB/USA) The walls ②–⑤ are screwed into the base plate ① from below, as shown in the illustration. The underside of the base can be recognised by the triangular blocks glued onto the corners of the frame. As the final wall, the middle part ⑥ is pushed in between the two rear walls ④ and ⑤ from above and is also screwed onto the base plate from below. 12 Allen screws and a key are enclosed.
The attic ⑩ is now placed on the walls in such a way that its edge strips are outside the surfaces of the walls.
Now, the two roof areas ⑦ and ⑧ are connected with one another at right angles on the ends with the drilled edge strips. The two triangles ⑨ are pushed onto the edge strips from the inside, thus creating a stable connection of the roof parts. The roof is positioned in such a way that the strips underneath the roof window on the right and left have contact to the attic.

Figure 4.4: Screenshot of the assembly instruction for the dollhouse 4246 provided by Selecta Spielzeug.

In the schematic drawing used in the assembly instructions (see Figure 4.4), the different parts that need to be assembled are displayed. Nine wooden parts are marked in the drawing. In the experimental condition, the roof was already assembled. This left five parts that were small (parts 2-6 in the assembly instruction corresponding to Figure 4.6 to Figure 4.9), a larger board (marked 1 in the instructions corresponding to Figure 4.5), and the ready assembled roof part (see Figure 4.10). In addition to this attic, the dollhouse that was assembled in the experiments had a first story as well (see Figure 4.11). This story consisted of the same parts labeled 1-6 in the instruction but without bevels (Figure 4.12 to Figure 4.15 display the original parts).

Figure 4.5: Board corresponding to object 1 in the instruction.

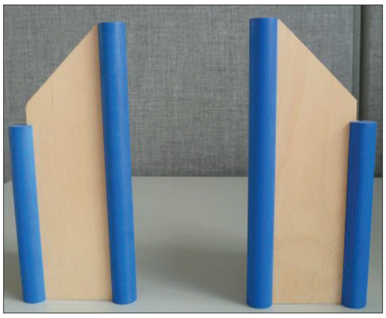


Figure 4.6: Wooden objects corresponding to objects 2 and 3 in the instruction.

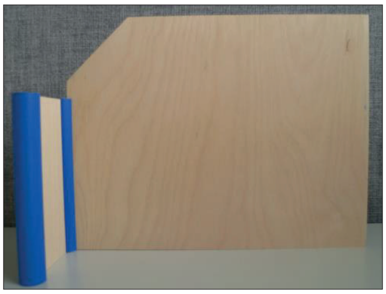


Figure 4.7: Wooden object corresponding to object 4 in the instruction.

Figure 4.8: Wooden object corresponding to object 5 in the instruction.

Figure 4.9: Wooden object corresponding to object 6 in the instruction.

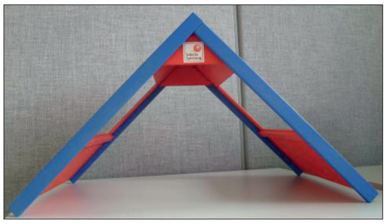


Figure 4.10: Roof part of the two-story doll-house.

Figure 4.11: Assembled two-story dollhouse.

Figure 4.12: Wooden objects for the first story corresponding to objects 2 and 3 in the instruction for the attic.

Figure 4.13: Wooden object for the first story corresponding to object 4 in the instruction for the attic.

Figure 4.14: Wooden object for the first story corresponding to object 5 in the instruction for the attic.

Figure 4.15: Wooden object for the first story corresponding to object 6 in the instruction for the attic.

4.4.1.4 Procedure

The procedure was divided into three parts, namely *introductory phase*, *experimental phase*, and *ending*. This structure is used in the description of the procedure as well.

In the *introductory phase*, the experimenter welcomed the participant in her office. She asked the participant to fill out a form of consent first. Then she told the participant about the general aim of the experiment by saying "Mit diesem Experiment wollen wir wissen, was Sie denken, während Sie die gestellte Aufgabe lösen." (In this experiment, we want to find out what you think while you solve the given task.) Then a general introduction was given stating that there were no time constraints and asking the participant to indicate when he/she is finished with the task.[88] It was important that the participant indicated the time at which he/she felt that the task was completed. Chi (1997) points out that experimenters may state that the task is completed and close the experimental session because they feel that the goal state is being reached but the participant may not share this evaluation. It is important to recognize that different goal states can be defined and in this experiment it was not the aim to reach the experimenter's goal state

[88] Original wording: „Es gibt keine zeitliche Begrenzung zur Ausführung der Aufgabe. Wenn Sie mit der Aufgabe fertig sind, sagen Sie das bitte." ("There is no time constraint in this task. If you are finished with the task please say so.")

but it was rather important to study the participant's representation of the goal state. Furthermore, the experimenter said that there was a video camera but that she was mainly interested in the speech recorded with a head set.[89]

After these general remarks the participant was instructed to think aloud while solving the task. Furthermore, he/she was told that the experimenter would be in the room but no interaction was possible.[90] As suggested by Ericsson and Simon (1993:377), the think aloud procedure was tested with two warm-up tasks. First, participants were asked to solve the equation 24×36.[91] Although most participants did not succeed in solving the equation correctly, the task was well suited to illustrate the difference between describing and think aloud. A second task was posed for participants who showed a tendency to describe rather than think aloud. In this task participants were asked to list as many foreign language teachers from their school time as possible. It was important that they mentioned their names as well as the subject that they taught.[92] If the participant felt content and the experimenter had the impression that the participant knew the difference between think aloud and describing, the participant was informed about the general procedure that was to follow.[93] Finally, the experimenter asked if there were any remaining questions regarding the experiment. In order to ensure that the content as well as the wording of the information provided in the *introductory phase* was

[89] Original wording: „Die Videokamera ist nur zur Sicherheit da, so können wir später einordnen, was Sie getan haben während des Sprechens. Uns geht es aber um die Sprache, die mit dem Headset aufgezeichnet wird." ("The video camera is only there to record what you are doing while you talk in case we want to know what you did. But we are mainly interested in the speech recorded with the head set.")

[90] Original wording: „Während Sie die Aufgabe lösen, bitte ich Sie, laut zu denken, das heißt Sie sagen alles, was Ihnen durch den Kopf geht. Ich werde nicht mit Ihnen sprechen. Ich bin nur als Zuhörer im Raum. Sollten Sie eine Weile nicht laut gedacht haben, werde ich Sie kurz erinnern. Ich werde dann sagen ‚Sprechen Sie bitte weiter.'" ("I ask you to think aloud while you are solving the task; this means that you say everything that comes to your mind. I will not talk to you. I will be in the room as a listener. If you are not thinking aloud for some time I will remind you. I will say 'Keep talking, please.'")

[91] This example is suggested by Ericsson and Simon (1993:378).

[92] This example is similar to the one proposed by Ericsson and Simon (1993:378) who suggest to ask participants "How many windows are there in your parent's house?". Since the experiment involves assembling a house, this example cannot be used because of a possible priming effect. Therefore, I looked for a question that each participant can relate to and which involves retrieval of information from memory. The question about foreign language teachers worked very well given the purpose to motivate participants to talk freely and verbalize all emerging thoughts.

[93] Original wording: „Wir gehen jetzt in einen anderen Raum, in dem das Experiment stattfindet. Ich werde Ihnen die Aufgabe vor der Tür stellen, gehe in den Raum, um die Geräte einzuschalten und bitte Sie laut zu denken, sobald Sie den Raum betreten." ("We will go to another room in which the experiment will take place. I will give the task instructions to you in front of the door. Then I will go into the room to start all technical devices. Please think aloud as soon as you enter the room.")

the same for each participant, the text was written down and read to the participants.

After the introduction the experimenter and the participant went to a conference room that was located on the same floor as the office. The experimenter read out the instructions of one of the three conditions in front of the door. The instructions were given to the participant before they went into the room to ensure that they did not have first associations before the experimental phase started. Additionally, this procedure ensured that the starting point of the experiment was the same for each participant. In the *verbal and visual goal* condition the picture was given to the participant who was asked to return it after 30 seconds. Then the experimenter went into the room, made sure to close the door, and started the recording devices (the video camera and the mini disk recorder).

The *experimental phase* started when the participant entered the room. During the assembly the experimenter was sitting at the back of the room and did not move or talk. If the participant kept quiet for longer than approximately 1 minute, the experimenter reminded him/her to keep talking. When the participant indicated that he/she finished the task, the experimenter walked over to him/her. Then the experimenter asked the participant to recall what he/she thought during the task and to indicate when he/she was uncertain about these memories.[94] The assembled object was still present on the table to assist the participant's memory. After this *retrospective report* the experimenter posed four guided questions addressing specific features of the assembly process. Table 4.4 provides the wording of the questions that were asked in the *short interview*.

After the participant answered the last interview question, the experimental phase was over. Then the experimenter turned off all recording devices and thanked the participant. If there were any questions about the experiment, the experimenter answered them. Then the participants received their course credit. After accompanying the participant to the door, the experimenter closed the experimental session by noting down information about the recordings and observations that she made during the assembly procedure. Furthermore, she took

[94] Original wording: „Nachdem Sie die Aufgabe erfolgreich gelöst haben, bitte ich Sie zu berichten, was Sie gedacht haben, während Sie die Aufgabe lösten. Falls Sie sich in Ihren Erinnerungen unsicher sind, sagen Sie das bitte. Wenn es Ihnen möglich ist, berichten Sie die Gedanken in der Reihenfolge, in der Sie Ihnen während des Zusammenbauens kamen." ("After you have successfully solved the task I ask you to please recall what you thought while solving the task. If you are uncertain about your memories please say so. If possible, tell your thoughts in the same order as they came to your mind during the assembly.")

pictures of the assembled house before arranging the objects in the initial order to start a new experiment.

condition	questions (original wording)	questions (translation)
all conditions	1) Haben Sie das **Stecksystem** wahrgenommen? Diente es Ihnen zur Orientierung? 2) Haben Sie die **Bohrlöcher** wahrgenommen? Dienten sie Ihnen zur Orientierung? 3) Haben Sie das **Label "Selecta Spielzeug"** wahrgenommen und in Ihre Überlegungen mit einbezogen?	1) Did you notice the connecting system? Did it guide your thoughts? 2) Did you notice the boreholes? Did they guide your thoughts? 3) Did you notice the label "Selecta Spielzeug" and did it influence your ideas?
underspecified goal	4) Wie kamen Sie auf die Idee, ein Haus zu bauen?	How did you have the idea to assemble a house?
verbal goal	4) An welche Art Puppenhaus dachten Sie als ich die Aufgabe stellte? Hat sich dies verändert nachdem Sie die Teile gesehen haben?	What kind of dollhouse came to your mind when I told you about the task? Did it change after you saw the given objects?
verbal and visual goal	4) Hatten Sie das gesehene Bild während des Aufbaus im Kopf? Veränderte es sich während des Aufbaus? Wenn ja, inwiefern?	Did you remember the picture during the assembly? Did it change during the assembly? If so in which way?

Table 4.4: Questions asked in the short interview in the respective experimental conditions.

4.4.1.5 Equipment and technical challenges

A video camera (Panasonic Mini DV, Model No. NV-GS27) was used to record the visual scene. It was fastened on a tripod and placed in a corner of the room to make sure that it did not constrain the participants in their movements. All audio files were recorded with a Sony MZ-RH10 mini disk recorder. For the first audio recordings a head-set was used. Participants carried the head-set around their neck and the mini disk recorder in their pockets. This arrangement caused some difficulties in cases in which participants did not have any pockets. This problem was solved by using a microphone that was attached to the mini disk player because this allowed to place the mini disk recorder near the table. In this set-up, the participants were not constrained by the equipment at all.

The audio files were transcribed using the free software f4[95]. The video recordings were converted into digital format by the Medienstelle at the University of Bremen.

[95] Available at http://www.audiotranskription.de/f4.htm.

The equipment and experimental design posed some challenges. First, some recordings could not be properly transcribed because the participants did not speak loud enough or the objects were placed in a manner that created much background noise. In some cases, the noise could be filtered out to some extent. In two cases the video camera did not work and in two additional cases the video tape was too short to record the entire *experimental phase*.

4.4.2 Experiment 2: Instructions

4.4.2.1 Design

The second experiment was designed as a follow up experiment. To study the influence of communicative intention it aimed at collecting data in which the speaker produced speech for a partner. As outline in sub-section 4.2.1.3, there are different levels that need to be distinguished when eliciting speech produced for an addressee. First, the researcher needs to decide about the *partner's presence* in the experiment. The partner may either be physically present or not. Second, the participant's knowledge about the *partner's knowledge state* needs to be considered. The participant may not know anything about the partner or the participants are told that their partner was an expert or has the same background knowledge. In the first case, the participant is likely to imagine a generic addressee whereas in the second case he/she is likely to imagine a specific addressee. The participant's assumed knowledge state about the partner is very important to the researcher if results by different participants are compared. Third, the possibility of *feedback* needs to be considered. Partner and participant may either be allowed to talk freely (concurrent feedback), not to talk at all (no feedback), or feedback is provided after the task is finished (delayed feedback). To the best of my knowledge, the influence of the later kind of feedback has not been tested empirically so far.

In the present study participants were asked to instruct a partner verbally on the assembly of the two-story dollhouse. In order to compare the results from the two experiments, it was necessary to devise an experiment in which the conditions are not too different from the first experiment. Thus the partner was not physically present in the room. However, in order to ensure that each instructor has the same addressee in mind, a specific addressee was introduced. Since there was no partner involved in the first experiment, the non-feedback scenario was selected for the second experiment. In order to motivate participants to provide good instructions, a delayed feedback was offered. Although no feedback was intended, preliminary analyses of the think aloud protocols revealed that the objects at hand were difficult to describe; therefore visual support was offered providing a shared workspace. Based on these considerations, it is proposed that the comparison between think aloud protocols and instructions allows to study the influence of communicative intention on verbalization.

In order to instruct the partner on the assembly, the instructors needed to know how to assemble the dollhouse themselves. There are two different ways of acquiring this knowledge; participants are either shown how to assemble the object or they assemble it themselves. In the second case, the assembly can be guided by a manual or unaided. Daniel and Tversky (2012) took this approach by asking their participants to assemble a TV cart providing them with the photograph of the completed TV cart on the box that normally contained all object parts. As the first experiment focused on verbalization in unaided assembly participants in the second experiment did not receive any external help either. Kiefer et al. (1993) distinguish two kinds of prior knowledge, namely action-oriented prior knowledge and static prior knowledge. In order to investigate the influence of the nature of prior-knowledge on reference production in an assembly task, they varied the procedure in which the instructors acquired knowledge about the assembly object in two conditions. In the first condition, a video of the assembly process was shown to the participants before asking them to assemble the object themselves[96] (*action-oriented prior knowledge*). In the second condition, participants were shown a picture of the assembled object only (*static prior knowledge*)[97]. In their study, Kiefer et al. (1993:21) report a significant effect of type of prior knowledge on nominal specificity used by instructors. Participants who assembled the object themselves prior to the instruction, produced more specific references than participants who were shown the picture only. Based on these findings, Mangold-Allwinn et al. (1995) assume that participants who interacted with the objects and assembled it themselves, created a richer representation of the different parts and the assembly structure during the assembly phase. This additional knowledge about object features, such as grooves in some screws (Kiefer et al. 1993:19), is expressed in more specific nominal references in instructions as compared to participants who acquired a static representation.

[96] The description of the experimental procedure is very short in Kiefer et al. (1993:19). It was difficult to find exact information about the study design. Apparently one part of the data was collected in an experiment that is described in Mangold-Allwinn et al. (1992). In this experiment assembly instructions were collected from participants who were provided with action-oriented prior knowledge.

[97] The terminology for the two kinds of prior-knowledge is not mentioned in Kiefer et al. (1993:19) but is introduced afterwards in Mangold-Allwinn et al. (1995).

Given that participants in the first experiment were recorded while assembling the dollhouse, i.e. during the acquisition of action-oriented prior knowledge, instructors needed to gain the same expert status in the second study. Although the majority of participants, who were provided with few prior information about the goal state, succeeded in assembling a house, the frequency of original dollhouses was much higher if participants were provided with much prior information (for more detail see the results on success reported in chapter 5). Based on this finding, instructors were provided with much prior information (*verbal and visual goal* condition) in order to assure that instructors succeeded in assembling the original dollhouse within few trials. The *knowledge acquisition phase* was divided into two parts, namely the *explorative assembly phase* and the *gaining expert status phase*. In the first phase, participants acquired action-based knowledge about this specific dollhouse assembly and in the second phase, they strengthened this knowledge. The second phase was introduced to make sure that instructors knew the individual assembly steps and could thus concentrate on the instructions afterwards. Success in the *knowledge acquisition phase* was the prerequisite for proceeding with the second task, i.e. the instruction. Additionally, it was assumed that quick success kept participants motivated to solve a second task.

Based on these careful considerations, it can be assumed that the design ensured that all instructors were provided with the same knowledge about the exact structure of the goal object, about the specific addressee, and about the general assembly steps.

4.4.2.2 Instructions

As outlined above, the experiment contained two sub-tasks. First, the participants were asked to assemble the dollhouse for themselves (referred to as *self-assembly task*). The instructions were identical to the instructions in the *verbal and visual goal* condition in the first experiment on unaided object assembly. However, the participant assembled the house silently while the experimenter was waiting outside the room. Second, participants were asked to verbally instruct a partner to assemble the same dollhouse (referred to as *instruction*). The partner, who was in another room, was able to see and hear the instructor via the Skype interface but no interaction was possible. The exact instructions were:

> „Das ist Lena. Wie Sie sehen, liegen hier dieselben Puppenhausteile noch einmal bereit. Wenn die Webcam angeschaltet ist, kann Lena Sie auf dem Bildschirm sehen. Wir gehen gleich zurück in den Versuchsraum. Dort bitte ich Sie, das Puppenhaus noch einmal aufzubauen und Lena dabei

> mündlich zu erklären, wie sie mit den Teilen hier das Gleiche tun kann. Über diese Webcam kann Lena Ihrer Beschreibung folgen. Sie baut also dasselbe Puppenhaus zusammen wie Sie. Allerdings kann sie keine Rückfragen stellen. Deswegen ist es wichtig, dass Sie alles genau erklären und beschreiben. Nach dem Versuch können Sie sehen, wie erfolgreich Ihre Beschreibung war, nach der Lena das Haus aufgebaut hat. Weil es uns vor allem um die Sprachdaten geht, haben wir einen MD-Spieler in dem Raum liegen.“ (“This is Lena. As you can see the same dollhouse parts are lying here as well. When the webcam is on Lena can see you on her screen. We will go back to the conference room in a minute. When we are back I ask you to assemble the dollhouse once more and instruct Lena on how to do the same with her object parts. Lena can follow your instructions via this webcam. She will assemble the same dollhouse as you do. But she cannot give any feedback hence it is important that your instructions and descriptions are accurate. After the experiment you can assess your success in instructing Lena. We are primarily interested in the verbal data therefore we have placed a MD player in the room.”)

The partner, who was always called Lena, was also addressed with an instruction, namely

> „Wir gehen jetzt in den anderen Raum zurück und ich stelle die Webcam an. Geben Sie mir ein Zeichen, wenn Sie uns auf dem Bildschirm sehen? Dann bauen Sie das Puppenhaus ganz genau so auf, wie es Ihnen von [Name des Teilnehmers] erklärt wird. Okay?“ (“We will go back to the other room now and I will switch on the webcam. Can you give me a sign when you see us on the computer screen? Then you will assemble the house exactly how it is described to you by [participant’s name]. Okay?”)

The partner’s task in the experiment was to follow the instructions and assemble the dollhouse. Since this task was not crucial to the experiment, the partner was a confederate. It was the same confederate in all experimental sessions.

4.4.2.3 Material

The same two-story dollhouse produced by ‘Selecta Spielzeug’ was used in this experiment. Therefore, the material was the same as described in section 4.4.1.3 Material. However, each object part was there twice, i.e. once in the conference room in which the instructor worked and once in the laboratory in which the partner assembled the dollhouse. The objects for the partner were not contained in a box in order to save time for the partner. It was assumed that it was easier for the partner to identify the objects when they were lying on the table.

As described for the *verbal and visual goal* condition, the black and white picture (see Figure 4.2 on page 110) of the ready assembled dollhouse was shown to the instructor in the self-assembly task.

4.4.2.4 Procedure

As mentioned above, the experiment consisted of two sub-tasks referred to as *self-assembly task* and *instruction*. The self-assembly task was further divided into an *explorative assembly phase* and the *gaining expert status phase*. However, before the experimental phase started, a general introduction was given. Experimenter and participant were in the experimenter's office for the *introductory phase*. First, the experimenter asked the participant to fill out a form of consent. Then a general overview of the procedure was given. Additionally, the participants were informed that there were no time constraints and that they needed to open the door to indicate that they had finished the task.[98] All remaining questions were answered.

Then the participant started with the *self-assembly task*. The participant and the experimenter walked to the conference room that was the same room as in experiment 1 (*unaided object assembly*). In front of the door the experimenter read out the instructions and handed the picture to the participant. Then she started the video camera inside the conference room. After 30 seconds the participant handed back the picture and entered the room. The experimenter stayed outside and started the time recording when the door closed. This was the beginning of the *explorative assembly phase*. After the participant opened the door indicating that he/she was finished, the experimenter entered the room as well. She evaluated the assembled house and either gave a positive feedback or an encouraging one. The positive feedback for participants, who were assigned the *advanced status* read as follows:

> „Sehr gut, Sie haben das Puppenhaus original getreu nachgebaut. Bitte bauen Sie es noch ein weiteres Mal zusammen, denn als nächstes werden Sie eine andere Person bei dem Bau des Puppenhauses mündlich anleiten.“ (“Well done. You have assembled the original dollhouse. Please assemble it once more since you will instruct another person on the assembly next.”)

Then, the experimenter and the participant left the room. The experimenter went back in and disassembled the house placing all objects back in the box. Entering

[98] Original wording: „Das Experiment findet in einem anderen Raum statt. Wir werden Sie während der Lösung der Aufgabe mit einer Videokamera aufzeichnen. Bitte verhalten Sie sich dennoch so natürlich wie möglich. Ich werde Ihnen die Aufgabe vor Betreten des Raumes stellen. Dann gehe ich in den Raum, um die Kamera anzustellen und bitte Sie anschließend hinein. Es gibt keine zeitliche Begrenzung. Wenn Sie die Aufgabe gelöst haben, öffnen Sie bitte die Tür.“ (“The experiment will take place in another room. We will video tape you while you solve the task. Nevertheless, I ask you to behave as natural as possible. I will read the task to you in front of the door. Then I will go into the room to start the video camera and ask you to come in afterwards. There are no time constraints. When you solved the task please open the door.”)

the room a second time, the *gaining expert status phase* started. The experimenter started the time recording when the door closed. She waited until the participant opened the door once more to indicate that the second assembly was finished as well. The encouraging feedback for participants who were assigned the *learner status* read as follows:

> „Ihr Puppenhaus entspricht leider noch nicht dem abgebildeten Puppenhaus. Ich zeige Ihnen noch einmal das Bild und stelle Ihnen dieselbe Aufgabe.“ (“Unfortunately your dollhouse does not resemble the pictured dollhouse yet. I will show the picture to you and read the same task to you once more.”)

Then the experimenter and the participant left the room. The experimenter handed the picture to the participant and disassembled the house placing all objects back in the box. After 30 seconds the experimenter took the picture back and the participant entered the room once more. The experimenter started the time recording again. She waited until the participant opened the door once more to indicate that the second assembly was finished. If the newly assembled dollhouse resembled the pictured one, the participant was assigned the *advanced status* and the procedure with positive feedback started. If the participant was assigned *learner status* again, he/she got one more opportunity with the encouraging feedback procedure. If no expert status was reached after a third *explorative assembly phase*, the experimenter did not proceed with the experiment.

After the expert status was gained, the participant faced the second task, i.e. instructing another participant. After positive feedback about the performance in the self-assembly phase, the experimenter pointed out that there was a webcam in addition to the previously introduced video camera. Then the participant was informed that the webcam was connected to a computer in another room in which another participant was waiting.[99] Then the experimenter and the participant walked to the other room that was located in the same hallway. After introducing the participant to the partner, the experimenter read out their task that mainly addressed the instructor. Then the experimenter read the instructions to Lena. If there were no questions concerning the procedure, the experimenter and the participant went back to the conference room. The experimenter switched on the webcam and connected the laptop to the computer in the laboratory while the participant disassembled the dollhouse. While the connection between the two

[99] Original wording: „Sie wurden bisher mit dieser Kamera aufgezeichnet. Es gibt hier im Raum aber auch noch diese Webcam. Wenn Sie angeschaltet ist, überträgt Sie in unser Labor. Dort wartet Lena.“ (“So far it was this camera that recorded your actions. But there is also this webcam in this room. If it is turned on, it is transmitting to a computer in our laboratory. Lena is waiting over there.”)

computers was established, the partner, i.e. the confederate, also started the video recording software (more detail in the next sub-section) that recorded the visual interface of Skype. When the partner indicated that the connection was fine, the experimenter left the room and the participant started the instruction. The experimenter waited outside the room until the participant opened the door to indicate that the instruction task was finished. The confederate wrote down the instruction time and stopped the video recording. After the MD player was turned off, the experiment was finished. In the end, participants received their course credit or monetary compensation and left the office.

4.4.2.5 Equipment and technical challenges

A video camera (Panasonic Mini DV, Model No. NV-GS27) was used to record the visual scene. It was fastened on a tripod and placed in a corner of the room to make sure that it did not constrain the participants in their movements. All audio files were recorded with a Sony MZ-RH10 mini disk recorder and a microphone attached to the recorder.

The program *Debut Audio Recording Software 1.42*[100] was used to record the visual Skype interface. The computer that was used by the confederate ran on Windows whereas the laptop that was positioned in the instructor's room ran on Mac OS X 10.6.8.

The audio files were transcribed using the free software f4[101]. The video recordings were converted into digital format by the Medienstelle at the University of Bremen.

Although the equipment and the software worked well in most cases, occassionally difficulties were faced concerning the recordings. Difficulties were encountered with the audio recording in two cases and with the video recording in one case.

4.5 Analysis

This section provides a summary of the quantitative and qualitative measures that were collected in the experiments and the way in which these are analyzed in the

[100] Available at: http://www.zdnet.de/debut_video_recording_software_download-39002345-63892-1.htm

[101] Available at: http://www.audiotranskription.de/f4.htm

different chapters of this thesis. No results or detailed descriptions of the analyses will be presented.

4.5.1 Data sets for analysis

Due to technical problems and individual difficulties with the task, six data sets were lost for the analysis. Furthermore, the data for four participants was not complete, i.e. audio and video recording. Table 4.5 provides a summary on the data sets in experiment 1 (*unaided object assembly*) in which some difficulties were faced, specifying the reasons. This table serves as a reference that will be pointed to when the data sets for the respective analyses are reported.

reason	number of participants	description	scope of loss of data
personal	1	The participant knew the doll house from a previous experiment.	complete
personal	3	They gave up and did not complete the task.	complete
personal	1	The participant said that he said everything he thought during the assembly hence he refused to provide a retrospective report.	partial (retrospective report)
technical	2	No audio and video recordings.	complete
technical	1	The video tape was too short.	partial (video data)
technical	2	The video recording did not work.	partial (video data)

Table 4.5: Summary of partial and complete data loss due to technical and personal reasons in experiment 1.

Due to technical and personal difficulties, four data sets were completely lost in the second experiment. Additionally, the video recordings of five videos could not be analyzed because the DVD was damaged and the original tapes had been re-used already. Table 4.6 provides a summary on the data sets in experiment 2 (*instructions*) in which some difficulties were faced, specifying the reasons.

reason	number of participants	description	scope of loss of data
personal	1	The participant did not reach the expert status.	complete
technical	3	The audio and video recording did not	complete

		work.	
technical	5	The DVD that contained the video recordings was damaged.	partial

Table 4.6: Summary of partial and complete data loss due to technical and personal reasons in experiment 2.

4.5.2 Qualitative and quantitative analysis

The basis for all of these studies is an in-depth qualitative analysis of the recorded verbal reports. In order to analyze the recorded speech, the data needs to be properly transcribed. Ericsson and Simon (1993) point out that "the verbalizations of adults can generally be transcribed into ordinary words" (Ericsson & Simon 1993:279). But as language encodes information at different levels (such as content, intonation, and pauses), decisions about the level of granularity need to be made prior to the transcription process. Generally, these decisions depend on the focus of the respective researcher. Nevertheless, Ericsson and Simon (1993) point out that transcripts should include as much information as possible so that other researchers can use the same data set to investigate new question (for a discussion of this view see chapter 3). In the case of the presented experiments the transcriptions needed to encode the semantic content expressed by words, pauses, and hesitation markers such as 'uh', 'uhm', 'well'. Regarding pauses the transcription was based on Ericsson and Simon (1993) who state that "only the duration of longer pauses need to be recorded, since these may indicate periods when the subject is not obeying the instruction to verbalize" (Ericsson & Simon 1993:279). This was operationalized in the experimental phase by reminding participants to keep thinking aloud if they fell silent for longer than one minute.

In order to analyze the information that is contained in the transcripts, the speech needs to be properly coded. The coding of data is subjective because the person analyzing the data defines what needs to be annotated and how it is annotated. In the present studies, I made the decisions about which categories were annotated and how. The annotation was done by one assistant and myself. There was one basic annotation scheme from which aspects have been selected and which was extended in the chapters focusing on different features of analysis, such as verbs and object references. In the basic annotation scheme each discourse unit was annotated with regard to nominal reference, more specifically nominal specificity, article, deictic references (anaphoric, exophoric, and ellipsis), verbs, and clause expansions. The categories of clause expansion were based on Halliday and Matthiessen (2004). Although the aspect of clause expansion was not analyzed in this thesis the intercoder agreement will be reported for completion. More detail on the annotation of features that are analyzed, such as nominal phrases, will be outlined in the respective chapter. The annotation scheme for nominal phrases including nominal reference, articles, and deictic reference is outlined in more detail in chapter 8. More detail about the annotation of verbs is provided in chapter 7.

The intercoder agreement for the general scheme was tested between two coders, who annotated six protocols each. These six protocols resemble 12.0% of the corpus and they contain 627 discourse units; this represents 13.7% of all annotated discourse units. During the training, also referred to as *pilot test* (Lombard et al. 2002:601), the greatest challenge was the segmentation of protocols into discourse units. After feedback and further training, this difficulty was overcome.

After the training the percent agreement for this basic scheme was very high, i.e. 98.8%. Krippendorff's alpha for the individual annotation categories was very high in most cases.[102] The lowest Krippendorff's alpha value was 0.84 (see Table 4.7). According to Neuendorf (2002:143) "coefficients of .90 or greater would be acceptable to all, .80 or greater would be acceptable to most situations, and below that there exists disagreement". Thus the obtained coefficients indicate that there was no significant disagreement between the coders.

annotation category	percent agreement	Krippendorff's Alpha
nouns unspecific	99.0%	0.958
nouns specific	100.0%	1
nouns constraints	99.0%	0.837
personal pronoun	99.8%	0.996
ellipsis of nouns	100.0%	1
articles definite	100.0%	1
articles demonstrative	100.0%	1
articles indefinite	100.0%	1
verbs all	100.0%	1
verbs modal	100.0%	1
ellipsis of verbs	100.0%	1
ellipsis of articles	100.0%	1
reference exophoric	99.8%	0.996

[102] Krippendorff's Alpha has been calculated using the tool ReCal which is available on the internet. (http://dfreelon.org/utils/recalfront/; last access 29.03.2013).

reference anaphoric	100.0%	1
ellipsis of reference	100.0%	1
elaboration: exposition	100.0%	1
elaboration: exemplification	100.0%	1
elaboration: clarification	100.0%	1
extension: addition	100.0%	1
extension: variation	100.0%	1
extension: alternation	100.0%	1
enhancement: temporal	100.0%	1
enhancement: spatial	100.0%	1
enhancement: manner	100.0%	1
enhancement: causal	100.0%	1

Table 4.7: Percent agreement and Krippendorff's Alpha for annotation categories within the general annotation scheme.

Insights on the general nature and structure of problem solving processes in unaided object assembly can be gained in a qualitative analysis of the general structure of think aloud protocols. On the most general level these protocols can be assumed to consist of an introductory sequence, a middle part, and a concluding statement. The content of each of these structural parts will be described, categorized, and analyzed. The middle part is assumed to contain the actual assembly activity and this part will be studied in terms of problem solving processes, i.e. known problem solving processes will be identified and new processes will be described. By means of the general frequency of the individual processes as well as recurring sequences of these processes, a general structure of problem solving in unaided object assembly will be described (see chapter 6). In a more fine-grained analysis the linguistic representation that is characteristic of the individual problem solving processes will be highlighted. In the analysis, the focus will be on the verb phrase, specifically modality and verb type, but general patterns of discourse markers will be highlighted as well.

The influence of prior information on the development of the mental representation of the goal object can be studied by analyzing object references. Object references reflect the speaker's object conceptualization, thus referential change throughout the assembly recorded in think aloud protocols will be analyzed (chapter 8). The influence of prior information on the problem solving process itself is investigated by analyzing the use and assigned function of task specific constraints, such as physical properties of objects, in think aloud protocols (chapter 9). In both chapters these findings will be contrasted with results obtained in instructions. This comparison highlights the influence of communicative intention on the development of the mental representation and the nature and function of constraints in unaided object assembly.

Additionally, the design offered a number of quantitative measures such as time, success, and length of verbal reports. More specifically assembly times were recorded in different conditions. First, experiment 2 (*instruction*) provided a baseline condition because participants did not think aloud while assembling the dollhouse. The task encountered by participants during the *explorative assembly phase* in experiment 2 and by participants in the *verbal and visual goal* condition was the same. The only difference was that participants in experiment 1 (*unaided object assembly*) were posed the additional task to think aloud. Second, times were recorded for participants who needed to perform the same task, i.e. assemble a dollhouse, but with different amounts of prior information about the goal object (experiment 1 *unaided object assembly*). A common measure besides time is length of verbal protocols; the weaknesses of protocol length are addressed in the next chapter that reports the evaluation of performance measures.

In addition to these mere quantitative measures, the frequency of the categories that are defined in the qualitative analysis will be statistically evaluated, i.e. in terms of descriptive and inferential statistics in the individual chapters.

5 Features of analysis: Performance measures

5.1 Introduction

Hussy (1993:23) discusses the advantages and disadvantages of measures that are not easily quantifiable, such as information about a person's feelings, and quantitative ones, such as performance time. He summarizes that verbal protocols are subjective in nature containing much information about participants' reasoning, their decisions, and other mental processes at work. However, as verbal reports are not quantitative[103], their content is not easily comparable between participants. Therefore, many researchers conclude their theoretical considerations by proposing to study the multifaceted nature of human problem solving by combining qualitative and quantitative research methods and measures (Schoenfeld 1985; Hussy 1993; Funke & Spering 2006).

[103] Quantitative methods are objective in the sense that their truth-value can be assessed (Funke 2003:21).

Although this thesis focuses on verbal protocol analysis, quantitative measures were collected as well. First, assembly and instruction times were recorded. Second, the assembled dollhouse could be evaluated in terms of success and the verbal protocols allowed assessing the experimenter's reminders in the task. Furthermore, the comparison between performance times and assembly success in a condition in which participants assembled the dollhouse silently and the same measures in a comparable condition in which participants thought aloud highlights whether verbalization has an effect or not. Additionally, it is investigated if the action-oriented assembly task encourages participants' spontaneous verbalization. Although there is no substantial body of evidence for self-talk[104], Merz (1969:131) reports that students tended to speak to themselves while solving math problems. This observation is supported by Goffman (1981) who proposes that we "speak to ourselves judgmentally about our own doings (...), and verbally mark junctures in our physical doings." (Goffman 1981:79) Goffman's (1981) assumption is based on self-observation and hearsay. Given that the investigated task involves much physical action, it may encourage spontaneous self-talk. In addition to recording the experimental conditions in which participants were instructed to think aloud, Dickson et al. (2000) audio taped the experimental session in which participants worked silently as well. This approach is unique, at least to my knowledge. The analysis of the tapes revealed that no participant spoke during the experiment when they were instructed to work silently (Dickson et al. 2000:222).

A different observation is reported in Cowley and Nash (2013). They report that a number of participants working on the river problem, which is also known as Hobbits and Orcs problem, started talking aloud without the instruction to do so. It is important to note that these participants were soldiers and their officer was also present during the trial. Cowley and Nash (2013) analyzed the transcript of one participant in detail and the extracts that are reported show that this specific talk aloud protocol resembles a think aloud protocol. It can be assumed that the occurrence of this very unique case of informative talk aloud is due to the fact that the officer was present which motivated the soldier to perform well and hope for encouragement along the way[105]. Additionally, this specific soldier is very likely to be highly fluent in verbalizing his thoughts and he apparently does not feel constrained by the test situation.

[104] Goffman (1981:79) defines self-talk as speaking to oneself in cases in which no other person is present either addressing oneself or an imaginary addressee.

[105] Cowley and Nash (2013) report that the soldier glanced at the officer after he made a suggestion about a possible solution.

Each of these performance measures (performance time and individual success) and possible influencing factors (the experimenter's reminders and the additional task to think aloud) will be discussed from a theoretical point of view in the following paragraphs. Based on this literature overview, research questions and expectations will be stated before the analysis procedure is described. After presenting the results, those will be discussed in the final section.

Funke (2003:20-21) differentiates between qualitative and quantitative methods and measures in problem solving research. Quantitative measures of behavior are objective when their truth-value can be assessed. Common quantitative performance measures are solution times (e.g. Knoblich et al. 1999; Funke & Spering 2006; Schelhorn et al. 2007) and individual success, either in the form of appropriate solutions (e.g. Merz 1969; Roth 1985; Gralla et al. 2012) or solution frequencies (e.g. Knoblich et al. 1999; Knoblich et al. 2001). Funke and Spering (2006:689) point out that solution times are the sum of durations of different subprocesses such as perceiving, decisions to act, and acting. Therefore, solution time encodes the overall performance time but it does not provide any information about individual processes that are involved in performing the task. In order to gain insights on the underlying cognitive processes, additional information such as verbal protocols or eye-movement needs to be considered as well. In eye-tracking studies, fixation rates and sequences, as well as pupil diameter are recorded that can provide insights about where attention is focused and for how long. Focus and duration of attention can be analyzed in terms of information processing, i.e. which information is processed and how (Funke & Spering 2006:691). In a similar fashion cursor movements and clicks on the keyboard (e.g. Ritter & Larkin 1994) can be analyzed.

Qualitative measures, in contrast, are subjective because their truth-value cannot easily be assessed. Funke (2003) argues that verbal reports are subjective because participants may consciously or sub-consciously change their thoughts before verbalizing them. Schoenfeld (1985:282) summarizes that participants may be influenced by different aspects of the experimental setting, such as the pressure of being recorded leading to the strain to say something regardless of whether it is true or not, the participant's belief about the nature of the experiment, and beliefs about the discipline itself. Concerning problem solving research, Schoenfeld (1985) points out that the participant's assumptions about the nature of the experiment may guide his/her choice of solution methods because "certain methods are considered 'legitimate' for solving problems in a formal setting, others not" (Schoenfeld 1985:282).

If quantitative measures are considered, the general question regarding the influence of think aloud on task performance needs to be raised and discussed. There are two levels that may be influenced by the task to verbalize everything that comes to the participants' mind, namely the *processing level* and the *performance level*. Summarizing the literature review offered in chapter 3, it can be assumed that think aloud does not influence mental processes at the *processing level* if participants are instructed to say out loud everything that comes to their mind. If participants are instructed to verbalize specific information, such as reasons for actions or choices, an influence on mental processing is assumed because information needs to be selected.

Studies focusing on the influence of think aloud on the *performance level* yield different results. Merz (1969) and Franz and Merz (1976) report several studies in which solution correctness was significantly facilitated by concurrent verbalization whereas solution times were longer than in the silent condition. Short et al.'s (1991) studies revealed that verbalization facilitated performance on spatial and verbal analogy tasks for fifth graders. However, no such effect was observed in their adult test group. Schooler et al. (1993) investigated the influence of concurrent verbalization on performance on insight and non-insight problems. An insight problem, according to Schooler et al. (1993), is characterized by three features, namely

> "(a) is well within the competence of the average subject; (b) has a high probability of leading to an impass (…); and (c) has a high probability of rewarding sustained effort with an "Aha" experience." (Schooler et al. 1993:168)

In their study no effect of verbalization on performance on non-insight problems was observed but a negative effect on performance on insight problems (for a discussion of these results see chapter 3). Another study revealed effects of think aloud on parts of the problem solving process as recorded in patterns of eye movement and fixation rates (Knoblich & Rhenius 1995). However, no significant effect of verbalization was observed on performance time. In sum, most studies report that participants thinking aloud took longer to solve the task than the silent control group. Success was influenced positively in some tasks but hindered in others. To date, only mental manipulation tasks have been studied but no investigation of performance in physical manipulation tasks was reported.

It is well known that people speak at different rates depending on speech context, i.e. if participants are nervous they tend to speak more and faster, but also on personal preferences, i.e. some people speak before they think whereas other people think about what they want to say for quite some time before actually saying it out loud. Additionally, people differ in the quantity of verbalization (Dominowski 1998:30). Thus van Someren et al. (1994) stress that "two protocols may differ in length not because of differences in thought processes, but because of different styles of verbalization" (van Someren et al. 1994:135). Ericsson and Simon (1993) address this effect as well. Reviewing a number of studies they propose that "some subjects may transform automatically, whereas others may require a concious effort for this transformation and verbal production" (Ericsson & Simon 1993:88). They assume that concurrent verbalization is easier for participants who think verbally than participants who prefer a visual code. The additional effort of transforming the visual code into a verbal one is assumed to result in fewer spontaneous verbalizations. At the level of training, Ericsson and Simon (1993:250) argue that those warm-up tasks that are based on oral information for which all participants are likely to produce concurrent reports[106] can help to reduce individual differences and foster fluent verbal reports. At the level of analysis, Chi (1997) proposes that the effect can be overcome by analyzing what people say rather than how much they say. More specifically,

> "this means that one would not count the number of words a person has spoken as an index of the amount of elaboration, for example, but use a more appropriate measure such as the number of independent ideas generated." (Chi 1997:306)

In the current study, the number of different problem solving processes that are generated by each participant would be a more appropriate measure in Chi's terms. However, the focus in this thesis is not on individual performance in terms of problem solving behavior but on the overall pattern of problem solving in an unaided assembly task. This issue is addressed in much detail in chapter 6.

Ericsson and Simon (1993) as well as Chi (1997) stress the importance of reminding participants to keep thinking aloud. It can be argued that reminders help to reduce individual differences because participants are continuously reminded to speak. This effect can only be gained if the experimenter has a clear guideline for reminding participants, e.g. if the participant remains silent for one minute. Importantly, Chi (1997:305) stresses that the problem of individual differences does not only exist for verbal protocol analysis but also in different terms for other kinds of dependent measures. Concluding this review, protocol length will be reported in this chapter on performance measures for completeness although it does not allow for valid conclusions on its own.

---

[106] They propose mental multiplication and anagrams (Ericsson & Simon 1993:250).

5.2 Research questions and expectations

Based on the literature review discussed in the previous section and a more detailed review in chapter 3, six research questions can be investigated by analyzing quantitative performance measures, namely assembly time and assembly success, and contextual factors, i.e. experimenter's reminders and additional task to think aloud.

- Research question 1: Does the additional task of verbalization have an influence on assembly time?

Based on the different findings reported in the literature, the additional task of verbalization may either have no effect on assembly time or slow assembly down. But it needs to be kept in mind that these findings are not based on tasks involving physical manipulation but on those requiring mental manipulation, e.g. match or reasoning.

- Research question 2: Does verbalization have an influence on individual assembly success?

In studies investigating mental problems, such as math equations, verbalization has been observed to have a positive effect. However, as pointed out before, there are no previous studies investigating problems involving physical manipulation. Thus, no specific expectations concerning the influence of verbalization on object assembly were formulated.

- Research question 3: Does prior information have an influence on assembly time?

As described in detail in chapter 4, Bransford and Johnson (1972) investigated the influence of prior knowledge on comprehension and recall. Their results revealed a positive effect of prior knowledge on accuracy of recall. Additionally, Dixon (1987b), who studied the influence of prior knowledge on reading time, reported that participants who were provided with prior knowledge were significantly faster in reading and understanding instructions on a physics problem. He assumes that prior knowledge fosters the creation of a mental representation of the problem state as newly encountered information can be structured faster if a preliminary representation exists already.

Based on these findings, it was expected that prior information had a positive influence on assembly time. In line with Dixon's (1987b) findings, participants who were provided with much prior information were expected to solve the task faster than participants who needed to create the goal structure from scratch, i.e. without a preliminary mental representation.

- Research question 4: Does prior information have an influence on individual assembly success?

There were two ways of defining success in the investigated assembly task. On the one hand, success could be defined differently for each of the three conditions. In all three conditions the goal state was the measure of success but the goal state was defined differently in the instructions.

For participants in the *underspecified goal* condition, the goal state was a sensible object in which all objects were included. In the *verbal goal* condition, participants needed to assemble a two-story dollhouse using all object parts. And in the *verbal and visual goal* condition, participants were successful when the assembled house matched the pictured two-story dollhouse. Following this argumentation the condition-specific success rate needed to be evaluated and compared to the success rates in the other conditions. This evaluation will be referred to as *condition specific success*. On the other hand, the same goal state could be reached in all three conditions. If participants used the boreholes (see Figure 5.1) and other physical constraints such as bevels (see Figure 5.2) and grooves (see Figure 5.3), they were guided to assemble the original two-story dollhouse. Following this argumentation, success could be evaluated at the same scale in all three conditions. This view will be referred to as *condition unspecific success*.

Figure 5.1: One of the boards clearly displaying the boreholes.

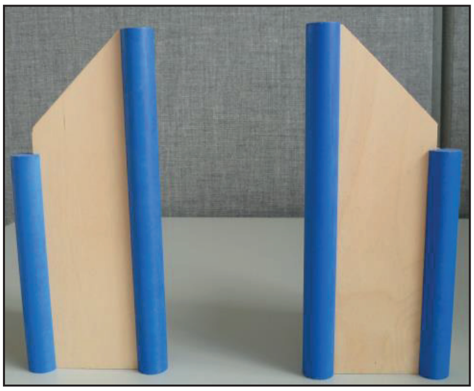


Figure 5.2: Two of the four objects with a bevel. Those objects need to be placed on the second story in order to position the roof part on top of them.

Figure 5.3: One of the two objects that have a notch on both sides. Other objects can be slid into these notches in order to be fastened.

On the other hand, evaluating the *condition specific success*, each participant had the same chance to succeed regardless of the specificity of prior information that was provided. However, evaluating the *condition unspecific success*, prior information was assumed to have a clear effect. Participants who saw the picture of the assembled dollhouse (*verbal and visual goal* condition) could be expected to succeed in assembling the original dollhouse more often than participants who were not provided with an external model. Participants who saw the picture were more likely to succeed because they could use either one or both of the following strategies: either matching the given objects to the memorized dollhouse or noticing and using the physical clues that were provided by the objects. Participants who did not know the exact nature of the dollhouse, in comparison, could only rely on the second strategy and were thus expected to be less likely to succeed.

- Research question 5: How often does the experimenter need to remind participants to keep thinking aloud?

Normally, reports about studies using think aloud protocols do not provide information on the number of reminders. Therefore, no literature based expectations could be formulated for this question. However, two alternative expectations were formulated based on general assumption about the influence of think aloud on task performance and the naturalness to think aloud.

First, based on Ericsson and Simon's (1993:85) postulate that verbalization does not influence basic performance measures and thought processes, it was assumed that the task to think aloud did not pose any extra effort on participants. Thus there should be no systematic differences in the number of reminders between conditions. Differences might rather be observed between participants because as Chi (1997:305) points out "some people are more verbose than others".

Second, an alternative expectation was also formulated. As "spontaneous thinking aloud is rare in everyday life" (Ericsson & Simon 1993:xiv) participants might forget verbalizing their thoughts when they were very engaged in the task or when they faced difficulties. If this was the case, participants in the *underspecified goal* and the *verbal goal* condition might need to be reminded more often assuming that their task of building a mental representation from scratch or matching the objects to an unspecific mental representation was more demanding than matching the given objects to an existing specific mental representation. Assuming that these participants were highly absorbed in performing the task and that they might encounter difficulties more frequently, they might forget to think aloud more frequently than participants in the *verbal and visual goal* condition.

- <u>Research question 6: Does the action-oriented nature of the assembly encourage spontaneous self-talk?</u>

As quoted above, Ericsson and Simon (1993) point out that spontaneous talk aloud can hardly be observed in everyday life. If people verbalize thought processes, this is often done in "social motivated verbalizations", for example by providing explanations (Ericsson & Simon 1993:xiv). However, based on personal experiences and hearsay as well as on Goffman's (1981) report, it is assumed that spontaneous self-talk can be witnessed in the investigated task. Goffman (1981:79) defines self-talk as speaking to oneself when no other person is present. This definition does not say anything about the form of self-talk, i.e. what is verbalized. The term self-talk seems to refer to verbalizations that are only sporadic and even less coherent than verbalizations when thinking aloud thus the analysis in this thesis focuses on instances of self-talk rather than think aloud, i.e. verbalization of trains of thought.

5.3 Analysis

5.3.1 Assembly time, instruction time, and protocol length

As noted above, there are different measures to assess performance. In the first experiment, performance time was recorded at different points. In the unaided assembly task *assembly time* was recorded, i.e. the time recording started when participants started the assembly and it ended when they indicated that they are done. In the second experiment, the main interest was in the instructions but the *assembly time* was recorded as well. The time recording started when the door closed behind the participant and it stopped when the participant opened the door to indicate that he/she was done (for more detail see section 4.4.1). Some participants needed more than one attempt to assemble the pictured dollhouse. Then more than one assembly time was recorded for that participant. Assembly time was also recorded for the last assembly prior to the instruction, referred to as *gaining expert status phase* (see section 4.4.2). Additionally, the time that participants took to instruct their partner was recorded. This temporal measure was referred to as *instruction time*. The time difference or similarity between participants in the *silent assembly* condition, i.e. without the task to think aloud, and in the equivalent *think aloud* condition can be assessed by comparing mean *assembly times* in both conditions. In order to ensure a high degree of comparability, the time recording for the first *explorative assembly* in the instruction scenario was compared to the assembly time recorded during unaided assembly. In both cases participants were provided with much prior information, i.e. the picture of the assembled dollhouse, they were given the same instructions, and they were presented with the same experimental setting. The only difference was in the task to think aloud during the assembly.

Altogether, 17 assembly times were analyzed in the *verbal and visual goal condition* (for more detail on participants in experiment 1 see section 4.3.1). For the instruction experiment all recorded assembly times were analyzed including those participants whose instructions could not be analyzed due to technical problems. This allowed for the analysis of 19 assembly times in experiment 2.

The number of words that were spoken during the assembly task might also highlight differences between conditions. Therefore, mean *protocol length* as well as the range of words that were recorded in the think aloud protocols was compared to investigate the influence of prior information on verbalization.

Protocol length was defined as the number of words that were recorded in the transcript. The length of all 50 think aloud protocols was assessed.

5.3.2 Assembly success

Whereas time and number of words are quantitative measures, *participant's success* is a qualitative performance measure. As such this measure can be quantified by rating the individual success of the assembled objects. For these means a 4-point scale was defined in the analysis. If participants assembled the original dollhouse, their assembled dollhouse was rated '3' (see Figure 5.4 and Figure 5.5). Participants who assembled a house which had two-stories (see Figure 5.6 and Figure 5.7) were rated '2'. Participants who assembled a one-story house (see Figure 5.8) were rated '1' in their success. Only those who built a one-story dollhouse by using all objects and arranging them in a meaningful way (see Figure 5.9) were rated '1'. Participants who failed to use all object parts were rated '0' which was equivalent to 'no success'. The participant who built the house displayed in Figure 5.10, for example, did not use the second board.

Figure 5.4: Assembled dollhouse that represents the original dollhouse [14]. [107]

Figure 5.5: Assembled dollhouse that represents the original dollhouse [17].

[107] The number in brackets refers to the participant's ID within the corpus. All of the pictures show houses assembled in experiment 1 (Unaided object assembly).

Figure 5.6: Assembled two-story house [16].

Figure 5.7: Assembled two-story house [12].

Figure 5.8: Assembled one-story house [52].

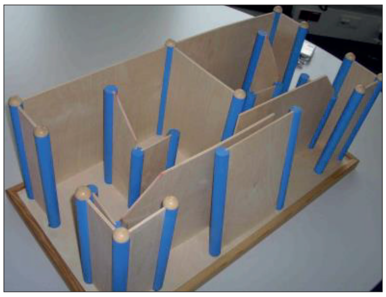


Figure 5.9: Configuration of objects in the assembled one-story house displayed in the previous picture [52].

Figure 5.10: Assembled house in which one board is missing [10].

5.3.3 Spontaneous verbalization and experimenter's reminder

Spontaneous instances of think aloud were annotated by watching the videos recorded during the *silent assembly phase* in the second experiment. The *first explorative assembly* (abbreviated by EA) was comparable to the condition faced by participants who were instructed to think aloud. These recordings were annotated by taking down all spontaneous utterances. Due to technical problems only eleven videos could be annotated (for more detail see chapter 4).

In order to capture the experimenter's involvement in the first experiment, the number of reminders was counted in the transcripts. For each participant the sum of reminders was taken down. Reminders could be easily extracted from the protocols since they were always phrased the same way, i.e. „Bitte sprechen Sie weiter." ("Please keep talking.") This phrase was introduced to the participants in the introductory phase of the first experiment (for more detail see section 4.4.1).

5.4 Results

5.4.1 Assembly time and protocol length

5.4.1.1 Assembly time

Participants who did not think aloud during the explorative assembly (EA) took between 3.11 minutes (fastest participant) and 20.17 minutes (longest assembly time) to finish their assembly; M = 8.72 minutes (SD = 4.42) (see Table 5.1). Participants who were instructed to think aloud during the assembly in the *verbal and visual goal* condition took between 3.04 minutes and 12.30 minutes; M = 6.22 minutes (SD = 2.15). The results revealed that participants who verbalized their thoughts during the assembly were significantly faster than participants who were silent, $t\ (26.67) = 2.20$, $p < .05$.

condition	minimum (minutes)	maximum (minutes)	mean (minutes)	standard deviation	range	N
underspecified goal	3.15	29.17	10.28	5.10	26.02	17
verbal goal	2.45	15.25	7.26	3.23	12.80	16
verbal and visual goal	3.04	12.30	6.22	2.15	9.26	17
1st explorative assembly	3.11	20.17	9.12	4.42	17.06	19

Table 5.1: Mean assembly time with maximum, minimum, standard deviation, and range; differentiated by condition (minutes).

Additionally, the results in experiment 1 revealed that prior information had a significant influence on assembly time. Participants who were provided with least specific goal information (*underspecified goal* condition) needed significantly longer to assemble the dollhouse (M = 10.28 minutes, SD = 5.10) than participants who were provided with specific background information (*verbal and visual goal* condition) (M = 6.22 minutes, SD = 2.15), $t\ (21.52) = 3.03$, $p < .01$. Participants who were told about the nature of the goal object (*verbal goal* condition) (M = 7.26, SD = 3.23) needed significantly less time for their assembly than participants who were provided no information about the goal object, $t\ (27.26) = 2.32$, $p < .05$. There was one outlier in each condition (see Figure 5.11).

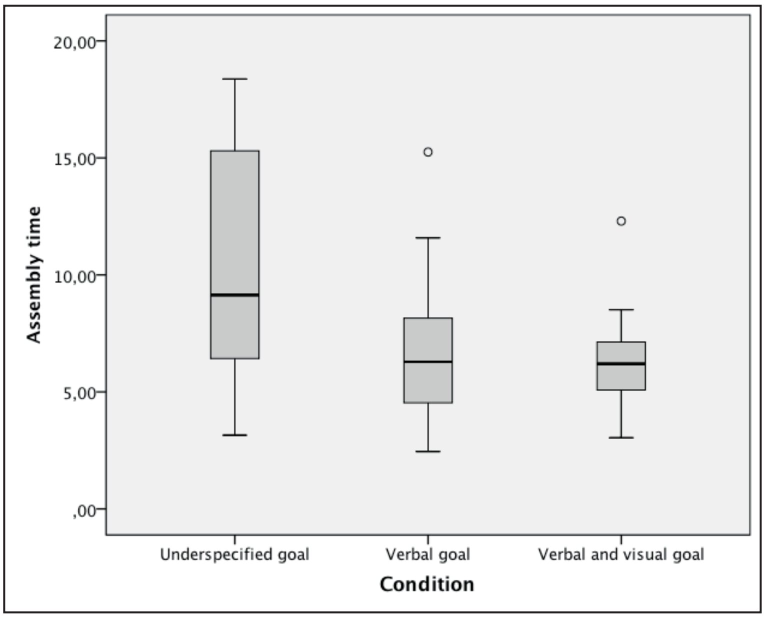


Figure 5.11: Assembly time (in minutes) with think aloud.

5.4.1.2 Protocol length and instruction time

5.4.1.2.1 Unaided object assembly

The shortest think aloud protocol was 104 words long and the longest one amounted to 2655 words. This was a clear outlier because it was 1490 words longer than the second longest protocol. Therefore, this protocol was excluded from the statistical analysis. Thus the longest protocol was 1165 words long; this resulted in a range of 1061 words between the shortest and longest think aloud protocol. The mean protocol length was 454.42 words (SD = 250.45).

The analysis revealed marked differences between conditions. Think aloud protocols by participants who were provided with *underspecified goal* information were longer than those by participants in the two conditions in which goal information was provided (see Table 5.2). Participants in the *verbal goal* and the *verbal and visual goal* condition used the same amount of words in their reports on average. The difference within condition, as expressed in the statistical measure *range*, was most pronounced in the two conditions in which participants were provided with little goal information, i.e. *underspecified goal* and *verbal goal* condition.

condition	minimum (words)	maximum (words)	mean (words)	standard deviation	range	N
underspecified goal	197	1165	548.35	296.98	968	17

verbal goal	104	1126	405.81	245.34	1022	16
verbal and visual goal	189	883	406.24	182.39	694	17

Table 5.2: Length of think aloud protocols by words with maximum, minimum, mean, standard deviation, and range differentiated by condition.

5.4.1.2.2 Instructions (protocol length and time)

The shortest instruction was 3 minutes long whereas the longest one took almost 10 minutes (9:53 minutes). On average, an instruction lasted 6:13 minutes (SD = 1.80). The transcripts of the instructions were between 319 and 984 words long, i.e. resulting in a range of 665 words. The mean length of the 16 protocols was 614.82 words (SD = 214.86).

5.4.2 Assembly success

5.4.2.1 Unaided assembly

Overall, only four participants did not succeed in solving the assembly task (8%). The majority of participants assembled a house (24 participants, 49%) or the original dollhouse (22 participants, 45%). Neither a picture of the assembled object nor a video is recorded for one participant (for an overview on missing data see sub-section 4.5.1).

Considering *condition specific success*[108], participants in the *underspecified goal* condition succeeded when they assembled an object using all objects. In terms of the rating scale this meant that either the assembly of a house or the original dollhouse expressed success. Participants in the *verbal goal* condition succeeded when they assembled a two-story dollhouse using all objects. In terms of the rating scale this meant that they succeeded when assembling a two-story house or the original dollhouse. Participants in the *verbal and visual goal* condition succeeded when they assembled the pictured dollhouse using all objects. In terms of the rating scale this meant that only the original dollhouse expressed success. Table 5.3 summarizes the raw frequency of each rating within conditions. Success rates (last column in Table 5.3) revealed that participants who were provided with most prior information were least successful.

condition	no success	one-story house	two-story house	original dollhouse	N	success (%)

[108] As described in the section on analysis this measure of success is rated in relation to the definition of the goal object as defined in the task instruction.

underspecified goal	3	3	7	4	17	82.35%
verbal goal	0	1	8	6	15	93.33%
verbal and visual goal	1	0	4	12	17	70.59%

Table 5.3: Assembled objects rated by success distribution within conditions (raw frequency).

Evaluating the *condition unspecific success*[109] the reverse trend showed. From this perspective prior information had a significant influence on assembly success, $L\chi^2$ (6, N = 49) = 11.68, p = .02. Participants who saw the picture of the assembled dollhouse (*verbal and visual goal* condition) succeed in assembling the original dollhouse more frequently than participants without this specific information (12 cases as compared to 4 cases) (see Figure 5.12). Participants who were provided with unspecific goal information assembled a house more often than participants who were provided with an external model (10 cases as compared to 4 cases) (see Figure 5.12). In this analysis standardized residuals did not reveal which dependent variable contributed most to the effect. Nevertheless, the effect was most evident in the frequency of original dollhouse assemblies in the *verbal and visual goal* condition with z = 1.6 (for more detail see Table 5.6 in the appendix).

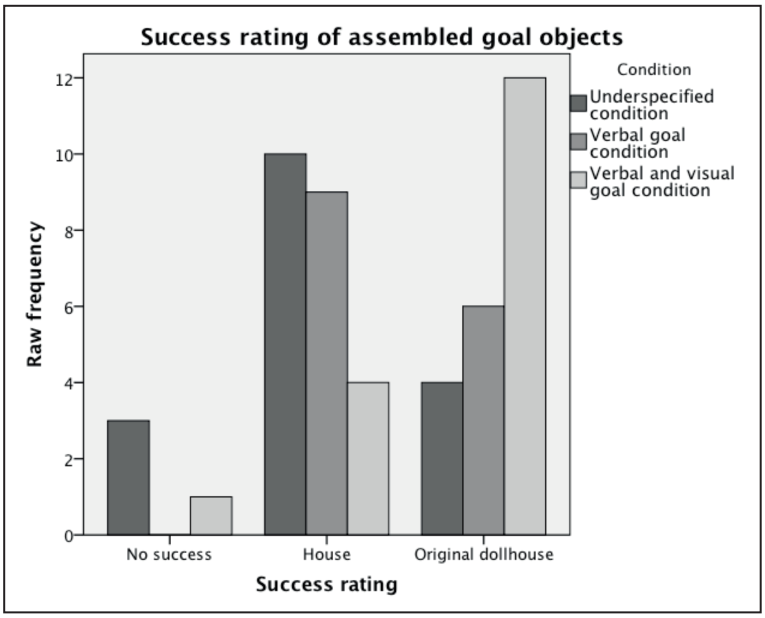


Figure 5.12: Distribution of categories of assembled products between conditions (raw frequency).

5.4.2.2 Instructions

Assembly success in first explorative assembly could be annotated for 15 participants. The majority of participants succeeded in assembling the original

[109] As outlined in the section on analysis this measure of success is rated in relation to the possibility to assemble the original dollhouse for all participants who paid attention to clues such as boreholes and physical constraints.

dollhouse (80.00%, 12 cases). Two participants did not to use all object parts and one participant assembled a one-story house.

Comparing the results in the silent condition to those in the corresponding think aloud condition, i.e. *verbal and visual goal* condition, no difference in assembly success could be observed (see Table 5.4). This finding was confirmed by a non significant likelihood ratio; $L\chi^2$ (2, N = 32) = 2.14, p = .34. All instructors succeeded in instructing their partner to assemble the original dollhouse.

condition	no success	house	original dollhouse	N
think aloud (Verbal and visual goal)	1	4	12	17
silent (1^{st} explorative assembly)	2	1	12	15

Table 5.4: Assembled objects rated by success distribution within the think aloud and the silent assembly condition (raw frequency).

5.4.3 Spontaneous verbalization and experimenter's engagement

5.4.3.1 Spontaneous instances of verbalization

Spontaneous instances of verbalization were rarely observed, i.e. six participants (54.55%) did not say anything at all during the assembly, three participants (27.27%) used hesitation markers, such as 'uhm' and 'tja', and two participants (18.18%) uttered the discourse particles 'also' (so) and 'so' once.

5.4.3.2 Experimenter's engagement

In the majority of cases, the experimenter did not need to remind the participant to think aloud (32 cases, 64%) (see Figure 5.13 and Table 5.7 in the appendix). If participants needed to be reminded, it was once in twelve cases (24%), twice in three cases (6%), three times, four times, and five times once each (2% respectively).

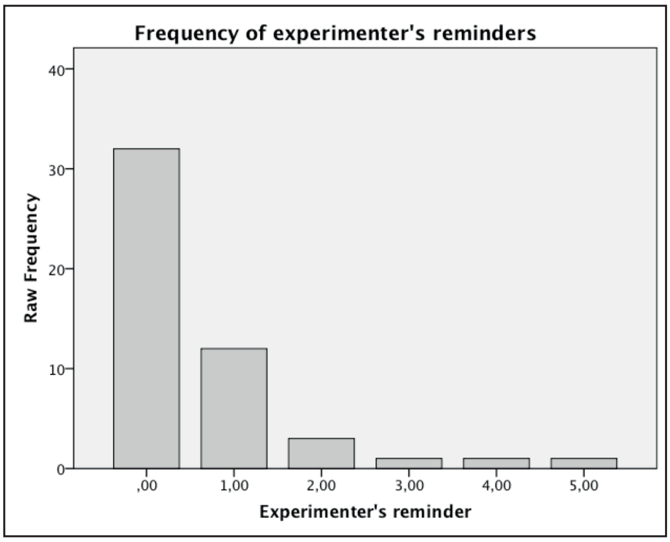


Figure 5.13: Number of experimenter's reminder (raw frequency).

Differences could be observed between conditions (see Figure 5.14). Participants in the *underspecified goal* condition needed to be reminded more frequently (M = 1.06 reminders, SD = 1.60) than participants in the *verbal and visual goal* condition (M = 0.29 reminders, SD = 0.47). Participants in the *verbal goal* condition were in between (M = 0.44, SD = 0.63). A more detailed look revealed that 58.82% of the participants in the *underspecified goal* condition did not need to be reminded (ten cases). However, participants who needed to be reminded three, four, and five times were also to be found in this condition (for more detail see Table 5.7 in the appendix). The greater range of experimenter's reminders in the *underspecified goal* condition shows well in Figure 5.14. Participants in the *verbal goal* condition needed to be reminded once or twice in 37.5% of the cases (six participants) whereas participants in the *verbal and visual goal* condition needed to be reminded least often, i.e. once in 29.41% of the cases (five participants).

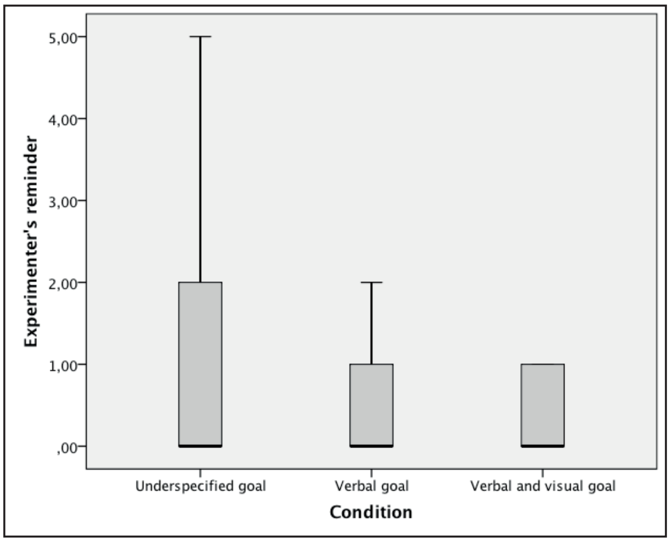


Figure 5.14: Raw frequency of experimenter's reminder by condition

In order to run a Chi-square test, the number of reminders was divided into three categories, i.e. *none*, *once*, *twice or more* (see Table 5.5). The observed differences between conditions reached statistical significance, $L\chi^2$ (4, N = 50) = 9.27, p = .05. Participants in the *underspecified goal* condition were reminded significantly more frequently than participants in the other conditions.

condition	none	once	twice or more
underspecified goal	10	2	5
verbal goal	10	5	1
verbal and visual goal	12	5	0

Table 5.5: Frequency of experimenter's reminders categorized in three groups.

5.5 Discussion

As pointed out in the general introduction (chapter 1), the aim of this thesis is two-fold. The first interest lies in the analysis of the structure and linguistic representation of the problem solving activity verbalized during unaided object assembly. The second interest lies in studying the influence of prior information on assembly performance and the development of the mental representation of the goal structure.

This chapter does not provide any insights on the processes involved in assembling the two-story dollhouse without help but it offers insights on the influence of prior information and communicative intention on assembly performance. First, the effect of those two independent variables on performance measures, i.e. assembly time and individual assembly success, will be discussed. By comparing the mean assembly time for assembly while thinking aloud to the mean assembly time in the silent condition significant differences were revealed. Participants in the *verbal and visual goal* condition needed 6.22 minutes which was almost 3 minutes faster than participants in the silent conditions. This result was unexpected as it suggests that verbalization facilitates assembly in the given task – a finding that has not been reported in the literature so far. It is important to note that prior studies investigating the effect of verbalization on task performance focused on mental problems, such as math or analogical reasoning, as opposed to problems involving physical manipulation such as the one investigated here. Thus, the facilitation effect might be specific to the task. This should be tested in further studies.

Surprisingly, no significant effect of verbalization could be observed on the dependent variable of assembly success. This finding suggests that verbalization facilitates understanding of structural relations between objects regarding their position within the goal structure. However, the same conclusions can be drawn without saying words out loud. Thus it seems that verbalization does not foster different conceptualizations and insights.

The second independent variable that was investigated was prior information. Prior information showed the expected effect on performance time. Participants who were provided with least specific prior information (*underspecified goal* condition) needed significantly more time than participants who were provided with more specific information. Participants who were shown the picture were fastest in their assembly. This trend was mirrored in the average number of words in the think aloud protocols. One possible explanation for these results can be found in Dixon's (1987b) observation that comprehension was fostered by a pre-existing situational representation. As pointed out in the introductory section, performance time does not allow for any insights on the nature of the facilitation effect. However, supported by Dixon's (1987b) findings it can be assumed that the specificity of the representation causes the observable effect. The findings suggest that if the structural representation is very detailed, i.e. the picture of the assembled dollhouse, newly encountered information can be integrated faster than when no such structure is available. In this case, the pre-existing specific information seems to foster the conceptualization of objects, thus revealing the structural relations and functions of individual parts faster.

Prior information also showed an effect on individual assembly success. The results were surprising because they revealed that participants who were shown the picture were least successful comparing the success rates within each condition, i.e. *condition specific success*. Participants who were provided with general domain information were most successful in achieving the goal state as stated in the instruction. These findings are unexpected because, theoretically, participants in the *verbal and visual goal* condition could have used two strategies to be successful. They may have matched the encountered objects to the remembered picture or they may have used the physical and logical constraints as outlined in section 3. The findings suggest that participants who did not remember the picture did not notice the different physical clues and vice versa, thus they had little chance to be successful. The observed results suggest that participants in the *verbal and visual goal* condition were too restricted by the very detailed goal state that they could not apply the second strategy.

If the amount of original dollhouse that was built in each condition is compared, all participants have the same chances to succeed. Participants who saw the ready assembled dollhouse were significantly more successful than all other participants. They succeeded more often in assembling the original dollhouse than participants who did not see the picture. This finding suggests that an external representation provides better and more reliable clues about objects than the objects themselves. It further suggests that the physical features of the objects were not as salient and/or informative to participants as expected. If they were salient, more participants in the two conditions without an external model would have built the original dollhouse as well. This observation rejects the assumption that has been discussed in the previous paragraph because it highlights that the second strategy has not only been rarely used by participants in the *verbal and visual goal* condition but it has been seldom used by all participants. A closer investigation of the nature and function of these task specific constraints will follow in chapter 9.

The frequency counts of the experimenter's reminders revealed that participants needed to be reminded up to five times during their assembly. Those participants that needed to be reminded three, four, or five times were all tested in the *underspecified goal* condition. Participants who were given most information, i.e. *verbal and visual goal* condition, needed to be reminded only once if at all. Overall only 46% of the participants needed to be reminded and most of them only once. The finding that participants in the *underspecified goal* condition needed to be reminded significantly more often than participants in the other conditions was unexpected from the literature point of view because so far it has only been assumed that differences are caused by participants' individual verbosity (e.g. Chi 1997; van Someren et al. 1994). The reported results, however, support the expectation that participants who needed to create the assembly structure from scratch faced a more demanding task and thus forgot to think aloud more frequently than participants who were provided with a specific structural representation.

Contrary to the expectations outlined prior to the analysis, no instances of spontaneous self-talk were observed in the silent assembly condition. However, the analysis revealed few cases of occasional verbalizations such as hesitations and discourse particles, namely 'so' and 'also' (so). These can be interpreted as externalizations of thought processes (e.g. Caron-Pargue & Caron 1991; Tenbrink 2008; Bégoin-Augereau & Caron-Pargue 2010). However, as they are not rich in content by themselves and no more context information is verbalized, this finding does not contribute to any further analysis. This result seemed surprising at first glance given different personal experiences. However, going back to Goffman's (1981:81) observation that self-talk is socially taboo, participants would need to be very engaged in their task to speak freely. Furthermore, the experimental situation strengthens the social taboo because as Schoenfeld (1985) pointed out participants feel that it is inadequate to behave in the laboratory the same way as at home in private. Thus, the observed lack of self-talk may be an artifact caused by the experimental situation.

5.6 Appendix

condition	no success	house	original dollhouse	N
underspecified goal	1.4	0.7	-1.3	17
verbal goal	-1.1	0.7	-0.3	15
verbal and visual goal	-0.3	-1.4	1.6	17

Table 5.6: Assembled objects rated by success distribution within conditions (standardized residuals).

number of reminders	underspecified goal condition	verbal goal condition	verbal and visual goal condition
none	10	10	12
one	2	5	5
two	2	1	0
tree	1	0	0
four	1	0	0
five	1	0	0
sum	17	16	17

Table 5.7: Raw frequency of experimenter's reminder in different conditions.

6 Features of analysis: Problem solving processes in unaided object assembly

6.1 State of the Art

On the most general level, human problem solving is defined as a search through a problem space. The problem space consists of the present state, the goal state if it is known, and all possible intermediate cognitive states that were generated in previous problem solving activities or which are generated during the current problem solving activity. In order to modify existing states and generate new states, operators are used that are generally known to the problem solver or which are newly learned. This view summarizes the main assumptions proposed by Newell and Simon (1972) in their seminal work on human problem solving; a more detailed summary is provided in chapter 2. In experimental settings, the participant needs to construct a problem space by

> "defining goals, rules, and other aspects of the situation – in some kind of space that represents the initial situation presented to him, the desired goal situation, various intermediate states, imagined or experiences, as well as any concepts he uses to describe these situations to himself." (Newell & Simon 1972:59)

The primary aim in this thesis is the identification and description of the exact problem solving processes, i.e. the construction of the problem representation and the activities there in, involved in object assembly. Most research in the area of problem solving has focused on mental problems in which the present state needs to be transformed into a well-defined goal state, e.g. in the Tower of Hanoi (for more information see chapter 2). However, Newell and Simon's (1972) theory suggests that problem solving activities are to be found in other tasks as well, i.e. tasks that do not suggest problem solving at first glance. Palmer (1974), for example, identified the characteristics of problem solving when participants solved a mental synthesis task. In this task, participants are asked to identify which two structural elements need to be combined to form the presented figure. In the participants' activities, Palmer (1977:466) observed explorative hypotheses, false leads, dead ends, backtracking, and fresh starts; processes that cover with the term problem solving.

As will be outlined in detail in the following sections, first time assembly is another domain in which problem solving activities are likely to be observed. Those may be first time assemblies of what Richardson (2007) calls "self-assembly" of furniture as well as toys. In many cases, manuals facilitate the assembly. A survey conducted in the United Kingdom by the Office for National Statistics[110] showed that

> "52% of the adults in the UK have assembled self-assembly furniture in the last 2 years. Of these, it was estimated that 3 million UK adults have damaged the item being assembled and 1.8 million received minor injury." (Richardson 2007:305)

These figures illustrate that self-assembly is part of many people's everyday live and that success is not always guaranteed. The observation that the goal object was damaged suggests that self-assembly might result in problem solving activities possibly because of misleading, erroneous, or missing manuals. In order to test this hypothesis, the structure of the problem solving processes reported during unaided self-assembly will be studied and described in detail in this chapter. After providing an overview on research about mental processes involved in assembly

[110] The results are based on 1295 face-to-face interviews in private households in Great Britain. The aim of the survey was "to identify any consumer issues that there may be with self-assembly furniture" (Richardson 2007:307). For more detail about the Omnibus Survey see Richardson (2007).

tasks (unaided as well as aided by manuals), research questions and expectations will be outlined. Then a detailed account of the procedure of analysis is provided. In the presentation of the results the focus will be on the overall structure of the assembly process in general and the description of the problem solving processes specifically. The chapter will conclude with a discussion of the results against the backdrop of a *general structure of mental processes involved in assembly* that is proposed based on the literature on mental and physical processes involved in assembly.

6.1.1 Introduction

Every task consists of a certain pattern of actions taken with the aim to change the current state of the environment in order to achieve an intended state, such as brewing coffee in the morning as an everyday task. Zacks et al. (2001) studied participants' perception of tasks involved in daily routines (making the bed and washing dishes) or less frequently performed ones (assembling a saxophone and fertilizing houseplants). In order to record participants' perception of these tasks, they asked them to segment continuous videos of these actions into the smallest or largest meaningful units. Their analysis showed that these activities are perceived of as hierarchically ordered sequences of actions on objects. Furthermore, they found that descriptions about these individual sequences, also termed events[111], "referred to accomplishments or achievements, activities that culminate in natural endings" (Zacks et al. 2001:41). Therefore, they conclude that "perception of unfolding events entails thinking about function, causes, goals, and ends" (Zacks et al. 2001:41). This was the same for familiar tasks as well as for less frequently performed ones.

Prior to Zacks et al.'s study, Norman (2002)[112] addressed the process of action planning and taking from another perspective. He evaluated the design and usability of everyday objects in order to study people's routine performance with them. He specifically looked at technical devices that are used on a daily basis such as computers, VCRs, or film projectors. His focus was on their design and ways to improve it in order to make the usage of everyday devices easier and more intuitive. In his popular book "The design of everyday things" Norman proposed a model of action planning and taking consisting of seven steps (Norman 2002:48), namely

[111] "An event is defined to be a segment of time at a given location that is perceived by an observer to have a beginning and an end" (Zacks, Tversky, & Iyer 2001:29).

[112] Norman (2002) is a reprint of his book first published in 1988.

a) forming a goal,
b) forming an intention,
c) specifying the action,
d) executing the action,
e) perceiving the state of the world,
f) interpreting this state with regard to the intended goal,
g) evaluating the outcome.

Norman (2002) stresses that these steps do not resemble discrete entities but may overlap and are repeated several times until the final goal is achieved. Thus, he calls it the 'action cycle'. The final goal may be an internal representation of an action (e.g. brewing a coffee) or an object (e.g. a chair). In order to achieve the intended goal, the current state of the world needs to be modified by taking specific actions. These actions have to be ordered in an action plan, which is then carried out successively or simultaneously. Planning is a mental process whereas the execution involves interaction between the subject and the physical object or environment. If the intended result is not achieved by the actions taken, the action plan needs to be modified and carried out again. Norman (2002) proposes that this revision and action taking is repeated until the goal representation matches the state of the external world. Although Norman (2002) took another approach and long before Zacks et al. (2001), his model suggests that function, causes, goals, and ends are not only important information in the perception of events but they are also driving forces in planning and performing events. If people are confronted with tasks for which they do not have a learned action plan ready, the formulation of goals becomes central and functions need to be conceptualized and defined along with the objects that are needed. Thereby the creation of an action plan turns into a problem solving activity.

A thorough and detailed review of the literature on traditional as well as recent theories on problem solving, mainly focusing on logic-based problems, is provided in the chapter 2. As the aim of this thesis is the identification and description of problem solving activities in an everyday task, the following introduction will focus specifically on theories and empirical findings in this area. Returning to Palmer's (1977:466) characterization of problem solving involving *exploratory hypotheses*, *false leads*, *dead ends*, *backtracking*, and *fresh starts*, it is suggested that Norman's model can be used as a reference for classifying the steps taken in solving problems posed by a construction task. The following comparisons are proposed: the *goal* might be an explorative hypothesis as well as a mental representation of an object or action. The *actions* taken can be false leads resulting in dead ends, which can be overcome by reorganizing the actions taken or creating a different plan of actions (fresh start).

This structure can be further elaborated by Ganier et al.'s (2000, cited in Ganier 2004:16-18) findings on cognitive activities involved in processing procedural instructions. They identified the following five basic cognitive activities:

1) setting a goal representation and holding it in working memory,
2) integrating information provided by the document, the equipment, and the user's prior information,
3) planning and executing actions,
4) monitoring and regulating activity, and
5) integrating novel information into long-term memory.

Even though their research focused on aided assembly, their model is very similar to the general steps proposed by Norman (2002). Furthermore, it highlights two important information resources, namely the activation and integration of prior knowledge as an aid, and the integration of new information in long-term memory. Those two components are interrelated in the sense that new information is always integrated into already existing informational networks. This is specifically important in construction tasks because "in assembly tasks, the object changes shape as it is constructed through execution of a specific sequence of steps" (Novick & Morse 2000:1243). In assembly tasks, prior knowledge about similar assembly tasks or general construction features as well as first associations serve as references and basis for creating a mental representation of an assembly. In aided construction tasks, manuals provide mental representations. If the reported findings are combined, the following general structure of problem solving in assembly emerges:

a) a goal is formulated, the problem solver has the intention to achieve the goal, and the goal is held active in working memory until the task is fulfilled,
b) the intention is reformulated into an action sequence – this action sequence is derived from the mental model that is created as a situational representation by integrating information presented in the manual, the equipment at hand, and prior knowledge,
c) the action is executed,
d) the changed state is perceived and interpreted with regard to the goal state,
e) the outcome of the interpretation process is evaluated,
f) newly acquired information is integrated into long-term memory.

Those processes may arise in a sequential order but they may also interact and new structures emerge when components are skipped or repeated consecutively. This general structure will be further elaborated by integrating findings obtained in experiments on object assembly with the aim of presenting a general model that is used as a reference point in the analysis and which can be supported or falsified by the results obtained.

6.1.2 Object assembly

6.1.2.1 Research on object usage and assembly

Helander and Willén (2003) were among the first researchers to look at assembly tasks from the assemblers' point of view by considering their mental work during assembly. In their paper, they identified three human factor principles involved in assembly that are based on the human information processing system, namely perception, decision making, and manipulation (Helander & Willén 2003:26-27). In product design, it is assumed that manipulation consists of the following elemental tasks: reach, grasp, move, position part, and insert. Positioning the part involves the tasks of alignment, orientation, and engaging the object into another object or a structure of objects (Helander & Willén 2003:26). This brief technical description of actions involved in an assembly task illustrates the great complexity inherent in such tasks even if the structure is known. This complexity stresses the importance of an accurate mental representation of the goal structure and the position of the individual parts (Helander & Willén 2003:26-29). The creation of an appropriate mental representation can be facilitated by well-designed assembly instructions, i.e. manuals.

Thus, the design of manuals has become a widely researched field in recent years. The balanced distribution of pictures and diagrams as visual aid (e.g. Glenberg & Langston 1992; Marcus et al. 1996; Norvick & Morse 2000; Heiser et al. 2004) as well as the design of diagrams (Heiser et al. 2003; Tversky et al. 2007) are among the most widely investigated topics. Based on previous findings by Tversky and colleagues (e.g. Zacks et al. 2001), Heiser et al. (2003) assume that people conceive assembly tasks as hierarchically organized sequences of actions in which "objects or significant object parts separate steps at the higher level and differentiated actions on the same object at the finer level" (Heiser et al. 2003:546). This assumption is supported by their subsequent studies on user's instructions on assembly tasks. Therefore, Tversky et al. propose that "the mental model of assembly is a hierarchical set of actions and subactions on objects or object parts" (Tversky et al. 2009:126). This observation is an empirical affirmation of Klix's (1976) proposal on the nature of cognitive processes. In his work, he states that action sequences consist of sub-actions that vary in duration and complexity depending on the overall complexity of the action sequence. Klix (1976:28) also postulates that although sub-actions may have a different nature, they are arranged in a hierarchical structure.

Additionally, Prabhu et al. (1995) identified hierarchical structures not at the task level but at the level of object structure. They distinguish between hierarchical and vertical product structures:

> "Vertical product structures may be build sequentially, with the placement of a new piece on the most recently placed piece. Hierarchical products require the construction of several subassemblies. The final product is assembled from the subassemblies." (Prabhu et al. 1995:150)

Prabhu et al. (1995) propose that the perceived structure guides the assembler's strategy in that "visual perception provides assembly heuristics" (Prabhu et al. 1995:150). Nevertheless, the aid provided by perceptual information is limited because the perceived heuristics do not need to be correct and may result in unsuccessful assembly. Although Prabhu et al. (1995) do not specify the strategies any further, the following descriptions for assembly strategies emerge. In vertical assembly the construction of the structure is perceived as a single unit that can be built in a bottom-up manner. In hierarchical assembly several subassemblies need to be performed before those are merged to build the final product. The fundamental difference according to Prabhu et al. (1995) is to be found in the information processing involved in perceiving these structures. In remembering hierarchical products, all subassemblies need to be maintained in working memory. This demand on memory capacity is important in the assembly process because assembly that is not aided by a manual involves "semiskilled work requiring manual ability, mechanical ability, and visual memory capacity" (Prabhu et al. 1995:151).

Further interests can be identified in the overall structure of manuals (e.g. Ganier 2004; Richardson et al. 2004) or the influence of users' aims on the structure and lexical choice in instructions (Maes 1999; Maes et al. 2004). Moreover the evaluation of assembly complexity has started to influence manual design recently because "it is the structure of the object to be assembled that ultimately impacts on assembly complexity" (Richardson 2011:190). To conclude, it can be noted that scanning the literature in search of reports about mental processes involved in assembly tasks, a substantial body of research on the design of manuals was revealed. However, surprisingly those studies only rarely mention or consider the assembler's mental processes involved in the task.

6.1.2.2 Experimental studies on object assembly

In this section, literature that addresses actual assembly performance will be reviewed. Although there is much research on manual design there is hardly any research on the actual assembly performed by people. Therefore, mental processes involved in assembly performance have not been studied systematically so far.

First attempts to bridge this gap are reported by Richardson and colleagues. In their paper, they present theoretical observations on object assembly before testing which task variables influence theoretical as well as actual object assembly. Based on the literature, Richardson et al. (2004) identified two phases in construction tasks. First, instructions need to be understood and a mental representation of the assembly is built. Second, the mental representation is used to facilitate the actual assembly process. Richardson et al. (2004) analyzed 40 instructions that came with self-assembly products[113]. In their analysis they identified four fundamental assembly operations, i.e. selection of components required for the next assembly procedure, rotation of the object to allow correct positioning, positioning to allow fastening, and fastening (Richardson et al. 2004:949). Those are also identified by Helander and Willén (2003), which support the contention that those components are central to assembly performance.

In a second step, Richardson and Jones (2005) ran an empirical study on manual aided assembly of eight different pieces of ready-to-assemble furniture (Richardson & Jones 2005:181). They were interested in different task determinants that influence thinking time (i.e. time spent on studying the manual and performing the first three fundamental assembly operations). Their results revealed that qualitative information, such as variety and asymmetry of components, had a significant effect on "assembly complexity and impacts on cognitive load" (Richardson & Jones 2005:183). Their study highlighted three additional indicators of assembly complexity, i.e. fastenings, fastening points, and selection. Richardson and Jones (2005) argue that the selection of object parts influences cognitive load because a great amount of different objects indicates numerous novel assembly steps. Each new assembly step requires the assembler to conceptualize the object and create a mental representation of the sub-assembly. If the same assembly steps need to be repeated, cognitive load decreases.

Besides observing people solve assembly tasks in experimental settings, Richardson was interested in people's actual experience with self-assembly furniture including

[113] Richardson (2007:307) explicitly defines self-assembly furniture to be "furniture that are sold 'ready to assemble'. [More specifically,] the purchaser must follow supplied instructions and assemble the item before it can be used".

problems as well as satisfying moments. Since there was no statistical data to answer this question, he conducted a survey in private households in Great Britain (Richardson 2007). From the numerous very interesting results only a selection will be presented here. The data showed that 52% of the British people (aged 16 and older) assembled some kind of furniture in the last two years. About half (i.e. 49.3%)[114] of those reported that they experienced the completion of the assembly to be satisfying. Mistakes (38.6%) and pleasure (34.8%) were common phenomena during the assembly (Richardson 2007:309). This part of Richardson's (2007) survey highlighted that self-assembly involves numerous obstacles as well as joyful moments; results that suggest that people are highly engaged in the assembly process. Additionally, Richardson's study also provided evidence that object assembly may easily result in problem solving activities. His survey revealed that 13.2% of the interviewees damaged the object they assembled and almost half of all interviewees (41.0%) admitted that they got angry with the assembly (Richardson 2007:309). Interestingly, 5.9% of the interviewees reported that they found the assembly too difficult to complete (Richardson 2007:309).

Tversky et al. (2009) were interested in the overall structure of assembly instructions given by participants after they had assembled the object for themselves. The object in question was a TV cart consisting of five parts (Tversky et al. 2009:122). Their analysis revealed a general narrative structure consisting of a beginning, middle, and end. In the middle part, the assembly procedure was described resembling a hierarchical structure of actions and subtasks (Tversky et al. 2009:125). This finding was the same for all three communicative modes that they investigated, namely gestures, words, and diagrams. All instructions followed the same narrative structure consisting of a beginning, middle, and end. In the beginning, some form of introduction to the task or to the object parts was provided. The middle parts consisted of step-by-step descriptions of procedures. The transitions between sub-tasks were either marked by words (e.g. first, next, after that, now) or gestures. In the end, all subjects indicated that the task was accomplished either by using deictic gestures or words to point at the assembled object or by circling the final diagram. Tversky et al. (2009) conclude that *beginnings*, *middles*, and *accomplishments* (also called events) are "perceived and conceived to be a hierarchical set of actions on objects accomplishing goals and subgoals" (Tversky et al. 2009:126).

In a recent publication, Daniel and Tversky (2012) were interested in the relevance of the different *events*. In order to test which information was considered crucial

[114] Interviewees were allowed more than one answer.

for the assembly, Daniel and Tversky compared instructions produced without time constraints to instructions produced under the constraint of brevity. In a qualitative content based analysis Daniel and Tversky (2012:305) identified minimal units of information. They classified those units according to six categories, namely *general comments*, *temporal comments*, *action*, *extrinsic action*, *intrinsic action*, and *description*. Since the authors were specifically interested in spatial information, the three categories containing information on action were distinguished based on their spatial content. Some information units contained action verbs without information on spatial orientation of objects. According to Daniel and Tversky (2012:307), this information about 'what-to-do' represented higher-level action statements. Extrinsic actions contained action verbs and information on the spatial reference system of the body or the world. Intrinsic actions, in contrast, contained information in which the object was used as a spatial reference system. Both action categories that contained information on the spatial reference system were considered to encode sub-goal information (Daniel & Tversky 2012:307). Any information about physical properties of the object was classified as descriptions.

In addition to this new classification of 'minimal units of information', Daniel and Tversky (2012) identified the same narrative structure of beginning, middle, and end as reported in Tversky et al. (2009). Instructors choose different ways of starting an instruction. They either started the assembly right away or they provided a list of the objects and tools that are needed for the assembly or they provided a general introduction to the task by explaining the global shape. Daniel and Tversky (2012:306) report an effect of the constraint of brevity on the starting section of the instructions. If participants were unconstrained in their time, they used all three strategies to the same degree. However, participants who were told to be brief started the assembly right away in the majority of cases. They also report an influence of brevity on the information level (higher-level vs. sub-goal action information). If participants were told to be brief, they used significantly fewer units that included sub-goal action whereas the amount of higher-level action information remained the same (Daniel & Tversky 2012:307).

Whereas Daniel and Tversky (2012) studied the structure of assembly instructions that were created for an imagined generic addressee, Rieser (1996) collected assembly instructions in an interactive experimental setting. More specifically, he was interested in the use of functional terminology for a physically present object in assembly dialogues. He was interested in the way in which instructors created and assigned meaning to objects. The data suggested that meaning was created by an ad-hoc concept of the object at hand. In the instructions, participants referred to features that are typically assigned to a well-known structure for the description and identification of physically present objects. Rieser (1996) called this concept *representational metonymy*. The analysis revealed different representational communicative functions of what Rieser called *representational metonymy*, such as identification of object parts and their correct orientation (Rieser 1996:13)[115].

Analyzing the general structure of the assembly dialogues, Rieser (1996) observed that instructors used two strategies for starting the dialogue. Instructors either used the 'global strategy' in which they mentioned the goal structure, i.e. airplane. Whereas other instructors chose not to mention the goal structure in the beginning but rather assembled the object in different sub-assemblies ('local strategy'). These sub-assemblies followed the structure 'identify parts, use parts, cover all holes, pay attention to the symmetry of the emerging structure' (Rieser 1996:11). Rieser (1996) reported an influence of introductory strategy on the frequency of *representational metonymy* in the instruction. The results revealed that instructors who chose the 'local structure' used *representational metonymy* less frequently than instructors who chose the 'global strategy'.

So far, literature from two domains has been reviewed, namely problem solving and object assembly. In a next step the relations between the reported findings are highlighted to propose a theory based structural representation of problem solving process in assembly tasks. This theoretical structure will be used as a reference point throughout the analysis in this chapter. First, the reported findings are summarized in Table 6.1.

problem solving process	reference in the literature	processes identified in assembly	reference in the literature
selection of goal	Newell and Simon (1972)	information on the task and/ or the goal structure	Daniel & Tversky (2012); Rieser (1996)
selection of method	Newell and Simon (1972)	activation and integration of prior knowledge, information provided by a manual, and information available in the environment	Ganier (2004)
execution of action	Miller et al. (1970), Newell & Simon (1972)	Execution of extrinsic, intrinsic, and higher level actions	Daniel & Tversky (2012)

[115] For a description of the concept of *representational metonymy* see chapter 8.

		4 basic assembly operations: select, orientate, position, and fasten components	Richardson et al. (2004)
		manipulation: reach, grasp, move, position part, insert part	Helander & Willén (2003)
evaluation of results	Miller et al. (1970), Newell & Simon (1972)	closing an instruction, e.g. by exemplifying the usability of the assembled structure	Daniel & Tversky (2012)
		integration of novel information into long-term memory	Ganier (2004)
exploratory hypotheses, false lead, dead end, backtracking, fresh start	Palmer (1977)		
		perception, decision making, and manipulation	Helander & Willén (2003)

Table 6.1: Summary of problem solving processes and assembly processes described in the literature.

Second, these findings are combined and integrated into Norman's (2002) action cycle to propose the following *general structure of mental processes in assembly*. The stages are at a general level comprising individual problem solving processes. In the final discussion, the adequacy of the proposed structure will be evaluated.

a) *Goal representation*: Goals can be formulated with regard to different reference points, i.e. either with regard to the overall task where the overall structure is already given, e.g. "Use all given objects and assemble a cupboard." Goals can also be formulated with regard to sub-assemblies, e.g. "Attach the doors onto the cupboard." The first example defines the goal on a *conceptual level* (goal concept cupboard) whereas the second example defines the goal at the *action level* (fasten doors on cupboard). Thus, two levels of goal information are identified: reference point and nature.

b) *Intention to achieve the goal*: Goals are either motivated intrinsically, i.e. the motivation to act resides in the problem solver, or extrinsically, i.e. a goal is presented by a third party.

c) *Reformulation of the intention into an action sequence*: (Explorative) hypotheses are formulated about the necessary actions. These hypotheses are based on the integration of prior knowledge, information provided in the manual and associations activated by the equipment at hand. In order to perform the actions, the objects that are needed have to be identified. The selection might be based on the mental representation of the object, by some external object representation such as a picture or a manual, or by general knowledge about assembly, e.g. screws are needed if there are boreholes. Temporal comments may indicate the hierarchical structure of the intended actions.

d) *Execution of pre-planned actions*: Planned actions are selected to complete the four basic operations in assembly procedures, i.e. select, orient, position, and fasten components. The execution may result in a sequence of sub-assemblies (vertical assembly) or as a bottom-up assembly (hierarchical assembly). If an operation does not work as planned, fresh starts (e.g. removing the object that has just been placed) might be necessary.

e) *Recognition and interpretation of changed state* with regard to the goal state: The interpretation might be positive leading to new plans of action or it might be negative (false lead).

f) *Evaluation of outcome* of the interpretation process: if the interpretation is positive and the final goal is not yet achieved, this outcome leads to the formulation of new plans of action. If the interpretation is negative, there are two possible outcomes, i.e. either the assembler has an idea as how to proceed and a revised action plan is formulated or the assembler does not know how to proceed then a dead end state is reached. Ideally, this dead end can be overcome by a fresh start.

g) *Integration of newly acquired information into long-term memory* except for procedures that are performed only once, such as furniture assembly (Ganier 2004:17).

In the following section, the questions guiding the investigation of verbalized mental processes will be stated and expectations will be outlined. All expectations build on the proposed general structure of mental processes in object assembly.

6.2 Research Questions and expectations

Based on the reviewed literature on assembly task performance and the literature on traditional and recent theories on problem solving, three basic research questions arise. Each of them is discussed in more detail regarding expectations in the following paragraphs.

- <u>Research question 1</u>: Which assembly strategy is preferred by assemblers working by themselves and which one is preferred when instructing someone else?

- <u>Research question 2:</u> Which general structure can be identified in think aloud protocols? How do assemblers start their task and how do they indicate the end? Is there a difference regarding starting and concluding phrases between participants who think aloud and instructors?

- Research question 3: Do think aloud protocols of unaided object assembly contain verbalizations of problem solving processes? If they do which ones are expressed? Are there other processes that are not described in the literature so far?

- Research question 4: Is there a specific recurring pattern of problem solving processes?

Research question 1 addresses the preference for assembly strategies. Prabhu et al. (1995) differentiate between hierarchical and vertical assembly proposing that the first strategy involves different sub-assemblies. If the selection of the strategy is based on perception, hierarchical assembly involves a higher cognitive demand because the sub-assemblies need to be remembered during the assembly. In the investigated experiment, the goal object is a two-story dollhouse in which the first story is identical to the second story. If participants recognize this parallel structure, they may choose a hierarchical assembly approach. Based on the assumption that is proposed in the literature stating that action plans are hierarchically structured, the vertical assembly strategy, i.e. bottom-up, is also likely to be frequently observed.

It can be assumed that prior information has an effect on the selection of the assembly strategy. Participants who saw the picture of the dollhouse might observe the parallel structure and thus, identify a hierarchical product structure, as described by Prabhu et al. (1995). Therefore, they are expected to assemble the dollhouse in a hierarchical fashion more frequently than participants who did not see the picture. It is further assumed that communicative intention will have an effect on selected assembly strategy. Based on the proposal that thinking about actions is hierarchically structured (Zacks & Tversky 2001), instructors are assumed to use the vertical assembly strategy more often than the hierarchical one to help their partners.

Research question 2 contains a number of sub-questions. The primary interest is in the general structure of problem solving in think aloud protocols. It is expected that participants will explicitly mark their starting point and indicate when they are finished with their task. Furthermore, preliminary analyses (Tenbrink & Gralla 2009) showed that the main activity of problem solving happened in the middle part of the protocols, i.e. between the introductory statement and the concluding phrases. Thus, the broader structure of beginning, middle, and end was called *global structure* and the specific structure of the assembly process verbalized in the middle was referred to as *local structure*. Based on this observation, the way in which assemblers start their task and how they indicate the end will be analyzed. Generally, it is expected that the formulation of the goal and the motivation to achieve it will be verbalized at the beginning of the assembly. These processes correspond to the steps a) and b) in the proposed general structure of mental processes in assembly. Given the experimental set-up, participants are expected to recite the task instructions when entering the room and starting the assembly because participants were presented with the task instructions in front of the conference room[116]. Since the instructions were presented orally and only once, it is assumed that participants recite the task instructions for themselves as a reminder of what to do. Additionally, it is expected that participants will refer to the task instructions in the end as well because they are explicitly told to tell the experimenter when they are done. Based on Richardson's (2007) findings, participants may also express enjoyment about the task in the closing sequence. The closing sequence of the task marks the final step of the assembly task. Therefore, the closing remark is likely to represent the state in which the current product is evaluated against the intended goal structure, i.e. step f) in the general structure of mental processes in assembly (see page 165). This evaluation may be verbalized in comments on the final product or the assembler's own performance.

Besides studying the initial and final comments of an assembler, the influence of communicative intention on the *global structure* was investigated. Based on the literature, instructors were expected to choose one of the following three options to start their instructions; provide a list of parts, offer a structural model of how the parts fit together, or give general advice on the task (Daniel & Tversky 2012:126). Similarly to Daniel and Tversky's (2012) second option, Rieser (1996:11) describes one strategy, referred to as 'global structure', in which the goal structure was introduced and used as a reference point in the following assembly[117]. Additionally, Rieser (1996) reports the so-called 'local structure' in which instructors did not provide any information about the goal structure but worked by sub-assemblies without an explicit introduction of the assembly goal. Concerning the endings of instructions, Daniel and Tversky (2012:126) observed that instructors either simply stated that the task was accomplished or they made suggestions about how the assembled object could be used. Other instructors closed by encouraging the assembler to be proud of their achievements. Similar findings were expected in the current study.

[116] For more detail on the procedure see chapter 5.

[117] This reference is linguistically expressed by use of functional terminology related to the conceptualized goal structure (e.g. airplane in Rieser's (1996) experimental design).

Research question 3 follows from the distinction between *global* and *local structure*. As the main focus of this thesis is on the representation of problem solving processes in unaided object assembly, this research question aims at identifying those processes that are specified in the middle section of the think aloud protocols. Besides identifying those problem solving processes that are described in the literature on problem solving, the interest is in describing and identifying processes that are expressed in the protocols but have not been identified and described in the literature so far. Think aloud protocols are expected to contain problem solving processes for two reasons. First, as outlined in detail in the theoretical section of this chapter, Richardson's (2007) omnibus survey suggests that the assembly of furniture can be very demanding and result in trial and error behavior. Second, assembly is often based on the guidance by assembly instructions but the experimental set-up explicitly poses the task of assembly without external help. Thus, participants need to construct a representation of the goal structure and integrate all given parts by themselves during the assembly. Furthermore, they need to update their mental representation continuously during the assembly. These steps are representative for a problem solving sequence.

The second part of research question 3 focuses on the description of the individual processes that are identified in the think aloud data. First, this analysis aims at identifying known processes, such as hypotheses, false lead, dead end, backtracking, and fresh start (Palmer 1977). The literature review on traditional and recent theories on problem solving, presented in chapter 2 of this thesis, showed that search and test processes are the driving force of problem solving. Thus, it is expected that representations of those processes can be identified. Regarding the specific task of assembly, Prabhu et al. (1995) characterize it "as semiskilled work requiring manual ability, mechanical ability, and visual memory capacity" (Prabhu et al. 1995:151). The mechanical ability is specified by Richardson et al.'s (2004) four basic assembly operations. Visual memory capacity is addressed in Daniel and Tversky's (2012) observation that instructors provide descriptions of objects. It is expected that object features will be described frequently in think aloud protocols.

Second, the detailed analysis of processes aims at identifying and describing those processes that are not reported in traditional literature on problem solving yet but that highlight mental processes involved in solving the assembly task. These additional processes may either be specific to the discourse genre of think aloud protocols or to the task of assembly which has a different structure than logic centered tasks such as chess. Given the specific genre of think aloud protocols, it is expected that they include information on participants' associations with regard to the task as well as the objects. These associations do not need to be related to the task directly but may refer to general associations, e.g. with a dollhouse. Van Someren et al. (1994:122) refer to these comments as 'talking about not-task related issues'; a classification that does not seem to be reasonable because associations are activated by the task performance which makes them associated to the task. Therefore, these processes are interpreted as processes activated by the problem solving activity in this thesis. Think aloud protocols may also contain meta-information about the task, e.g. about the design, the time it takes etc. as well as comments on oneself (e.g. 'I am not comfortable' in van Someren et al. 1994:122).

Research question 4 builds on the findings of research question 3. It investigates if specific and recurring patterns of problem solving processes can be identified across the majority of protocols. In traditional accounts of problem solving, such as the TOTE unit represented by Miller at al. (1970), it is assumed that problem solving processes are structured in cycles of hypotheses → execution → evaluating → new hypotheses etc. Preliminary analyses of eleven protocols by Tenbrink and Gralla (2009) revealed a comparable structure (see Figure 6.1). The circle between hypothesis → action → evaluation → positive → hypothesis represents the most frequently observed cycle. The other circles were less frequently observed. Instances of *dead ends* and *fresh starts* were only rarely identified.

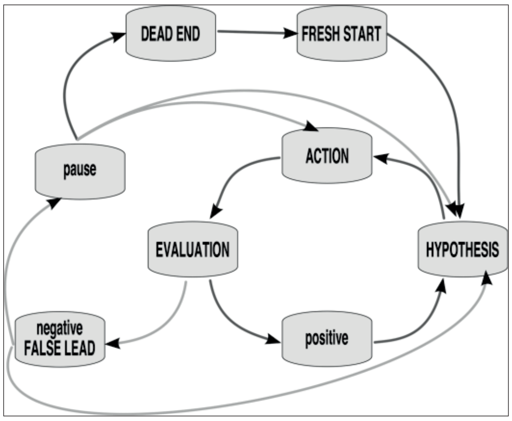


Figure 6.1: Process model derived from eleven think aloud protocols (Tenbrink & Gralla 2009:9).

6.3 Data Analysis

As outlined in the description of the different research questions the analysis focuses on two levels of description, i.e. qualitative and quantitative descriptions.

First, the different processes that are verbalized in the recorded think aloud protocols are identified in a qualitative content-based analysis.

Second, the identified processes are categorized according to categories described in the literature and new categories. These new categories will be described in detail.

Third, patterns of these processes are highlighted and investigated in a quantitative frequency analysis.

After this general overview a more detailed description of the respective procedures is provided.

6.3.1 Qualitative analysis

Based on a content analysis, it was identified and described what was verbalized in the think aloud protocols. The final coding scheme that comprised a number of general processes was evolved in a number of steps that will be outlined in the following paragraphs.

6.3.1.1 Coding scheme for processes

Since this analysis focuses on the content and sequential order of problem solving processes, it was important to be very precise in their description. In a first step the content of each process unit was summarized in few words for five protocols in a first run (see Figure 6.2 for an example). The sequences were annotated according to the temporal markings provided by the participant's silences. During this analysis, a first impression was gained of the topics that were addressed by participants (e.g. object features), of their mental activities (e.g. memory of picture, expected vs. encountered objects), and actions (e.g. sort and place objects). Van Someren et al. (1994:39) also suggest using pilot protocols to gain an impression of the vocabulary and phrasing that is used by participants.

In a second step, the categories that are described in the literature (hypothesis, plan for action, action, evaluation, dead end, false lead, and fresh start) were annotated. If there were cases that could not be classified according to those categories, those were grouped based on their content. From these groupings, new categories emerged based on their function that was linguistically expressed or arose from the immediate context. This second step in the analysis resulted in a more general coding scheme that was iteratively expanded with the protocols to come (see Figure 6.3 on p. 172 for an impression). After approximately 20 protocols the final coding categories were identified[118] and a more abstract coding scheme was derived (see Figure 6.4 on p.175).

	A	B	C	D
1	**Ablauf**	**Zeit**	**Verbalisierung**	***Kommentar***
30	KOMMENTAR ZU DEN MENTALEN PROZESSEN	01:14	im Augenblick schießen mir grade alle möglichen Ideen durch n Kopf	
31				
32		01:18	was ich bauen könnte aber keine nimmt so richtig Gestalt an	
33				
34	KONKRETE HANDLUNGSPLANUNG	01:24	ich glaub ich muss erstmal n paar Teile einfach so ähm zueinander stellen um dann	
35				
36		01:28	irgendwie Vorstellungen von konkreten Formen zu haben die ich bauen könnte	
37				
38	ÜBERLEGUNG DES GRUNDBAUPLANS (GRUNDFLÄCHE)	01:31	will ich die Teile aufeinander stapeln oder	*Referenz auf bekannte Modelle von Häusern (Bungalow vs. Etagenhaus)*
39				
40		01:35	will ich irgendwie alles in einer Ebene flach bauen	
41				
42	ÜBERLEGUNGEN ZUR STABILITÄT	01:39	kann man nicht halt auch [unverständlich] zusammen stecken	*Referenz auf bekannte Modelle der Stabilisierung (stecken oder so bauen, dass das Gewicht gleich verteilt ist)*
43				
44		01:46	oder muss ich versuchen die so aufeinander zu stapeln dass das gerade so sich ausbalancieren lässt	

Figure 6.2: Example of an explorative content analysis of one protocol [10][119] containing information on the event structure ('Ablauf'), temporal sequence ('Zeit'), transcribed verbalization ('Verbalisierung'), and comments ('Kommentar').

[118] The final list of process categories will be described in detail in 6.4.3. Therefore, no more detail is provided here.

[119] The numbers in square brackets always refer to the protocol ID from which the example was extracted.

A	B	C	D
HYPOTHESE/ HANDLUNG/ FALSE LEAD	00:03:05-5	und das muss in die andere Etage / also das / könnte so sein dann müssen da jetzt / [?] Wand rein also das muss auf jeden Fall nach oben// nee	*physical o. logical constraint*
DENKPAUSE (5 sec.)	00:03:09-1	hm	
EVALUATION des Objekts/ HANDLUNG	00:03:26-4	also [?] aber es fehlt die Rückwand / ist da ne Rückwand// das ist ds Dach und das hier ist auch Dach dann sortier ichs gleich beiseite	*Gesamtüberblick wird angestrebt*
	00:03:33-0	und der Rest müsste dann auch Dach sein und das auch	*logical constraint*
SUCHE NACH EINEM BAUTEIL - LOGISCHE ÜBERLEGUNGEN - HYPOTHESE	00:03:46-0	und weil die anderen Schrägen haben also ist das der zweite Stock / dann bleibt diese F Stücke für den ersten Stock [?] // ach das war die Rückwand	*logical constraint -> Ausschlussverfahren*
HANDLUNG			
HYPOTHESE	00:04:02-0	vorne war es noch so auf / und dann ich seh auch grad es fehlt auch keine mehr / dann könnten das hier die äh Doppelschienen sein die vorne links und rechts waren	*Bild als Orientierung - Abgleich mit dem gebauten Objekt*

Figure 6.3: Screenshot of a more general analysis of one protocol [24] containing information on the event structure (first column), temporal sequence (second column), transcribed verbalization (third column), and comments (fourth column).

One important step in the analysis is the segmentation of the protocols. The initial analysis showed that segmentation based on naturally occurring pauses[120] was not convenient because one verbalization might contain more than one thought. The analysis in most studies in discourse analysis is based on minimal units of meaning referred to as discourse units or figure (e.g. Martin & Rose 2003). Although the terminology varies, the content is defined similarly; it contains the process, i.e. the action, the participants, and possible circumstances. One information unit can be extended by linking it to subsequent information units by means of clause expansion as in example (1). In this example, the expansion is introduced by 'but' hence it is called adversative extension (Halliday & Matthiessen 2004:405).

[120] If participants did not think aloud for 30 seconds, the experimenter reminded them to keep talking.

(1) He pointed his arrow || but saw nothing. (Halliday & Matthiessen 2004:364)[121]

For the analysis, the think aloud protocols were segmented into discourse units first. These basic units were extended in some cases as in example (2) in which a conclusion is drawn as marked by 'hence'.

(2) „okay so sieht das schon aus || als würde die hier oben sein || deswegen wird das wohl hinten oben unterm Dach sein" ("okay this already looks || as if this is up here || hence this will be above in the back underneath the roof") [8]

Sometimes more than one discourse unit (including expansions) encoded one process. Example (3) illustrates this case. In this example, three discourse units encode one *hypothesis*, i.e. that of not knowing what is up and what is down. Newell and Simon (1972) introduce so-called 'problem-solving episodes'. These are categories that combine individual problem solving steps to describe them on a general level, such as 'solving the equation' (Ericsson & Simon 1993:272). Van Someren et al. (1994:20) define 'episodes' as a category that combines sequences of elements to represent a single element in the model. Although this definition of 'episode' seems adequate to represent the combination of discourse unit into one processes, such as *hypothesis*, it is too close to the term proposed by Newell and Simon (1972) to account for the current level of description. Since the analysis in this thesis focused on a more fine-grained level of individual processes than Newell and Simon (1972), the term *process unit* is introduced. In cases in which one process contained more than one discourse unit one *process unit* was defined, such as in example (4) that encodes one *aside*.

(3) „was ist unten || was ist oben || weiß ich noch nicht" ("what is up || what is down || I do not know") [11]

(4) „hm das ist schön zu spielen || und bezahlt zu werden || (pause) || das würde meinem Sohn auch gefallen damit zu spielen" ("hm this is nice to play || and be paid for it || (pause) my son would also like to play with this") [11]

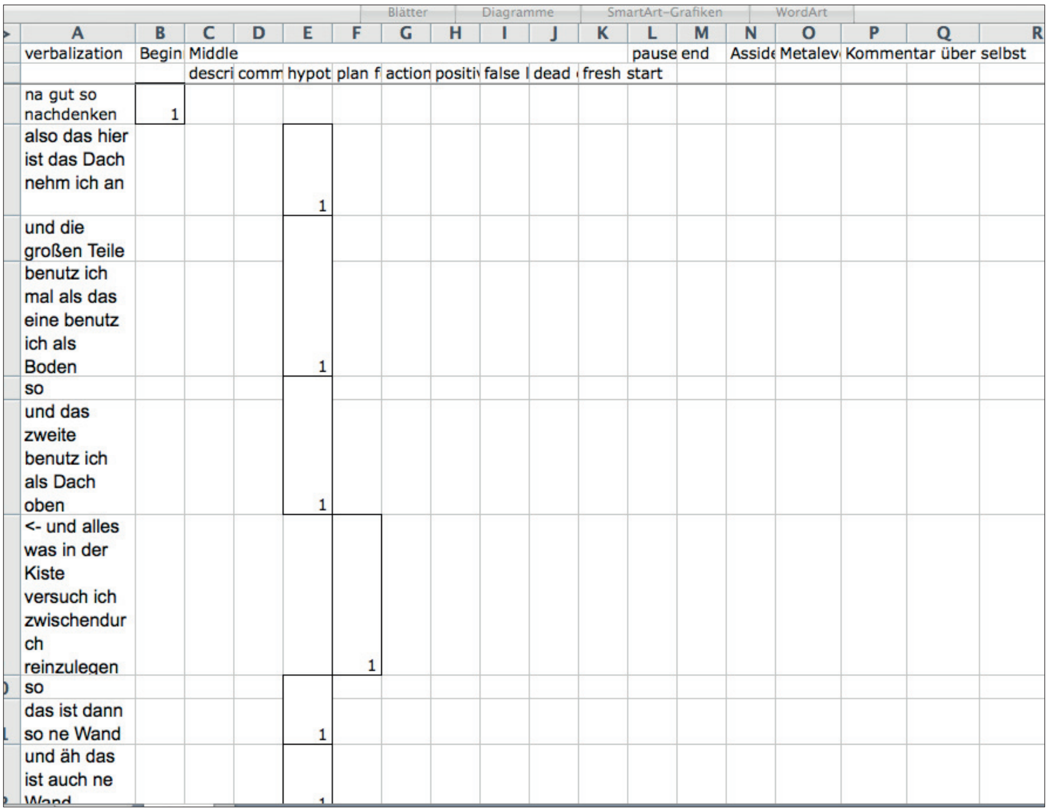
All protocols were annotated a second time based on this coding scheme. The quantitative analysis was based on this final coding scheme. In this final scheme, the global structure of beginning, middle, and end could be annotated as well as the fine-grained structure of the individual problem solving processes that were verbalized in the middle part of the protocols.

[121] The symbol || indicates the ending of one discourse unit. This notation will be adapted in the following examples.

The coding scheme contained the following categories: beginning, middle, pause, end, aside, meta-level comment, and comment on self. The beginning was defined as the introductory sequence of the protocol. The end was defined as the closing sequence of the protocol analogous to the beginning. The remaining units were defined as the middle part. The units in the middle part were annotated by the following nine process types that are defined as in the coding guideline for the coder:

a) *action*: verbs of doing, sometimes verbs are omitted, e.g. „und das dann so rum“ (“and then this this way”) [20],
b) *comment on object parts*: comment on shape, color, or salient characteristics, such as „aha sind das hier so runde Dinger“ (“oh these are such round things here”) [22], or comments on object structures „das ist jetzt schon irgendwie komisch“ (“now this is weird here somehow”) [22],
c) *dead end*: state of frustration or helplessness, i.e. having no further ideas,
d) *description of mental state*: comments on the current mental state, such as „das weiß ich nicht“ (“I don’t know that”) [36], or rhetoric questions addressed to oneself, such as „wieso ist denn da nur eins unten?“ (“why is only one down there”) [20],
e) *false lead*: this is a negative evaluation of a previous action or hypothesis,
f) *fresh start*: objects are disassembled and newly re-assembled,
g) *hypothesis*: hypotheses about what might be,
h) *plan for action*: future action is either indicated by tense, modal verbs, or the semantic sense of the utterance. In the example „also mach ich jetzt am besten die zweite Etage“ (“Thus, I build the second story now”) [24] the future sense can be inferred knowing that assembling a story involves more than one action,
i) *positive evaluation*: positive evaluation of a previous action or hypothesis.

Asides were defined as utterances that were not directly related to the task, such as references to childhood memories. *Comments at the meta-level* were defined as evaluations or general comments on the instructions, such as „also laut denken“ (I alright think aloud) [8], or the experimental setting. The category *comment on self* was exemplified by one example, namely „ich hoffe das hier wird nirgendswo gezeigt weil irgendwie ist das gerade richtig peinlich“ (“I hope this will not be shown anywhere because somehow this is very embarrassing right now”) [38]. One example of the final coding scheme is displayed in Figure 6.4.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
verbalization	Begin	Middle									pause	end	Asside	Metalev	Kommentar über selbst		
		descri	comm	hypot	plan f	action	positi	false l	dead	fresh start							
na gut so nachdenken	1																
also das hier ist das Dach nehm ich an				1													
und die großen Teile																	
benutz ich mal als das eine benutz ich als Boden				1													
so																	
und das zweite benutz ich als Dach oben				1													
<- und alles was in der Kiste versuch ich zwischendurch reinzulegen					1												
so																	
das ist dann so ne Wand				1													
und äh das ist auch ne Wand				1													

Figure 6.4: Example of the final coding scheme [12] with the transcribed verbalization in the first column and all processes categories in columns B to P.

To observe the precision of the definition of the categories and their applicability, two coders annotated according to this final coding scheme and their agreement was assessed. Both coders annotated 8 protocols; this represented 16% of the entire corpus. Overall, 851 processes were annotated in these eight protocols; this represents almost one third (30.8%) of all annotated processes. The percent agreement was 75.4% and Krippendorff's Alpha was 0.71. According to Lombard et al. (2002:593) "the criterion of .07 is often used for explorative research." The presented approach can be classified as explorative research because, to my knowledge, no research investigating problem solving processes in object assembly was reported elsewhere so far. Furthermore, Landis and Koch (1977:165) propose that Cohen's Kappa values of .61 to .80 represent substantial agreement and values of .81 to 1.0 represent almost perfect agreement. Therefore, it can be concluded that the intercoder agreement is sufficient to propose that the coding scheme is defined well enough to be used by different coders.

6.3.1.2 Assembly strategy

Inspired by Prabhu et al.'s (1995) findings on perception of assembly structures and its influence on the assembly process, the videos were annotated with regard to assembly structures. It was annotated if participants assembled the objects by following a vertical assembly strategy or a hierarchical one. This coding of assembly strategies was done for videos recorded in experiment 1 (unaided object assembly) and in experiment 2 (instruction). *Assembly strategies* were defined either as hierarchical or vertical. As outlined in the introductory section, participants following a *vertical strategy* assembled each story after the other. This could be observed when participants built the second story after building the first floor and placing the second board on top. In *hierarchical assembly* both stories were built in parallel and placed on top of each other in a final step. Participants who built houses that had one story and the roof part on top (see Figure 6.5) were classified as vertical assembly as well because the second board and the roof part were placed on top of the first floor after its completion.

Figure 6.5: Picture of a one-story house by one participant. [52]

In experiment 1 (*unaided object assembly*), all 50 videos could be annotated. However, as outlined in section 4.5.1, not all videos could be annotated for experiment 2 (*instructions*).

Participants in the *instruction* experiment assembled the dollhouse multiple times (for more detail on the procedure see chapter 4.4.2). Assembly strategies were annotated for the *explorative assembly* (EA), the *assembly prior to the instruction* (PI), and during the *instruction* (I). The opportunity to analyze the same assembly at different states of knowledge acquisition, i.e. learner status in EA and advanced learner status in PI, as well as with and without the influence of communicative intention is unique to my knowledge. Unfortunately, only 11 videos showing the silent assembly in experiment 2 (*instructions*) could be annotated but 16 instruction videos were annotated.

6.3.2 Quantitative analysis

Based on the final coding scheme, the distribution of each process category in all protocols was calculated. This frequency analysis provided an overview of the importance and prominence of the individual process categories. Furthermore, it allowed for a classification of topics that were addressed in the opening and closing sections of the assembly.

However, as problem solving is a dynamic process the analysis went beyond mere frequency accounts towards the identification of emerging and recurring patterns of problem solving processes. Example (5) illustrates the sequence *dead end – fresh start – hypothesis* that is described in the literature. By extracting all sequences with this pattern, the question of the frequency of this sequence can be investigated. Example (6) illustrates which process could follow a *fresh start*. In this specific case, it was the process of *planning an action* and an insight gained by an observation regarding specific object features. By running a frequency analysis, it could be investigated if this was a typical pattern for following fresh starts. Example (7) included the process of *comment on object features* that has not been described in the literature so far. Taken together these examples suggest that *hypotheses* and *actions* are frequently verbalized processes. To verify or falsify this first impression a frequency analysis was run.

(5) „das ist ja wirklich blöd / jetzt bin ich n bisschen überfordert wie ich das da abgeschrägt habe dann ist das richtig doof von mir [Dead end] ok nochmal sortieren [Fresh start] die Schrägen müssen irgendwie unter dieses Dach und zwar so dass das Dach da rauf passt [Hypothesis]“ [23] (“this is really bad now I feel a little overwhelmed about the way I placed this this is really stupid of me [Dead end] alright sort again [Fresh start] the bevels need to go under this roof somehow and in a way that the roof can be placed on top” [Hypothesis])[122]

(6) “ich mach alles nochmal rückwärts [Fresh start] ich tausche die Stockwerke glaube ich ja oder [Plan for action] ja doch klar [Positive evaluation] weil hier ist es so abgeschrägt das Dach [Comment on object features]“ [47] (“I will do it in reverse order again [Fresh start] I will change the stories I think yes or [Plan for action] yes certainly [Positive evaluation] because the roof is bevelled here” [Comment on object parts])

(7) „so da fehlt so'n bisschen Stabilität drin [Hypothesis] bisschen was dazwischen tun [Action] was passiert dann hm hab ich zwei geteilte Räume [Comment on object features] aber es ist irgendwie zu viel Zwischenplatz“ [Negative evaluation] [19] (“so this misses some stability here [Hypothesis] put something in between [Action] what happens then hm I have two

[122] In some cases, it is very difficult to translate the German phrases. Thus, the translations are not literal but rather content based ones.

separated rooms [Comment on object features] but somehow there is too much space in between" [Negative evaluation])

In order to extract recurring patterns of process sequences in the protocols and to compare their frequency, a script was written by André Scholz[123]. The program, which was used on a Mac OS X 10.6.8, was called 'find pattern' (the script is provided in the appendix). In order to make the process sequences readable for the program, they needed to be coded into simple one letter strings; example (8) represents an imaginary process sequence and example (9) represents the corresponding input string.

(8) process sequence as annotated in the protocol: beginning – hypothesis – action – positive evaluation – action – false lead – hypothesis – end

(9) input string for 'find pattern': BHAEAFHN

The tool was used for identifying patterns of processes in the data. In order to extract patterns, the user enters one letter or a string of letters that the program needs to identify in the available input strings. This procedure was selected because the literature suggests the occurrence of specific patterns of processes in problem solving, such as 'test – operate – test – evaluate' (Miller et al. 1970) which could be translated into 'hypothesis – action – hypothesis – evaluation' in terms of the defined processes for this analysis.

Besides extracting the exact sequence of processes, the program can also assist data mining by the function 'greedy'. By choosing this function, the identified string in the input strings does not need to match the entered sequence exactly. In the case of example (9) as an input string, the pattern of BH<u>H</u>AE<u>E</u>AFH<u>H</u>N would also be identified and saved in the output file. In this sequence the underlined processes are not part of the input string. Additionally, the program counts the number of occurrences of the provided input string.

In order to test if expected process sequences can be identified and to highlight additional recurring patterns, all coded processes were translated into input strings for 'find pattern'. Then the program was run to identify and highlight recurring sequences in the available input strings. First, the program was used to search for patterns that are described in the literature on problem solving. Second, a search method referred to as *data mining*[124] was used to extract frequent combinations of

[123] Special thanks to André Scholz who wrote this program in 2009.

[124] This term has been chosen to describe the procedure that has been applied in this thesis because it also aims at discovering previously unknown patterns in the large data set of annotated processes. But it needs to be stressed that the analysis does not consider statistical significance because the aim of this exploratory approach is to identify recurring

longer sequences. The procedure was as follows, starting the search for two-process sequences all possible combinations of annotated processes were entered as two-letter strings and their frequency was identified. The same procedure was run with three-process sequences, i.e. three-letter strings.

To illustrate the procedure of *data mining*, let's assume the following example: the frequency analysis showed that *hypothesis → positive evaluation* is a frequent two-process string, encoded as HE. In order to extract the frequency of possible three-process sequences, the letters encoding all other processes are added to the two-letter string. In this case, the three-letter input strings would be HEH[125], HEF, HEE, HED, HES, HEO, HEM, HEA, HEI. Then their respective frequency was extracted. After the frequency counts for all three-process sequences were run, the most frequent ones were identified. Let us further assume that the frequency analysis revealed HEA to be a very frequent sequence. To this three-letter string one more process was added. Therefore, all possible four-letter strings were entered: HEAH, HEAF, HEAE, HEAD, HEAS, HEAO, HEAM, HEAA, HEAI. The frequency of all of these four-letter strings was extracted. Assuming that HEAH was a frequent four-process sequence one more process was added to this sequence. This procedure was repeated as long as the results seemed reasonable, in the sense that it captured a large proportion of the data. Following the same procedure as in the previous examples, the following five-letter strings were tested: HEAHH, HEAHF, HEAHE, HEAHD, HEAHS, HEAHO, HEAHM, HEAHA, HEAHI. Frequencies for all of these sequences were extracted. In the end, it showed that sequences with five processes at most could be extracted. Due to the huge amount of data that was generated by this procedure, only the most frequent sequences will be reported in the results section.

6.4 Results

In this section, the results of the performed analysis are presented. The presentation starts at the general level and proceeds to the most fine-grained level, i.e. the individual processes. In 6.4.1 the results of the general assembly strategy are presented. In 6.4.2 the findings on the global structure, i.e. beginning and ending

patterns in this specific data set but it does not claim to draw conclusions reaching beyond the specific context investigated.

[125] For a better overview, abbreviations are used. H: hypothesis, E: positive evaluation, F: negative evaluation, D: dead end, S: fresh start, O: comment on object features, M: description of mental state, A: action, I: plan for action.

sequences are reported. Categories describing introductory and closing statements are introduced and a frequency analysis is reported for the categories. In 6.4.3 the results of the analysis of problem solving processes are presented. In addition to the detailed description of each process category, the quantitative results on the frequency of the different processes are presented along with emerging patterns of those processes.

6.4.1 Assembly strategy

The results on the distribution of assembly strategies in unaided object assembly revealed that assemblers had a clear preference for vertical assembly (32 cases, 65%) over hierarchical assembly (15 cases, 31%). One participant started vertically but changed to hierarchical assembly in the course of the assembly. No influence of prior information on assembly strategy could be observed (see Table 6.2).

condition	vertical	hierarchical	both	none	N
underspecified goal	10	6	0	1	17
verbal goal	10	5	1	0	16
verbal and visual goal	12	4	0	0	16
sum	32	15	1	1	49

Table 6.2: Assembly strategies within conditions (raw frequency).

Interesting observations were made in the analysis of assembly structures in experiment 2 (*instructions*) when the silent assemblies were considered. As outlined in sub-section 6.3.1, the data allows for the investigation of assembly strategies at three different times within the instruction experiment, namely during the *first explorative assembly* (EA), in the *assembly prior to the instruction* (PI), and in the *instruction* (I) itself. The condition in EA was comparable to the *verbal and visual goal* condition in the *unaided object assembly* experiment because instructors were shown the picture and instructed to assemble the pictured doll-house for themselves but without the task to think aloud.

The analysis of the instruction videos revealed that the great majority of instructors used the vertical assembly strategy as well (12 cases, 80%). Only three instructors chose a hierarchical assembly approach (20%). One participant did not assemble the dollhouse during the instruction; thus, no assembly strategy could be annotated for this participant. This trend was not reflected in the results on assembly strategies as observed in EA. The results rather revealed that six participants assembled the house hierarchically (54.5%) and 5 (45.5%) participants chose the *vertical assembly* strategy (see Table 6.3).

The comparison of the results recorded in the *verbal and visual goal* condition in experiment 1 and those observed in EA in experiment 2 revealed a strong trend of verbalization on assembly strategy. The likelihood ratio was $L\chi^2$ (1, N = 26) = 3.37, p = .066. Participants assembling the dollhouse without thinking aloud used the *hierarchical assembly* strategy more often than participants who thought aloud.

assembly time	vertical assembly	hierarchical assembly	both strategies	sum
first explorative assembly (EA)	5 (54.5%)	6 (45.5%)	0	11
assembly prior to the instruction (PI)	7 (63.6%)	2 (18.2%)	2 (18.2%)	11
instruction (I)	12 (80.0%)	3 (20.0%)	0	15

Table 6.3 Raw frequency and percentage of assembly strategies observed at different times within experiment 2.

Observing this great difference between strategies chosen in EA and instructions, the assembly strategies for PI were annotated as well. The observed assembly strategies in PI highlighted an interesting phenomenon. From those six participants who assembled the dollhouse in a hierarchical fashion during EA only two used the same strategy in PI. Two participants started of hierarchically but switched to a vertical procedure during the assembly and one participant started with the vertical strategy in PI right away (participant 10). One participant changed from hierarchical to vertical assembly but she did not assemble the dollhouse once more during the instruction (participant 8) (see Table 6.4 for an overview).

To summarize, six participants did not change their assembly strategy during the experiment whereas four participants changed from hierarchical to vertical assembly. One participant instructed only verbally; thus, no assembly strategy could be observed in I. The results revealed a strong trend of changing from hierarchical to vertical assembly during consecutive assemblies. However, this trend did not reach statistical significance, $L\chi^2$ (4, N = 37) = 8.99, p = .06.

participant ID	first explorative assembly (EA)	assembly prior to instruction (PI)	instruction (I)
2	hierarchical	starts hierarchically change to vertical	vertical
3	vertical	vertical	vertical
4	vertical	vertical	vertical
5	hierarchical	hierarchical	hierarchical
6	hierarchical	hierarchical	vertical

8	hierarchical	vertical	no assembly
9	hierarchical	starts hierarchically changes to vertical	hierarchical
10	hierarchical	vertical	vertical
15	vertical	vertical	vertical
20	vertical	vertical	vertical
21	vertical	vertical	vertical

Table 6.4: Summary of assembly strategies observed in first explorative assembly, assembly prior to instruction, and in the instruction.

6.4.2 Global structure in think aloud protocols and instructions

This part of the results section focuses on the presentation of the structure of the initial and final section of each protocol, i.e. the beginning and ending. The presentation will be structured as such and a comparison will be made between think aloud protocols and instructions.

6.4.2.1 Beginning

- Introductory sequences in think aloud protocols

The introductory part of each of the 50 think aloud protocols was classified based on content analysis. Six categories could be identified (see Table 6.5 for an overview)[126].

Instruction related introductory phrases were most common (21 cases, 42.0%). Nine participants (18.0%) started their assembly without any further remarks about the task or the objects whereas eight participants (16.0%) conceptualized some of the given objects at that early stage. Three participants (6.0%) referred to their expectations and equally many verbalized what they saw (6.0%). Three participants (6.0%) verbalized their mental states at the beginning of the assembly. And three phrases (6.0%) did not fit into any of the categories.

[126] No intercoder agreement is reported for this coding because the coding was done by Gralla only. There were no financial resources for training another coder on this specific coding scheme.

category label	description	example
instruction related	Parts of the instruction were repeated, such as what to do (e.g. think aloud, assemble a sensible object) and objects mentioned (e.g. box).	a) „also laut denken“ (“alright think aloud”) [8] b) „ok also die Materialien in der Kiste dürfen ja auch verwendet werden” (“ok the objects in the box can also be used”) [20] c) „ok ein zweistöckiges Puppenhaus ohne Schrauben“ (“ok a two-story dollhouse without screws”) [53]
con-ceptualization	Object parts were conceptualized within the goal domain after first perception.	a) „… es sieht n bisschen aus wie n Puppenhaus” (“it kind of looks like a dollhouse2) [23] b) „ok das ist das Dach“ (“ok this is the roof”) [36] c) „ja da sind zwei Stockwerke“ (“yes there are two stories”) [49]
expectation	The assembler started by verbalizing an expectation.	a) „da fehlt aber was oder” (“there is something missing, isn’t there”) [48] b) „gut Gedankenexperiment schön” (“well a mental experiment fine”) [57]
action	The assembler started by verbalizing an action.	a) „mhm alles raus hier“ (“mhm everything out here”) [19] b) „ok ich hab jetzt hierdrei Teile die ich jetzt sinnvoll miteinander verbinden werde“ (“alright I have got three parts here that I will connect in a sensible way”) [35] c) „so dann nehm ich erstmal den Boden“ (“so I take the floor first”) [40]
perception	Verbalizations of a visual impression.	a) „ok ich sehe einen Tisch“ (“ok I see a table”) [38] b) „oh nein die anderen Sachen sind hier drin“ (“oh no the other objects are in here”) [34]
mental state	Verbalizations of the participant’s mental state.	a) „da bin ich jetzt erstmal neugierig“ (“now I am curious”) [13] b) „ah ja also mein erster Gedanke ist dass mir das Holz gefällt“ (“ah yes well my first thought is that I like the wood”) [29]

		c) „also gut diese um Gottes Willen also ähm“ (“This phrase cannot be translated but it expresses the speaker’s puzzlement and some kind of difficulties that are not sepcified any further. ”) [25]
others	Phrases that did not belong to any of the previous categories.	„so so dies ist kein Gesitze“ (“so this is not about sitting”) [16]

Table 6.5: Summary of categories identified for initial sequences of think aloud protocols.

As motivated in chapter 4, three conditions were distinguished in the first experiment based on prior information that was provided about the goal object. Participants in the *underspecified goal* condition were not given any specific information about the goal object. Participants in the *verbal goal* condition were told that they need to assemble a two-story dollhouse. Moreover, participants in the *verbal and visual goal* condition were additionally shown a picture of the assembled dollhouse.

In their introductory phrases participants in the *verbal goal* condition tended to start their assembly right away (25.0%) if they did not refer to the instructions (see Table 6.6). Participants in the *verbal and visual goal* condition, in contrast, tended to conceptualize objects (29.4%) when they did not refer to the instructions. Participants in the *underspecified goal* condition, in contrast, only showed a tendency towards mention of instruction related information (35.3%) but for none of the other categories.

category label	underspecified goal condition	verbal goal condition	verbal and visual goal condition
instruction	6 (35.29%)	8 (50.0%)	7 (41.18%)
conceptualization	1 (5.88%)	2 (12.5%)	5 (29.41%)
expectation	2 (11.76%)	0	1 (5.88%)
action	3 (17.65%)	4 (25.0%)	2 (11.76%)
perception	2 (11.76%)	0	1 (5.88%)
mental state	2 (11.76%)	1 (6.25%)	0
others	1 (5.88%)	1 (6.25%)	1 (5.88%)
sum	17 (100%)	16 (100%)	17 (100%)

Table 6.6: Distribution of categories of initial phrases between condition (raw frequency with percentage in brackets).

- Introductory sequences in think aloud protocols

In addition to the think aloud protocols, the transcripts of the instructions were analyzed with regard to the introductory and finishing sequences. In the analysis of introductory sequences three categories were identified, namely *action*, *instruction*, and *address*. If instructors started the assembly without providing any information about the task and the goal object, these phrases have been labeled *action* (see examples (10) and (11)). In other cases, instructors told their partner explicitly what their task was; these phrases were annotated as *task* (see examples (12) and (13)). In some cases, these phrases started with a personal address (see example (14)). Furthermore, in one case the instructor asked for clarification by asking "also hörst du mich jetzt" (so you hear me now?) [8], this case was annotated as *other*.

(10) „also man nimmt die beiden Bauplatten erst einmal auseinander um die eine waagerecht vor sich hinzustellen" ("so one needs to take the two boards apart in order to put one in vertical position in front of oneself") [4]

(11) „also zunächst einmal die Teile sortieren" ("so first of all sort the objects") [5]

(12) „so ich versuch dich mal so anzuleiten ganz langsam wie ich das Haus das letzte Mal zusammengebaut hab" ("so I try to instruct you on how I assembled the house myself the last time") [16]

(13) „so die Aufgabe ist es dieses Haus zusammenzubauen" ("so the task is to assemble this house") [20]

(14) „hallo Lena ich soll dir jetzt erklären wie man dieses Haus aufbaut ähm ich geh mal davon aus dass du das gleichzeitig dann auch da drüben aufbaust" ("hello Lena I have to tell you how to assemble this house uh I assume that you assemble it at the same time over there") [12]

Instructors showed a clear preference for starting the assembly right away (nine cases, 56.25%), either with an address (four cases) or without (five cases). Six instructors (37.5%) provided their assemblers with information about the task. Half of them addressed their partner.

6.4.2.2 End

- Concluding sequences in think aloud protocols

The final sections of all 50 think aloud protocols were analyzed in the same fashion as the initial sections. The analysis revealed four categories to describe the verbalizations appropriately (see Table 6.7 for examples). Some participants ended the assembly by evaluating the assembled structure or their own abilities. These cases were annotated as *evaluation*. Some evaluations referred to information that was provided in the instructions, such as the picture or the task to assemble a sensible object. To distinguish this specific kind of evaluation from the more general category it was called *instruction*. Some participants determined the end of their task by referring to features of the construction (*construction*). And some participants did not provide any further information about why and how they determine the ending of the task; those cases were annotated as *end*.

category label	description	example
evaluation	The protocol ends with a final evaluation either of the product, the assembler's abilities, or the instructions.	a) „vielleicht n bisschen locker [pause] besser kann ichs nicht" ("maybe it's a little loose [pause] I cannot do it better") [23] b) „aber sieht aus wie'n Haus" ("but looks like a house") [37] c) „Aufgabe gelöst" ("task solved") [40]
instruction	The current state is evaluated against information provided in the instructions, e.g. the picture or the constraint to assemble a sensible object.	a) „also ich finds sinnvoll wenn man vorne reingreifen kann [pause] tja gut" ("I think it makes sense when you can reach inside from the front [pause] well") [22] b) „ok sah das ungefähr so aus [pause] ja ich glaub schon [pause] ok fertig" ("ok did it look like this [pause] yes I think so [pause] ok done") [24]
construction	The assembler finishes based on information about the construction.	a) „hoffe, dass es alles nicht umfällt und dann bin ich fertig [pause] fertig" ("I hope that this will not collapse and then I am done [pause] done") [12] b) „ich würd sagen dass es hält so" („I would say this works") [16]
end	Phrases that express the ending of an action but without any further information. Some participants referred to themselves or the object in order to determine the ending.	a) „ich bin fertig" ("I am done") [32] b) „es ist n zweistöckiges Familienhaus so [pause] ja" ("it is a two-story house [pause] yes") [50]
none	Participants did not explicitly indicate the ending of the assembly.	

Table 6.7: Summary of categories identified for final sequences of think aloud protocols.

The distribution of these categories revealed that assemblers had a clear tendency to end the assembly without much further information about why they were done (26 cases, 52.0%) (see Figure 6.6). Final sequences that contained evaluations of the assembled product or the assembler's abilities were second most frequent (eleven cases, 22.0%). No difference between conditions was observed.

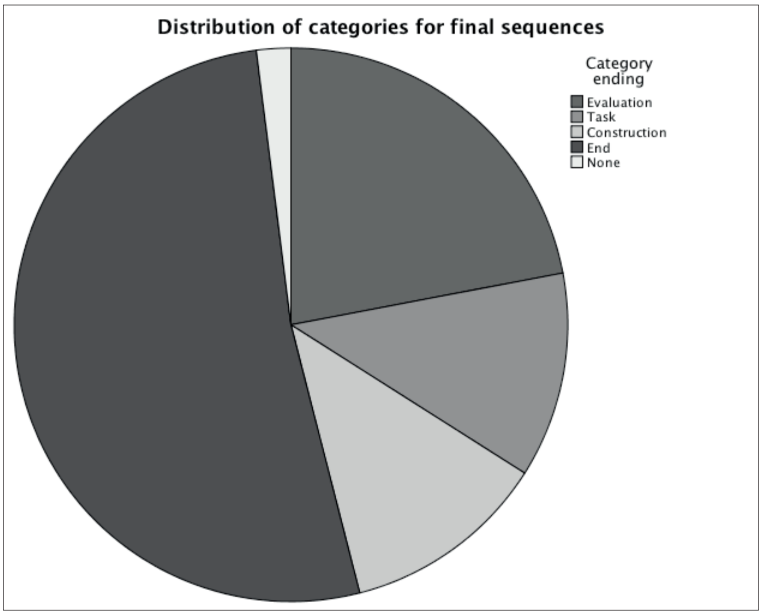


Figure 6.6: Distribution of categories on final sequences in think aloud protocols.

Taking a closer look at the two strategies that were most frequent, i.e. *end* and *evaluation*, interesting trends were revealed. The detailed analysis of the entity that was referred to when ending an assembly highlighted that assemblers either referred to themselves (e.g. "*ich* bin fertig"), to the product (e.g. "es ist n zweistöckiges *Familienhaus*"), or provided no such information (e.g. "fertig"). Comparing the distribution between these three options, a clear tendency towards the assembler determining the end was revealed (see Table 6.8). This trend was reflected in the data for participants in the *underspecified goal* and *verbal and visual goal* condition. However, the trend was different for assemblers in the *verbal goal* condition. Those provided no further information to almost the same degree as they referred to themselves as the determining factor (four and three cases respectively). Interestingly, only participants in the *underspecified goal* condition referred to the product as the determining factor.

category label	overall	underspecified goal condition	verbal goal condition	verbal and visual goal condition
subject centered	18 (69.23%)	8 (80.00%)	3 (42.86%)	7 (77.78%)
product centered	2 (7.69%)	2 (20.00%)	0	0
no further information	6 (23.08%)	0	4 (57.14%)	2 (22.22%)
sum	26 (100%)	10 (100%)	7 (100%)	9 (100%)

Table 6.8: Distribution of sub-categories of the category *end* between conditions (raw frequency with percentage in brackets).

The analysis of the nature of the entity that was evaluated highlighted that assemblers either referred to the product (e.g. „aber sieht aus wie'n Haus“ (“but looks like a house”) [37]), to their own abilities („vielleicht n bisschen locker (pause) besser kann ichs nicht“ („maybe it's a little loose (pause) I cannot do it better“)[23]), or to the instructions („Aufgabe gelöst“ (“task solved”) [40]). Among those options, the first one was the most frequently used one. It was equally distributed between conditions (see Table 6.9).

category label	overall	underspecified goal condition	verbal goal condition	verbal and visual goal condition
product	9 (81.82%)	3 (75.00%)	3 (100.00%)	3 (75.00%)
capabilities	1 (9.09%)	1 (25.00%)	0	0
instruction	1 (9.09%)	0	0	1 (25.00%)
sum	11 (100%)	4 (100%)	3 (100%)	4 (100%)

Table 6.9: Distribution of sub-categories of the category *evaluation* between conditions (raw frequency with percentage in brackets).

- <u>Concluding sequences in instructions</u>

The analysis of 16 instructions revealed that the categories *evaluation*, *end*, and *none* could be adapted to categorize the final sections. The distribution revealed a similar pattern as in the think aloud protocols, i.e. half of all instructors ended without any further information (eight cases) as in examples (15) and (16). Five (31.3%) instructors evaluated the final product (example (17)) and two did not explicitly mark the end of the instruction (12.5%).

(15) „und schon sollte das Haus fertig sein“ (“and the house should be finished already") [6]

(16) „wenn ich jetzt alles richtig gemacht hab und und Glück hatte dann hast du wahrscheinlich jetzt auch n einigermaßen brauchbares Häuschen bei dir stehen“ (“and if I did everything right and and had some luck then you will have a somewhat sensible house standing there with you”) [3]

(17) „ja und dann hoffe ich, dass du das gleiche Haus hast wie ich“ (“well and then I hope that you have the same house as I do”) [2]

Instructors who ended the assembly without any further information referred to themselves as the determining factor most of the times, five cases (62.5%). One such example would be „so und ich bin fertig (pause) ich hoffe du auch“ (“so I am done (pause) I hope you are too”) [5], and 2 participants (25.0%) referred to the final product. One participant did not provide any further information. Those instructors that ended with an evaluation, exclusively referred to the product that was assembled.

6.4.3 Local structure of the middle part

In the following section the results regarding the local structure of the assembly process will be presented. First, the identified processes will be described in detail with examples for illustration. Second, the frequency analysis of these processes will be reported. Third, the observed patterns of these processes will be reported and illustrated.

6.4.3.1 Description of process types

In this sub-section, the identified processes will be presented and their content will be described in detail with examples to illustrate the linguistic representation of the respective process. First, processes that are directly related to the problem solving process will be presented. In a second step, those that are activated by the problem solving activity will be described.

6.4.3.1.1 Problem solving processes

For an overview, the following processes will be described in detail in the next sections: description of mental state, comment on object parts, hypothesis, plan for action, action, positive evaluation, false lead (negative evaluation), dead end, and fresh start.

- Description of mental state

The first process category is called *description of mental state* because it contained verbalizations of the speaker’s current mental state. Example (18) illustrates that assemblers verbalized difficulties with regard to the structure of the goal object. In example (19), in contrast, the assembler expressed an emotion, i.e. being curious.

(18) „ich kann aus den Teilen momentan wenn ich die so einzeln sehe noch nicht irgendwie schließen wie die zu was sinnvollem Ganzen zu-

sammengefügt werden könnten“ (“by looking at the individual parts I cannot think of a way to combine them into a sensible structure”) [10]

(19) „ich bin neugierig darauf was für Teile sich hier in der Kiste befinden“ (“I am curious about the parts that are in this box here”) [10]

Example (18) represented the verbalization of the process of consideration whereas examples (19) and (21) highlighted that assemblers might lack information. The difficulties that might arise are illustrated once more in examples (20) and (22).

(20) „und frag mich gerade ob ich alle einsetzen will oder ähm ob ich vielleicht doch n paar weglasse“ (“right now I am asking myself if I want to use all parts or uhm if I want to leave some unused”) [10]

(21) „und was da steht ich weiß es nicht“ (“and what is written there I do not know”) [18]

(22) „n bisschen schwierig (...) schwierig find ichs auch weil ich nicht ob diese Noppen eher besser nach oben sollten oder nach unten“ (“a little difficult (...) I find it difficult because I don’t if these parts shall be up or down”) [18]

These examples demonstrate that this category contained the assembler’s statements about his/her consciousness („das *weiß* ich nicht“ (“this I do not know”), „ich kann nicht *schließen*“ (“I cannot conclude”)), that the speaker referred to himself/herself as the source of information („*frage* ich *mich*“ (“I ask myself”), „finde *ich*“ (“I think”), „*ich* hoffe“ (“I hope”)), and that emotions were verbalized („ich bin *neugierig*“ (“I am curious”)).

- Comment on object features

The second category that was defined contained *comments on object features*, such as shape (example (26), (29), (30)), color (example (27)), or other physical properties (example (23)-(25), (28)).

(23) „es gibt wohl keine Stecker“ (“apparently there are no connecting devices”) [16]

(24) „so hier ist seltsamerweise nicht so'n Keil dran“ (“surprisingly there is no such notch here”) [16]

(25) „phhw so was is hier mit dieser Kante“ (“puh and what about this edge”) [16]

(26) „so ähm hier sind diese runden Halbkugeln drauf“ (“so uhm there are these half like balls up here”) [16]

(27) „wieder andersrum denn irgendwie ist das blau“ (“the other way around because somehow it is blue here”) [19]

(28) „weil die Bauteile unten Löcher haben dass da eigentlich n Dübel rein müsste“ (“because the object habe holes on the bottom into which connecting devices shall be inserted normally”) [18]

(29) „das sieht aus wie Ecken“ (“this looks like an edge”) [27]

(30) „ok die Zweier hab ich hier in die Mitte getan“ (“allright those two object things I have placed in the middle”) [32]

In example (23), the participant noticed that no connecting devices were provided. In order to draw such a conclusion, the participant needed to have noticed the boreholes as features of the given parts. This is stated more explicitly in (28). A more direct way of expressing a feature was provided in example (24) in which the assembler noticed that a feature (i.e. a groove) that he/she previously encountered was not characteristic for the current object.

In example (26), the assembler mentioned the ball-like objects that are attached to half of all provided objects (see Figure 6.7). In example (28) another participant also refers to the object displayed in Figure 6.7. This example illustrates the different ways in which the same object was described. The object that was referred to as ‘Zweier’ (two) in example (29) is shown in Figure 6.8. In the picture, it can be observed that the object has two column-like parts on each the side and it can be assumed that the reference ‘Zweier’ refers to this specific object feature.

Figure 6.7: Picture of one wooden object that has ball-like objects attached to it.

Figure 6.8: Picture of the wooden object that has been referred to a ‘Zweier’.

- Hypothesis

The third category that was identified was one in which participants formulated *hypotheses*. These concerned numerous different aspect of the assembly process, e.g. at the level of conceptualization, hypotheses about functions of objects are proposed (see example (31)[127]). At the level of planning, assumptions about actions that are to be taken were verbalized (see examples (34) and (35)). At the level of physical manipulation, hypotheses expressed possible configurations (see example (33)) and position (see examples (32), (36), (37)).

[127] In *hypotheses* all examples are extracted from one protocol to highlight the richness of information that is considered by one assembler.

(31) „vermutlich sind das die Böden und das hier sind Wände“ (“presumably those are the floors and those over here are the walls”) [16]

(32) „wahrscheinlich so damit die Wände hier so stehen“ (“probably this way in order for the walls to stand like this”) [16]

(33) „vielleicht kann man das irgendwann verkeilen“ (“maybe one may fasten that at some point”) [16]

(34) „und dann ähm kommt das natürlich alles aufs zweite“ (“and then uhm this goes onto the second”) [16]

(35) „und ich muss jetzt natürlich irgendwie so anordnen, dass das obere Stockwerk auch da drauf hält ohne zu wackeln“ (“and I need to arrange somehow in a way that the upper story fits onto it without bouncing”) [16]

(36) „vielleicht gehört die auch hier rein“ (“maybe this goes in here”) [16]

(37) „das gehört ja wahrscheinlich hier oben“ (“probably this belongs up here”) [16]

The examples highlight common linguistic representations that were characteristic for verbalizations categorized as *hypothesis*. As represented in the examples, the following modal adverbs were frequently identified in *hypothesis*: ‘vermutlich’ (probably), ‘vielleicht’ (maybe), ‘wahrscheinlich’ (probably), ‘natürlich’ (certainly), and ‘bestimmt’ (definitely). The modal adverbs ‘vermutlich’, ‘vielleicht’, and ‘wahrscheinlich’ mark that the information expressed in the utterance is an assumption. More specifically, ‘vermutlich’ signals that information is based on expectations and speculations (Brinkmann 1972:401). ‘Vielleicht’ is often referred to as a hedge because it weakens what is said by stressing that it is only one possibility. The modal adverb ‘natürlich’, on the contrary, marks information to be a fact (Brinkmann 1972:401). It signals that what is said confirms an expectation and can thus be classified as a marker of affirmation (for more detail see chapter 7). The modal adverb ‘bestimmt’ modifies the given information by emphasizing the speaker’s certainty about it (Brinkmann 1972:401).

Additionally, many occurrences of *‘irgendwie’* (somehow), ‘irgendwo’ (somewhere), ‘irgendwas’ (something), and ‘irgendwelche’ (some) were identified. By choosing ‘*irgend*’ (some), the speaker stresses the indefiniteness of what is being said (Erben 1972:217), i.e. it might be just any thing. This expresses uncertainty about an assignment.

- Action

The next two categories share the same goal, i.e. action taking. But the first one expresses the mental simulation of an action whereas the second one refers to the execution. Verbalizations of the first kind were categorized as *plan for action* whereas the verbalizations of *actions* were categorized as such. The first examples illustrate planning processes and those will be contrasted with instances of action execution.

Example (38) was classified as a *plan for action* because assembling the second story involves numerous sub-processes of placing objects. In example (39), the assembler verbalized the plan to assemble the second story the same way as he/she assembled the first story. This plan involved numerous sub-processes as well. In example (40) the assembler wanted to place the second board on the object that he/she assembled so far but he/she needed to stabilize the construction first. This utterance contained a number of plans. The speaker of example (41) expressed a clear intention to act right away by positioning the roof on top.

(38) „also mache ich jetzt am besten die zweite Etage“ (“thus, I work on the second story now”) [24]

(39) „ok dann mach ich das oben noch mal genauso“ (“ok then I will do it the same way again for the upper one") [32]

(40) „aber bevor ich jetzt hier auf so wackeliges Gerüst noch ne zweite Platte stell muss ich mir irgendwie n Rezept ausdenken dass das halten kann“ (“but before I put a second board on this loose construction I need to have an idea about how to keep it fastened”)[128] [27]

(41) „aber jetzt werd ich das Dach oben auf setzen“ (“but now I will put the roof on top”) [27]

In verbalizations of actions, in contrast, the action was expressed to start at the moment of speaking by verbs of action[129] such as ‘take’, ‘put’, and ‘position’, for example. In example (42)[130], the execution of ‘take out’ started at the moment of speaking and ended when everything was removed from the container[131]. The same pattern described the action expressed in example (43); a specific object (‘that’) was put into a specific location (‘in the corner’). Then the action was completed. In examples (45) and (46) the assembler was not specific about the action that he/she performed in terms of the schema described before he/she rather stated that he/she will try something. In examples (44) and (47)-(49) the location was central whereas example (50) exemplified a specific action.

[128] In cases in which it is very difficult to translate the German phrases the translations are not literal but content based ones.

[129] In chapter 7 of this thesis the class of verbs will be defined as *verbs of doing and happening* but for the illustration at this point the less specific terminology of ‘action verbs’ seems adequate.

[130] All examples in the *action* category, except for the last one, were extracted from one protocol to illustrate the range of actions that were verbalized.

[131] It is very likely that the participant referred to the box and wanted to ‘take out’ the objects that were contained in it but since this reference was not clearly stated the subordinate term container was used at the description level.

(42) „dann *hol* ich erstmal alles raus“ (“then I take everything out first”) [27]

(43) „das *stell* ich jetzt also mal in die Ecke“ (“this I out in the edge now”) [27]

(44) „muss ich jetzt irgendwie hierher *stellen*“ (“I need to put it here somehow”) [27]

(45) „und die anderen Teile *ausprobieren*“ (“and try the other parts”) [27]

(46) „mal *gucken*, ob das irgendwie hinhauen kann“ (“let’s see if this can work”) [27]

(47) „also *stell* ich doch alles wieder so hin wie vorher“ (“then I will put everything the way it has been before”) [27]

(48) „jetzt diese zwei Dinger aufeinander *stapeln*“ (“now place those two things on top of each other”) [27]

(49) „*mach* ich das alles auf das wackelige Fundament“ (“I put everything on the loose basement”) [27]

(50) „ich kann jetzt gerade noch die Seiten *ausrichten*“ (“now I can even out the sides”) [16]

The analysis also highlighted characteristic linguistic representations for verbalizations categorized as *action*. Two temporal markers, namely ‘jetzt’ (now) and ‘als Erstes’ (first), and the discourse marker ‘so’ were frequently identified. Both temporal markers stress that the speaker is referring to the present situation by what he/she is saying. Erben (1972:195) describes that one function of ‘so’ is the acceptance of a conclusion that was previously drawn. In the context of the verbalization of *actions*, the use of ‘so’ can be assumed to highlight the confirmation of a previous state and the shifting focus to the new state. This assumption is supported by Tenbrink’s (2008) finding that ‘so’ frequently highlights the ending of sub-processes.

- Evaluation

After presenting *action planning* and *action taking*, the two categories of evaluation will be presented. The super-ordinate category *evaluation* contains two categories. These can be distinguished based on the conclusions that are drawn, namely positive and negative evaluation. Negative evaluations were also called *false lead*. First, examples of positive evaluations will be reported before examples for false leads will be given.

(51) „auf einmal wird’s alles n bisschen klarer“ (“suddenly everything becomes clearer”) [27]

(52) „so jetzt sind schon mal irgendwie einigermaßen die Löcher abgedeckt“ (“so now the holes are covered somehow”) [27]

(53) „also hier scheint schon was zu passen / das ist schon mal was / so“ (“alright something seems to fit here / this is something to start with / so”) [27]

(54) „aber so ähnlich wie auf dem Bild sieht das aus“ (“but it looks similar to that on the picture”) [27]

(55) „doch die passen hier rein“ (“yes they fit in here”) [27]

(56) „jetzt wackelt der nicht mehr so doll“ (“now it does not bounce that much anymore”) [34]

(57) „das passt“ (“this fits/ works”) [44]

(58) „und dann ham wir schon mal das Erdgeschoss zusammen gebaut“ (“and then we have assembled the basement already”) [58]

(59) „doch hält [pause] Wahnsinn irre“ (“it works [pause] great fantastic”) [58]

These examples illustrate that participants evaluated different aspects of the assembly task. They evaluated their own understanding of the task or the structure that they were to assemble (see example (51)). Assemblers who saw the picture of the assembled dollhouse evaluated their matching success as illustrated in example (54). Furthermore, participants evaluated the possibility that objects could be inserted or fastened (see examples (53), (55), and (57)) and if they represented a stable structure (see example (56)). Participants also evaluated sub-assemblies (see example (58)). Additionally, evaluations revealed insights about the assembler’s personal intrinsic goals, i.e. those that were not defined by the task. Example (52), for example, revealed that the participant wanted to cover the holes and that he/she succeeded.[132]

The same parameters were also evaluated negatively, i.e. goals were not achieved (see examples (60), (65), and (68)) or only in an unsatisfactory way or sub-assemblies were not stable (see examples (61) and (64)). Furthermore, assumed positions were not correct (see examples (62), (66), and (67)). However, there were also cases in which it was not clear what was evaluated by looking at the written transcript only (see example (63)).

(60) „das sieht aber auch nicht richtig gut aus – das sieht eher bescheuert aus“ (“this does not look very good – it looks rather stupid”) [58]

---

[132] The analysis of retrospective reports revealed that some participants used the boreholes in the boards to guide their choices regarding the position and orientation of the different objects. This specific participant explicitly mentions the strategy in his retrospective report by saying „irgendwann ist mir dann die Idee gekommen dass vielleicht die Bretter auch irgendwie durch die Löcher vorgeben wo welches Teil hin muss ich hatte zuvor aber erstmal die Ecken erkannt und die in die Ecken im in der Platte gestellt das dann revidiert weil die längeren Teile nicht so recht dazwischen passten und da wieder ja auf die Löcher in der Platte geachtet äh und äh versucht die mit den Löchern in den Teilen überein äh stimmen zu lassen **also Loch auf Loch**“ (at some point I had the idea that the boards somehow suggest which object needs to go where because of the holes before this I noticed the corners and put them in the corners in the boards then reversed this because the long things did not really fit in between and then paid attention to the holes in the boards uh and uh tried to match the holes with the pieces uh quasi **hole on top of hole**)

(61) „oh ah jetzt hab ichs kaputt gemacht“ (“oh now I broke it”) [58]

(62) „nee nee quatsch die gehörn hier noch nicht drauf“ (“nonsense they do not belong here”) [58]

(63) „auch irgendwie nicht“ (“no somehow not”) [27]

(64) „nee wenn ich das jetzt aufeinander stapel dann fällt alles auseinander“ (“no if I stack this then it will fall apart”) [27]

(65) „das was ich jetzt hier vorhabe nicht zu passen“ (“this what I plan to do not fit”)[133] [27]

(66) „nee dann sind diese runden Pinökel an diesen tragenden Flächen das geht nicht“ (“no then those round things are at those supporting surfaces this does not work”) [27]

(67) „also so kann ich die Ecken hier nicht aufstellen“ (“thus, I cannot place the edges here”) [27]

(68) „also kann das doch wieder nicht hinhauen“ (“thus, this does not work again”) [27]

Positive evaluations were frequently characterized by the following linguistic representations: the discourse marker ‘so’ and the affirmative particle ‘gut’ (good). ‘Gut’ was used along with different particles such as ‘ganz gut’ (alright) or in its superlative, i.e. ‘besser’ (better). Based on Erben’s (1972) interpretation of the discourse marker ‘so’ as signaling the acceptance of a conclusion, it can be argued that ‘so’ is also used as a marker of affirmation.

Negative evaluations, on the other hand, were frequently characterized by negations, such as ‘nein’ (no), ‘nee’ (nope), and ‘nicht’ (not). Additionally, numerous instances of ‘irgendwie’ (somehow) were identified. As outlined in the sub-section focusing on the description of the category *hypothesis*, this phrasing indicated uncertainty of assignment.

- Dead end

At some stages within the assembly participants did not know how to proceed in their assembly; for some participants this resulted in a *dead end* state. The following examples are provided to illustrate the level of frustration expressed in those statements.

In some utterances, participants expressed the state of helplessness (see examples (69), (71), (73), and (74)). Some participants verbalized reasons for this state, i.e. not knowing how to proceed (see examples (72) and (77)) and some stated the consequence, i.e. giving up (see example (70)). One participant verbalized that there was no progress, such as in examples (75) and (76).

[133] The German phrase is elliptic, a verb is missing. Thus, the elliptic character is adapted in the translation.

(69) „ich krieg die Krise“ (“I am lost”) [32]

(70) „mittlerweile bin ich fertig zum Aufgeben“ (“by now I am ready to give up”) [32]

(71) „scheiße ähm (pause) ähm ich hab keine Ahnung ich glaub ich werd nie fertig“ (“darn uhm (pause) uhm I have no clue I think I will never get this done”) [37]

(72) „ich weiß überhaupt nicht wo ich anfangen soll“ (“I do not know where to start”) [37]

(73) „ich weiß nicht mehr weiter“ (“I do not know how to go on”) [48]

(74) „ok ich steh aufm Schlauch“ (“this phrase means that the spaker has no clue”) [50]

(75) „äh ja ich bin gerade relativ ich komm nicht voran“ (“uh well I am relatively at the moment I do not progress”) [54]

(76) „und ich komm gerade überhaupt nicht voran weil ichs nicht verstehe wie es funktionieren soll“ (“I do not progress at the moment because I do not understand how this should be working”) [54]

(77) „und ich sehe momentan weiß ich garnicht wie ich weiter vorgehen soll“ (“and at the moment I see I do not know how to proceed”) [54]

- Fresh start

Some of those participants, who did not know how to proceed with the assembly, felt the need to take a *fresh start*. Fresh starts were rarely identified in the data, possibly because this process was seldom expressed linguistically or participants rarely chose this strategy. As linguistic representations of *fresh start* were rare, all occurrences are provided as examples.

In examples (78), (83), and (85) participants verbalized the decision to start over again. Sometimes those plans were made explicit by stating that objects were disassembled as in examples (79), (80), (81), and (84). Other participants stated that they performed the actions in reverse order to recreate a previous state (see examples (82) and (86)).

(78) „ok vielleicht fang ich noch mal von vorne an hier“ (“ok maybe I start over again”) [62]

(79) „ok noch mal die zweite Etage runter nehmen“ (“ok take down the second story once more”) [62]

(80) „und noch mal alles runter“ (“and everything down once more”) [62]

(81) „dann noch mal das Dach runter“ (“then the roof down once more”) [62]

(82) „noch mal wieder zurück“ (“back again”) [59]

(83) „also machen wir das ganze noch mal wieder von vorne“ (“thus, we will do all of this again from the start”) [59]

(84) „also noch mal das weg“ (“thus, this away once more”) [14]

(85) „ok noch mal sortieren“ (“ok sorting again”) [23]

(86) „obwohl doch ich mach alles noch mal rückwärts“ (“but yes I will do everything in reverse order once more”) [47]

6.4.3.1.2 Further activated processes

Besides those processes that reflect the problem solving activity directly, three additional categories were identified. These processes were activated by the problem solving activity. One category contained all verbalizations that reflected associations that arose while solving the task. This category was called *aside* (see examples (87), (89), (90), and (94)). The second category included all participants’ comments on the experiment. This information was subsumed in the category *meta-level comments*, such as in examples (92) and (93). The third category that was identified contained comments about the participants themselves and was thus called *comment on self* (see examples (91) and (95)). In some cases, more than one type of category was combined as illustrated in example (88). This example started with a comment at the meta-level about the task of think aloud. This reflection lead to an insight that was not related to the task at hand but of a general nature thus it was classified as an *aside*.

(87) „wenn Häuslebauer in der Realität so ähm an nen Hausbau rangehen würden glaube ich würden wir alle doch ähm lieber in Höhlen leben“ (“if constructors uhm approach building a house like this in reality uhm we would rather live in caves again I believe”) [10]

(88) „es hält wirklich auf sich bewusst zu werden was man eigentlich gerade denkt jetzt merk ich erst wie vielen glaub ich unbewusst stattfindet“ (“getting conscious about what you are think at the moment really slows you down now I realize how much happens unconsciously I think”) [10]

(89) „ich denke dass es mir gut tut meine Gedanken zu formulieren ähm ich glaub es macht mich auch langsamer aber dafür werd ich mir bewusster was ich hier eigentlich tue“ (“I think it is good for me to verbalize my thoughts uhm I think it slows me down but I get more conscious about what I am doing”)[134] [10]

(90) „ich glaub räumliches Denken war noch nie so meine Stärke aber genau dafür lohnt sich das ja auch um irgendwie dieses räumliche Denken n bisschen zu schulen“ (“I believe spatial thinking as never been my strength but that is exactly why it is good to practice spatial reasoning a little”) [10]

(91) „ich hoffe das wird hier nirgendswo gezeigt weil *irgenwie ist das gerade richtig peinlich*“ (“I hope this will not be shown anywhere because for some reason Thus, is really embarrassing here right now”) [38]

(92) „ich frag mich ob man mich versteht wenn das Holz so laut ist“ (“I am curious if I can be understood even though the wood is so loud”) [13]

[134]Interestingly, only one participant commented on the task of thinking aloud. The participant did so twice in the assembly highlighting that this topic was prominent in her mind. Therefore, both remarks are included.

(93) „ich glaub ich denk zu leise versteht man die Hälfte bestimmt nicht“ (“I think I am thinking to quietly half of it can probably not be understood”) [13]

(94) „so beim Haus da denk ich ja erst recht wieder an mein Kinderzimmer an spielen aufm Spielplatz wo n Holzhaus war“ (“so a house I am thinking about my room as a child about playing at the playground at which there has been a wooden house”) [13]

(95) „na ja Handwerk ist nicht so meine Sache“ (“well handcrafting is not my strength”) [13]

Table 6.10 summarizes the identified problem solving processes and specifies if these are directly related to the problem solving process or if they are only activated by it. Additionally, the table indicates if these processes were previously described in the literature or if they were first identified in the present data set.

process category	problem solving activity or activated by problem solving activity	previously described in the literature or new
description of mental state	problem solving activity	new
comment on object features	problem solving activity	new
hypotheses	problem solving activity	previously described
action	problem solving activity	previously described
positive evaluation	problem solving activity	previously described
negative evaluation	problem solving activity	previously described
dead end	problem solving activity	previously described
fresh start	problem solving activity	previously described
meta-level comments	activated by problem solving activity	previously described
comment on self	activated by problem solving activity	previously described
aside	activated by problem solving activity	new

Table 6.10: Summary of identified problem solving processes and those that were activated by the problem solving activity indicating their status as previously described or new.

6.4.3.2 Frequency of process types

In this section, the results of the frequency analysis of the processes defined above are presented. The presentation will follow the same structure as the previous sub-section. First, the results of the distribution of the problem solving processes will be outlined. Second, the distribution of the other three processes, namely *aside*, *meta-level comments*, and *comment on self* will be presented.

The analysis of 50 think aloud protocols resulted in the coding of 2765 processes that belong to the categories *description of mental state*, *comment on object features*, *plan for action*, *action*, *evaluation*, *dead end*, *fresh start*, and *hypotheses*. Processes were summarized by participant to evaluate the influence of prior information on frequency of process types.

The general distribution revealed that *hypotheses* were verbalized most frequently (976 cases, 35.3% of all cases) whereas *dead end* (26 cases) and *fresh start* (9 cases) were only rarely identified, i.e. both < 1.0% of all cases. *Actions* were the second most frequent process category (510 cases, 18.4%). *Positive evaluations* and *descriptions of mental state* were equally frequently found (325 and 321 cases respectively). Furthermore, a comparable number of cases of *negative evaluations* and *comments on object features* were verbalized (280 and 273 cases respectively) (see Table 6.11).

process category	raw frequency	mean (SD) raw frequency	percentage
hypothesis	976	19.52 (13.19)	35.30%
action	510	10.20 (7.89)	18.44%
positive evaluation	325	6.50 (5.33)	11.75%
description of mental state	321	6.42 (5.24)	11.61%
negative evaluation	280	5.60 (5.30)	10.13%
comment on object features	273	5.46 (4.46)	9.87%
plan for action	45	0.90 (1.89)	1.63%
dead end	26	0.52 (0.99)	0.94%
fresh start	9	0.18 (0.66)	0.33%
sum	2765		100%

Table 6.11: Distribution of process categories overall conditions (raw frequency and percentage); sorted by frequency from most to least frequent.

The analysis of distribution of problem solving processes within conditions revealed a highly significant effect of prior information on frequency of problem solving processes, χ^2 (16, N = 2765) = 45.2, p = .00. *Evaluations* were more often identified in the *underspecified goal* condition than in the other two conditions (see Table 6.12). Participants in the *underspecified goal* condition verbalized positive and negative evaluations at an almost equal frequency. Participants in the *verbal and visual goal* condition, in contrast, verbalized negative evaluations significantly less frequently (M = 4.2, SD = 1.01) than all other participants, z = -2.2, p < .01 (see Table 6.18 in the appendix for standardized residual).

process category	underspecified goal condition	verbal goal condition	verbal and visual goal condition
hypothesis	21.71 (4.44)	15.56 (2.15)	21.06 (2.49)
action	11.82 (2.53)	8.88 (1.72)	9.82 (1.39)
positive evaluation	8.12 (1.83)	5.94 (0.79)	5.42 (0.99)
description of mental state	8.59 (1.54)	4.13 (0.91)	6.41 (1.11)
negative evaluation	7.76 (1.75)	4.75 (0.76)	4.24 (1.01)
comment on object features	6.06 (1.41)	4.88 (1.00)	5.41 (0.82)
plan for action	0.94 (0.57)	1.25 (0.52)	0.53 (0.26)
dead end	0.82 (0.27)	0.63 (0.30)	0.12 (0.81)
fresh start	0.06 (0.06)	0.31 (0.25)	0.18 (0.13)

Table 6.12: Mean raw frequency of process categories between conditions (standard deviation in brackets).

Significant differences were also revealed in the distribution of *descriptions of mental state*. Participants in the *underspecified goal* condition described their mental state more frequently ($M = 8.6$, $SD = 1.5$) than participants in the *verbal goal* condition ($M = 4.1$, $SD = 0.9$), $z = -2.0$, $p < .01$. Additionally, participants in the *underspecified goal* condition verbalized *comments on object features* more frequently ($M = 6.1$, $SD = 1.4$) than participants in the *verbal goal* condition ($M = 4.9$, $SD = 1.00$).

The same trend was revealed with regard to the frequency of *hypothesis* that was more frequently verbalized by participants in the *underspecified goal* condition ($M = 21.7$, $SD = 4.4$) than by participants in the *verbal goal* condition ($M = 15.6$, $SD = 2.2$). This trend is visually displayed in Figure 6.9.

A reverse trend was observed with regard to *planning an action*; participants in the *verbal goal* condition verbalized plans significantly more often ($M = 1.3$, $SD = 0.5$) than participants in the other two conditions, $z = 2.3$, $p < .01$. A reverse trend was observed with regard to *planning an action*; participants in the *verbal goal* condition verbalized plans significantly more often ($M = 1.3$, $SD = 0.5$) than participants in the other two conditions, $z = 2.3$, $p < .01$. *Dead end* states were verbalized least frequently by participants in the *verbal and visual goal* condition ($M = 0.1$, $SD = 0.8$), $z = -2.2$, $p < .01$.

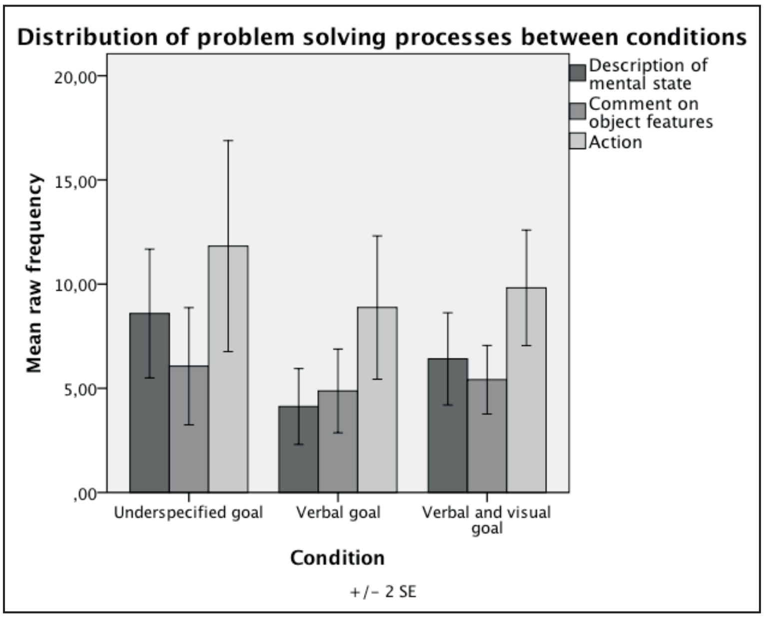


Figure 6.9: Distribution of the process categories *hypotheses*, *actions*, and *description of mental state* between conditions (mean raw frequency).

In addition to the analysis of problem solving centered processes, a frequency analysis has been run for processes that are not directly involved in the problem solving process but rather activated by the task. Overall, the results revealed that comments on the *meta-level* (55 cases) were clearly more frequent than *comments on self* (22 cases) and *asides* (14 cases) (see Table 6.13). The great majority of these processes were verbalized in the *underspecified goal* condition, i.e. 62 processes as compared to 17 processes in the *verbal goal* condition, and 12 processes in the *verbal and visual goal* condition. The general distribution is not different between conditions, $L\chi^2$ (4, N = 91) = 7.1, p = .13.

process category	asides	comment on meta-level	comment on self	Sum
underspecified goal condition	13 (92.9%)	36 (65.5%)	13 (59.1%)	62 (68.13%)
verbal goal condition	1 (7.1%)	11 (20.0%)	5 (22.7%)	17 (18.68%)
verbal and visual goal condition	0	8 (14.5%)	4 (18.2%)	12 (13.19%)
sum	14 (100%)	55 (100%)	22 (100%)	91 (100%)

Table 6.13: Distribution of process categories between conditions (raw frequency with percentage in brackets).

6.4.3.3 Sequences of problem solving processes

6.4.3.3.1 Process sequences assumed in the literature

In an initial step, the frequency of sequences of problem solving processes that are often assumed in the literature were identified in the think aloud data. In order to assess the frequency of each combination, all possible combinations of annotated processes were extracted; Table 6.14 illustrates this procedure for the case of *hypotheses*. The analysis revealed that *hypotheses* were most frequently followed by *evaluations* (negative evaluations 19.0% and positive evaluations 18.2%). In 30% of the elicited sequences hypotheses were followed by *actions*. Examples (99) and (100) illustrate the sequence *hypothesis* → *evaluation* and examples (96) to (98) illustrate the sequence *hypothesis* → *action*.

(96) „und diese Eckteile erklär ich einfach zu Trennwänden für Toilettenkabinen oder ja ne Küchenzeile [Hypothesis] mal sehen ob das hinkommt [Action]“ (“and those edgy things I declare to be separating walls for the toilets or yes a kitchenette [Hypothesis] let’s see if this works” [Action])[135] [10]

(97) „das ist schon mal das Dach alles klar [Hypothesis] dann hol ich erst mal alles da raus [Action]“ (“this is the roof al right [Hypothesis] then I will take everything out from there” [Action]) [27]

(98) „ich brauch noch ne Platte [Hypothesis] ok also die Platte darunter wieder weg [Action]“ (“I still need a board [Hypothesis] ok this board away from down there again” [Action]) [27]

(99) „so macht auch glaube ich nicht so wirklich Sinn [Hypothesis] aber geht ja darum wie es hinterher aussieht [Positive evaluation]“ (“I think this way it does not really make sense [Hypothesis] but it is all about what it looks like in the end” [Positive evaluation]) [27]

(100) „und dann wird der letzte Teil da hin passen [Hypothesis] und das sieht auch aus wie witzig ja [Positive evaluation]“ (“and then this last object will fit over there [Hypothesis] and this looks funny yes” [Positive evaluation]) [27]

sequence of problem solving processes	raw frequency	percentage
Hypothesis → Action	154	30.20%
Hypothesis → Negative evaluation	97	19.02%
Hypothesis → Positive evaluation	93	18.24%
Hypothesis → Dead end	3	0.59%
Hypothesis → Fresh start	2	0.39%

[135] The respective process category for the preceding process unit is specified in square brackets in the original and in the English translation. The number in square brackets after the original encodes the participants ID within the corpus.

Hypothesis → Comment on object feature	70	13.73%
Hypothesis → Description of mental state	80	15.69%
Hypothesis → Plan for action	11	2.16%
SUM	510	100.00%

Table 6.14: Raw frequency of all possible combinations of two-process sequences with *hypothesis* as the first process with the other categories that were identified.

Sequences that started with an *action* were frequently succeeded by *hypotheses* (38.09%) or *positive evaluations* (22.62%) (see Table 6.15).

sequence of problem solving processes	Raw frequency	Percentage
Action → Plan for action	7	2.08%
Action → Hypothesis	128	38.09%
Action → Negative evaluation	46	13.69%
Action → Positive evaluation	76	22.62%
Action → Dead end	0	0%
Action → Fresh start	1	0.30%
Action → Comment on object feature	35	10.42%
Action → Description of mental state	43	12.80%
SUM	336	100.00%

Table 6.15: Raw frequency of all possible combinations of two-process sequences with *action* as the first process with the other categories that were identified.

The same approach was adapted to evaluate the frequency of a three-process sequence that is commonly assumed in the literature (e.g. Newell & Simon 1972), namely *hypothesis* → *action* → *evaluation*. The results, displayed in Table 6.16, revealed that this sequence was indeed the most frequently observed one in think aloud data if negative and positive evaluations are subsumed into one category of *evaluation*. Examples (101) to (105) illustrate verbalizations of this sequence.

(101) „dann wär das glaub ich unten hinten [Hypothesis] mal ausprobieren [Action] so gut passt [Positive evaluation]“ (“I think this would be down there in the back [Hypothesis] let’s try it [Action] so this works fine” [Positive evaluation]) [36]

(102) „dann kann man da ja noch Wände reinbauen [Hypothesis] ah erstmal wieder abbauen [Action] passt [Positive evaluation]“ (“then one may built in walls there [Hypothesis] ah disassemble first [Action] works” [Positive evaluation]) [53]

(103) „theoretisch ist doch zumindest ne Tür da [Hypothesis] würd ich das so machen [Action] dann sind die aber wieder nicht verbunden [Negative evaluation]“ (“theoretically this is a door there [Hypothesis] I would do it like this [Action] the those are disconnected again” [Negative evaluation]) [19]

(104) „zumindest ist da ja kein Durchgang oder so [Hypothesis] kann man hier reinstellen [Action] nee wieder andersrum [Negative evaluation]“ (“at least there is no pass through or the like [Hypothesis] one my place it here [Action] no the other way around again” [Negative evaluation]) [19]

(105) „vielleicht muss doch alles aufeinander gestapelt werden [Hypothesis] wenn ich jetzt hier die Ecken darunter packe [Action] nee dann sind diese runden Pinökel an diesen tragenden Flächen [Negative evaluation]“ (“maybe everything needs to be stacked on top of each other [Hypothesis] if I put the edges down here now [Action] no then those round things are at those supporting surfaces” [Negative evaluation]) [27]

sequence of problem solving processes	Raw frequency	Percentage
Hypothesis → Action → Hypothesis	44	26.19%
Hypothesis → Action → Positive evaluation	34	20.24%
Hypothesis → Action → Negative evaluation	18	10.71%
Hypothesis → Action → Dead end	0	0
Hypothesis → Action → Fresh start	1	0.60%
Hypothesis → Action → Comment on object feature	9	5.36%
Hypothesis → Action → Description of mental state	21	12.5%
Hypothesis → Action → Action	38	22.62%
Hypothesis → Action → Plan for action	3	1.79%
SUM	168	100.00%

Table 6.16: Raw frequency of all possible combinations of the three-process sequence starting with the two-process sequence hypothesis → action.

6.4.3.3.2 Further process sequences

Besides identifying process sequences that are described in the literature, the corpus was analyzed regarding frequent patterns that have not been described so far. In the literature no distinction is made between positive and negative evaluations but this was done in this analysis to investigate whether *negative evaluations* were followed by different processes than *positive evaluations*. The analysis revealed that *negative evaluations* were most frequently followed by *hypotheses* (see left column of Table 6.17) and *positive evaluations* were most frequently followed by *hypotheses* as well (see right column of Table 6.17). The process type hypothesis was also the most frequent one following *dead end* states (see Table 6.19 in the appendix).

sequence of problem solving processes	raw frequency	sequence of problem solving processes	raw frequency
Negative evaluation → Positive evaluation	27	Positive evaluation → Dead end	2
Negative evaluation → Dead end	4	Positive evaluation → Fresh start	2
Negative evaluation → Fresh start	3	Positive evaluation → Comment on object features	30
Negative evaluation → Comment on object features	22	Positive evaluation → Description of mental state	33
Negative evaluation → Description of mental state	25	Positive evaluation → Action	52
Negative evaluation → Action	31	Positive evaluation → Plan for action	4
Negative evaluation → Plan for action	1	Positive evaluation → Hypothesis	75
Negative evaluation → Hypothesis	91	Positive evaluation → Negative evaluation	21
SUM	204	SUM	219

Table 6.17: Frequency of all possible combinations of the category *negative evaluation* and *positive evaluation* with the other categories that have been identified.

- Process sequences as observed in the data

Additionally, it was analyzed which processes followed process types that were not described in the literature but identified in the data. The analysis highlighted that sequences starting with a *comment on object features* were also frequently succeeded by *hypotheses*. The same trend was observed for two-process sequences that started with *descriptions of mental states* (see Table 6.20 in the appendix).

The same procedure was applied to search for frequent three process sequences expressing plans and actions. The search was not constrained to patterns that represent the exact string of processes but included strings that contained repetitions of the individual processes in the respective position to detect new patterns. In all extracted patterns the string that contained each process only once was the most frequent one. Thus, for a more structured overview only those strings are exemplified. The analysis highlighted that the sequences *hypothesis* → *action* → *hypothesis* (see examples (106) and (107)) and *hypothesis* → *action* → *action* (see examples (108) and (109)) were very frequent.

(106) „sie bilden überhaupt kein System [Hypothesis] na gut die Dinger n draufsetzen [Action] an dem muss das jetzt so so könnte das halten [Hypothesis]“ (“they do not represent a system [Hypothesis] well put those things on top there [Action] now needs to at this now so so it could hold like that” [Hypothesis]) [19]

(107) „und es muss ne zwote Etage hier noch rein [Hypothesis] erstmal alles rausholen [Action] achso das hießt hier n Fundament also muss das [Hypothesis]“ (“and there needs to be a second story in here [Hypothesis] take everything out first [Action] alright this means this basement here thus, this needs” [Hypothesis]) [18]

(108) „so und jetzt hab ich hier noch zwei komische Wände [Hypothesis] mh also noch bisschen hier rauf [Action] und das so stellen [Action]“ (“so and now I still have to walls [Hypothesis] mh thus, some more up here [Action] and place this like this” [Action]) [19]

(109) „irgendwas war noch in der Mitte [Hypothesis] dann kommen wir jetzt zu den beiden anderen Seiten [Action] ok also erstmal da drüben drauf [Action]“ (“something was in the middle there [Hypothesis] then we will start with the two sides here [Action] ok for a start there on top” [Action]) [19]

Furthermore, the sequence *action* → *hypothesis* → *action* (see examples (110) and (111)) was frequently identified in the data (43 times, 41.0% of all possible combinations starting with *action* → *hypothesis*) (all results are reported in Table 6.21 in the appendix).

The two-process sequence *hypothesis* → *negative evaluation* was most frequently followed by *hypothesis* (30 times, 40.0% of all possible combinations starting with *hypothesis* → *false lead*) (all results are reported in Table 6.22 in the appendix). Examples (112) and (113) illustrate such a three-process sequence.

(110) „die kann man so einzelnd verschieben [Action] ist hier noch ein Teil übrig doch das gehört dann vielleicht zum Dach [Hypothesis] versuche ich glaube ich mal [Action]“ (“one can slide those individually [Action] there is one part left here well this may belong to the roof [Hypothesis] I will try it I think” [Action]) [13]

(111) „ich nehm das trotzdem mal [Action] ist wahrscheinlich nicht so gedacht [Hypothesis] ich machs aber trotzdem mal [Action]“ (“I will take this anyways [Action] probably this is not mend to be like this [Hypothesis] I will due it anyways” [Action]) [11]

(112) „ähm ok sind das dann die Seitenwände [Hypothesis] nee die sind wahrscheinlich seitlich so [Negative evaluation] und das kommt dann bestimmt seitlich so [Hypothesis]“ (“ok those are the walls on the side then [Hypothesis] no those are probably on the sides like this [negative evaluation] and this probably belongs on the sides like that then” [Hypothesis]) [50]

(113) „und das irgendwie zusammen [Hypothesis] nee das macht ja auch keinen Sinn [Negative evaluation] oder ist es so rum [Hypothesis]“ (“and this together somehow [Hypothesis] no this does not make sense [Negative evaluation] or does it belong this way” [Hypothesis]) [50]

The application of the search method for identifying four-process sequences revealed that four combinations were frequently used. All of them contained the combination of *action* and *hypothesis*. Thus, unsurprisingly, the most frequently used combination was *hypothesis* → *action* → *hypothesis* → *action* (17 times, 41.46% of all possible combinations) (see Table 6.23 in the appendix) and in reversed order as well, i.e. *action* → *hypothesis* → *action* → *hypothesis* (11 times, 28.95% of all possible combinations) (see Table 6.24 in the appendix). Examples (114) and (115) illustrate the sequence *hypothesis* → *action* → *hypothesis* → *action* and examples (116) and (117) exemplify the pattern *action* → *hypothesis* → *action* → *hypothesis*.

Additionally, the sequence *hypothesis* → *action* → *action* → *hypothesis* has been identified 13 times (38.2% of all possible combinations) (see Table 6.26 in the appendix). The sequences *action* → *hypothesis* → *action* → *positive evaluation* was identified 11 times (29.0% of all possible combinations) (see Table 6.24 in the appendix). The sequence *hypothesis* → *action* → *positive evaluation* → *hypothesis* was also identified in 11 protocols (39.3% of all possible combinations) (see Table 6.25 in the appendix). Example (117) illustrates the pattern *action* → *hypothesis* → *action* → *positive evaluation*, whereas example (118) exemplifies the sequence *hypothesis* → *action* → *positive evaluation* → *hypothesis*.

(114) „da ist auch noch irgendwas Stift oder [Hypothesis] gut also das kommt hier unter [Action] dann muss ich [Hypothesis] nehm ich einfach irgendeins [Action]“ (“there is something a pin or [Hypothesis] well this belongs down here [Action] then I need to [Hypothesis] I take a random one” [Action]) [18]

(115) „so vorne waren an beiden Seiten so zwei parallele Säulen [Hypothesis] muss ich jetzt irgendwie hier her stellen [Action] irgendwas war noch in der Mitte [Hypothesis] dann kommen wir jetzt zu den beiden Seiten [Action]“ (“so in the front there have been two columns in parallel [Hypothesis] I need to put here somehow [Action] something was in the middle [Hypothesis] then we will start with the two sides” [Action]) [27]

(116) „so das da drauf [Action] eigentlich müsste hier glaube ich eine Ecke hin [Hypothesis] tue ich da in die Mitte [Action] zweistöckig [Hypothesis]“ (“so this on top there [Action] actually there needs to be an edge here I think [Hypothesis] I put it in the middle [Action] two stories” [Hypothesis]) [26]

(117) „so das kommt hier hin dann hält das logischerweise auch besser [Action] hätte man auch vorher drauf kommen können [Hypothesis] so den Rest getan [Action] jetzt hält vermutlich auch das zweite Stockwerk

besser [Hypothesis]“ (“so this belongs here and then this stays fixed in a better way [Action] one may have had this idea before [Hypothesis] so the rest done [Action] probably the second story will stay fixed better as well now” [Hypothesis]) [26]

(118) „ich leg jetzt erstmal alles was sich ähnlich sieht zusammen [Action] zum Teil soll dieses Gebäude nämlich auch noch irgendwelche Wände haben [Hypothesis] und jetzt möchte ich erstmal versuchen irgendwie das erste Stockwerk aufzubauen [Action] hm das geht [positive Evaluation]“ (“now I will put everything together that looks similar [Action] this building shall have walls partially as well [Hypothesis] and now I will try to assemble the first story [Action] hm this works“ [positive evaluation]) [23]

(119) „bin ich genauso weit wie vorher [Hypothesis] wenn ich den jetzt hier so dazwischen [Action] so gut [Positive evaluation] und dann hier weiter [Hypothesis]“ (“I know as much as before [Hypothesis] if I this one in between here [Action] so well [Positive evaluation] and then continue here” [Hypothesis]) [19]

Although the number of identical patterns decreased drastically by adding one more process, the ‘data mining’ technique was applied to identify frequent five-process sequences as well. The results were in line with the previous findings highlighting that the processes *action* and *hypothesis* were part of all process combinations that were most frequently used. They frequently occurred within different patterns. Those were the following sequences

- *hypothesis → action → hypothesis → action → hypothesis (five times, 31.3% of all possible combinations). This sequence is exemplified in example (120) (all results are reported in* Table 6.27 *in the appendix).*
- *action → hypothesis → action → positive evaluation → hypothesis* (six times, 85.7% of all possible combinations) (all results are reported in *Table* 6.29 in the appendix). This sequence is exemplified in example (121).
- *hypothesis → action → action → hypothesis → action* (six times, 54.6% of all possible combinations) (all results are reported in *Table* 6.28 in the appendix) This pattern is exemplified in example (122).

(120) „hm das sind doch bestimmt zu wenig Teile denn nach meinen Plänen geht das so nicht [Hypothesis] dann nehm ich den und stell den und stell den ähh so da hin [Action] nee Dach was weiß ich [Hypothesis] das kommt da hin [Action] das hat bestimmt was mit dem Dach zu tun [Hypothesis] (“hm those are probably to few parts because according to my plans this does not work [Hypothesis] then I take this one and place it and place it uh this way over there [Action] no roof I don’t know [Hypothesis] this belong there [Action] probably this is related to the roof“ [Hypothesis]) [61]

(121) „gut also das kommt hier runter [Action] dann muss ich [Hypothesis] nehm ich irgendeins [Action] so gut [Positive evaluation] noch einmal so [Hypothesis]“ (“alright this goes down here [Action] then I need to

[Hypothesis] I take just anyone [Action] so well [positive evaluation] this way once more" [Hypothesis]) [18]

(122) „schon ganz andere Teile da sein [Hypothesis] erstmal den ganzen Krämpel raus [Action] so erstmal alle Teile und [Action] ja das Dach spar ich mir [Hypothesis] und erstmal fang ich hier mit den beiden Platten an [Action]“ ("different parts here [Hypothesis] first of all everything out [Action] so first all objects and [Action] yes the roof I disregard [Hypothesis] and first I will start with these boards over here" [Action]) [14]

6.5 Discussion

Research question 1 addressed the use of vertical and hierarchical assembly strategies under different conditions. Participants who built the second story on top of the first story followed a *vertical structure*, i.e. in a bottom-up manner. Participants who built the first and second story in parallel and placed the second story on the first story after it was completed, in contrast, followed a *hierarchical assembly structure*. In theses cases the overall structure was assembled by means of at least two subassemblies, i.e. first and second story. In experiment 1 the influence of prior information on the choice of assembly strategy was addressed. In experiment 2 the choice of assembly strategy was studied with regard to the influence of assembly trials because instructors assembled the dollhouse for themselves in an explorative assembly before their product was evaluated and they were told to assemble it once more to gain expert status. Furthermore, the assembly during the instructions could also be analyzed since the instructions were video taped as well (for more detail see chapter 4).

Observations in think aloud protocols and in instructions revealed that participants assembling the dollhouse for themselves used vertical as well as hierarchical assembly. Participants who thought aloud during the assembly had a clear preference for vertical assembly (65%). Contrary to prior expectations, no influence of prior information on assembly strategy was observed. In contrast to Prabhu et al.'s (1995) proposal that perception of the product structure functions as a heuristic for the assembly procedure, these findings suggest that assembly structure was not guided by visual perception. One reason for these diverging findings might be the constraint that participants saw the picture but did not have it as a reference throughout the assembly. If the structure was not immediately salient in the 30 seconds in which the picture was presented, then the prerequisites might have been the same for all assemblers.

A trend was observed that suggests an influence of verbalization on choice of assembly strategy. Participants who assembled the dollhouse after seeing the picture but without the task to think aloud during the assembly chose the *hierarchical assembly* strategy more often than participants with the same prior information but who thought aloud. One possible explanation might be that participants observed the *hierarchical structure* of the dollhouse and used it as a guide in their assembly. Based on the proposal that hierarchical assembly involves more memory capacity to store the different sub-assemblies than vertical assembly (Prabhu et al. 1995), the findings suggest that these demands could be met by participants in the silent assembly condition but not by those who thought aloud.

Interestingly, instructors showed a marked preference for vertical assembly (80%) in instructions. This preference was more marked than the preference observed in experiment 1. Comparing the occurrence of vertical assembly in first explorative assembly (EA) and in instructions (I) an influence of communicative intention can be assumed. This impression was supported by a minor statistically significant effect of influence of communicative intention on assembly strategy.

This hypothesis was tested by annotating a third assembly, i.e. the assembly prior to the instruction task. As described in more detail in chapter 4, participants who assembled the dollhouse correctly were told to assemble it once more in preparation for a subsequent instruction (PI). Focusing on the use of *hierarchical assembly* the analysis revealed that two participants changed the strategy during the assembly and one participant chose the *vertical* strategy right away. Two participants kept using the vertical assembly. One of them choose the *vertical assembly* for the instructions as well. The other participant did not assemble the dollhouse during the instruction. This additional analysis highlighted that the observed difference between first time assembly (EA) and instructions was not an effect of communicative intention. The results rather suggest an unconscious learning effect. As participants knew the structure of the dollhouse, they did not need to construct the two stories in parallel but knew which objects belonged to which story. Thus, they sorted them, either physically or mentally, prior to the assembly and built the dollhouse in a *vertical assembly*. This finding suggests that *vertical assembly* is the default assembly strategy in assembling a two-story house.

Research question 2 addressed the general structure of the think aloud protocols and the instructions. The content analysis revealed distinct categories to describe the beginnings of think aloud protocols. First of all, participants referred to parts of the *instruction*, e.g. what needed to be done (e.g. „also laut denken“ (“alright think aloud”)) and which objects were mentioned (e.g. „ok also die Materialien in der *Box* dürfen ja auch verwendet werden“ (“ok the material in the *box* can be used as well”)). Second, participants *conceptualized* the objects they were presented with (e.g. „ok das ist das *Dach*“ („ok this is the *roof*”)). Third, participants started the task right away without any initial remarks (e.g. „ok alles raus hier“ („ok everything out here”)). The distribution of these processes revealed that the majority of participants started their assembly by referring to the *instruction*. The second most frequent categories were *actions* and *conceptualizations* of objects. The observation that assemblers most frequently recited parts of the instruction might be an effect of the experimental design. As participants were presented with the task before entering the room, their rehearsal may represent a strategy for memorization. The frequent identification of this strategy highlights the participants’ motivation to accomplish the task by paying close attention to the task constraints. This result might be an effect that is enhanced by the experimental situation because as Schoenfeld (1985) pointed out participants always want to fulfill the experimenter’s real or assumed expectations. The expectations are generally encoded in the instructions. Therefore, participants might pay close attention to them reciting them not to forget. Interestingly, the data revealed that participants who were provided with an external model, i.e. the picture of the goal object, tended to conceptualize objects in the introductory sequence more often than all other participants. This finding suggests that the external model encouraged first attempts of object conceptualization within the goal structure as early as in introductory remarks.

Instructors were observed to either start the assembly by actions right away (e.g. „also zunächst einmal die Teile sortieren“ („alright sort all parts first”)) or to provide some information about the task (e.g. „so die Aufgabe ist es dieses Haus zusammen zu bauen“ (“so the task is to build this house together”)). This finding suggests that communicative intention does not have an effect on the way in which the assembly was introduced to an addressee. The high frequency of beginnings without any introductory information can be explained by the fact that instructors assembled the dollhouse at least two times prior to the instruction. Thus, they may have understood the instructions as a new assembly in line with the other ones. Alternatively, the results might reflect an effect of the experimental design. When the instructor was introduced to his/her partner the experimenter stated that their task was to assemble the dollhouse[136]. It might be possible that some instructors did not feel the need to restate the task again.

[136] The exact instruction can be found in chapter 4.

The identified categories for beginnings in instructions correspond to the two strategies introduced by Rieser (1996), i.e. providing a global or local structure. Assemblers who started the assembly right away apparently choose the same strategy as instructors who did not provide any information about the goal structure but started with sub-assemblies as reported by Rieser (1996). Instructors who provided some information on the task offered a conceptualization of the task and the goal structure, a strategy that has been observed by Rieser (1996) as well. The strategy to offer lists of object parts as described in Daniel and Tversky (2012) was not observed in the data. One important difference between the experimental designs reported by Daniel and Tversky (2012) to that of Rieser (1996) and the one adapted in this thesis regards the instructor's knowledge about the addressee. Whereas Daniel and Tversky (2012:305) asked their participants to instruct "someone else" on the assembly, instructors in the current experimental setting as well as in Rieser's (1996) setting knew the addressee. As instructors did not know anything about the person who would read their instructions in Daniel and Tversky's (2012) experiment, the instructors apparently adopted the general schema of an instruction, i.e. provide introductory information such as objects needed and goal structure, before providing detailed assembly steps. By choosing this approach, the instructor can expect that the majority of readers will be able to follow the instructions. In cases in which an addressee is known, there is no need to follow a formulaic structure; the instructor is less constrained and may start without much introduction.

Regarding the closing of the assembly phase, it was expected that assemblers as well as instructors verbally indicated the task completion. Based on the proposed *general structure of mental processes in assembly*, it was assumed that assemblers evaluated the final product against the goal structure. It was expected that these evaluations were verbalized as comments on the final product as well as on personal performance.

The analysis of the closing parts of the think aloud protocols revealed four strategies, namely evaluations, instruction related endings, construction related endings, and endings without any further information. The first category comprised phrases in which the assemblers *evaluated* the product or their abilities (e.g. „vielleicht n bisschen locker aber besser kann ichs nicht" ("maybe a little loose but I cannot do it any better")). If participants determined the ending in relation to an external point of reference, these were either the instructions (e.g. „ok sah das ungefähr so aus ja ich glaub schon" ("ok did it look like this yes I think so")) or some feature of the construction, such as stability (e.g. „hoffe dass es alles nicht umfällt dann bin ich fertig (pause) fertig“ (“I hope that it will not collapse then I am done (pause) done”)). If no such information was verbalized, the phrases were coded as *ending* (e.g. „ich bin fertig“ (“I am done”)). The distribution of the identified categories highlighted the assemblers’ tendency to state that they were done without providing any further information about what determined this end (e.g. „ich bin fertig“(“I am done”)). If they provided information, they evaluated the assembled product. It is very likely that this tendency reflects one part of the initial task instructions as those explicitly stated „Wenn Sie mit der Aufgabe fertig sind, sagen Sie das bitte“ (“If you are finished with the task, please say so.”). No difference between assemblers was observed on closing sections with regard to prior information.

Instructors either finished with *evaluations* or no further information (*ending*). Their endings did not differ from those observed in unaided object assembly in experiment 1. Contrary to the findings reported in Daniel and Tversky (2012), no suggestions on how to use the dollhouse or encouragements to be proud of the assembly that was completed could be identified. This difference is interesting because the results on the opening sequences of instructions suggested that participants in Daniel and Tversky’s (2012) study followed the general script of an instruction. In instructions accompanying objects or technical devices no encouragement to be proud will be provided because generally instruction texts are impersonal and action oriented (Engel 1996:132), i.e. not addressing the reader personally. The results suggest that instructors in the current experiment knew about this general convention and thus concluded the assembly by focusing on the assembled product without any comments about personal feelings.

Whereas the introductory and closing sections of the protocols were analyzed so far, research question 3 focused about the content of the middle part. Eight problem solving processes and three activated processes were identified in the analysis. The eight problem solving processes resembled the expected categories and two additional ones. No instances of backtracking were identified though. Overall, the following processes were identified: *comment on object features*, *hypotheses*, *action* and *plan for action*, *evaluation* (positive and negative), *dead end*, and *fresh start*. Additionally, *descriptions of mental state* were verbalized. The later one was not described in the literature before. The identification of this process suggests that participants followed the instruction to verbalize everything what came to their minds. Therefore, this process might be specific to think aloud protocols.

Additionally, the problem solving activity resulted in the activation of the expected additional processes, namely *meta-level comments*, *comments on self*, and *asides*. Their distribution revealed that *meta-level comments* were most frequently and *asides* least often verbalized. The high frequency of *meta-level comments* can be interpreted as one feature of think aloud protocols as those have been identified in previous studies as well (Gralla et al. 2012). The low frequency of asides was surprising at first glance because the task to assemble a dollhouse was expected to activate childhood memories. However, keeping in mind that participants might have played with dollhouses but probably never assembled one themselves before, these memories might not be prominent during the assembly. The findings also suggest that participants were very engaged in the task without pondering about other topics.

The observed frequencies of the eight problem solving processes revealed the components of the general search and test procedure as proposed by Newell and Simon (1972). In the data, *hypotheses* were the most frequently verbalized processes followed by *evaluations* (negative and positive combined) and *actions*. The formulation of hypotheses represents the verbalization of search processes because the analysis revealed that hypotheses expressed conceptualizations of objects and assembled structures as well as ideas about how to proceed. Taking actions and evaluating the outcome complement the search & test sequence because they represent sub-processes of a test procedure. The comparison of process frequencies between conditions revealed a significant effect of prior information on *evaluations*, *descriptions of mental state* and *dead end states*. Participants who were given only underspecified goal information verbalized *negative evaluations* and *describe their mental state* more frequently than participants who were provided with an external model (*verbal and visual goal condition*). Furthermore, in these protocols more verbalizations of *dead end states* were identified.

These results can be interpreted as indicators of the more complex cognitive task that was posed by constructing an assembly structure from scratch; those participants formulated ideas that were not working and needed to be rejected as expressed in a negative evaluation. Participants who were provided with an external model seem to reject their hypotheses less frequently. One explanation might be that they used the memorized goal structure to access possible actions and concepts of objects – a strategy that seemed to result in acceptable and satisfactory new states in more cases than with participants in the *underspecified goal* condition. An alternative explanation might be that participants who needed to construct a representation of the goal structure focused on each step and if a hypothesis did not work, they verbally rejected it. Participants who held a representation of the goal structure may have had different ideas and if one did not work, they tried another one. Thus, their focus was on the alternative hypothesis rather than on the rejection of the first hypothesis. Therefore, they might not have verbalized negative evaluations as frequently. This second explanation could be tested by comparing the frequency of the process sequence *action* → *negative evaluation* and *action* → *hypothesis* between these two conditions. This analysis was not run so far because the analysis of process sequences was performed across conditions and not within.

The results revealed that prior information also influenced the frequency of task-associated topics. Participants who were least constrained in their goal concept verbalized processes that were activated by the problem solving activity markedly more often than participants who knew about the nature of the goal. This finding suggests that participants who needed to define the goal object reflected on the task more frequently whereas participants who were presented with a well-defined problem adapted a pre-defined structure without engaging in free associations. The results further suggest that offering a goal concept, as imprecise as only mentioning its nature, activated a pre-defined structure already.

Theoretical models of problem solving behavior assume a generic structure of sequences for problem solving processes, such as *hypothesis* → *action* → *evaluation* → *hypothesis* → *action* → *evaluation*. If the evaluation process confirms that the goal state is achieved, the task is done. However, empirical research on sequences of problem solving suggests that this assumption is too simplistic. Wedman et al. (1996), for example, studied problem solving processes involved in analogical transfer by comparing think aloud protocols of successful and unsuccessful problem solvers. In their introduction, the authors already caution that “analogical problem solving steps do not necessarily occur in a strict serial order” (Wedman et al. 1996:53). Their finding revealed that processes occurred in various sequences. The reported exemplary process graph (Wedman et al. 1996:55) illustrates very nicely that processes might be missing (here: exploring) and sequences do not represent the proposed theoretical model exactly. Furthermore, Wedman et al. (1996) do not offer any generalizations of their process graphs. This can be interpreted as an indicator of too much variance for extracting a general pattern of process sequences.

Therefore, the question of recurring patterns of problem solving processes is addressed in research question 4. The analysis of sequences of problem solving processes was performed with two aims in mind. First, process sequences as described in the literature were extracted. Second, additional recurring sequences of processes were identified. With regard to the first aim, the analysis of two process sequences revealed two frequent patterns, namely *hypothesis* → *action* and *hypothesis* → *evaluation*. These sequences represent parts of the process sequence that is assumed in the literature (e.g. de Groot 1969; Miller et al. 1970; Dörner 1987). This is also the case for the most frequently verbalized three processes sequences. The patterns *hypothesis* → *action* → *hypothesis*, *hypothesis* → *action* → *positive evaluation*, and *hypothesis* → *action* → *action* represent parts of the structure but not in the exact order. The last example highlights the common phenomenon that processes occurred more than once before another process was verbalized.

With regard to the second aim, the analysis revealed that sequences consisting of more than three processes were very difficult to identify because there was so much variability in the combination of processes. Keeping in mind that the frequency of each combination decreases as the number of possible combinations increases, the method was adapted by accepting that each process might occur multiple times at the respective spot. Under these revised constraints, the analysis revealed frequently recurring patterns of *hypothesis* → *action* → *hypothesis* → *action*, *action* → *hypothesis* → *action* → *hypothesis*, *hypothesis* → *action* → *action* → *hypothesis*, and *hypothesis* → *action* → *positive evaluation* → *hypothesis*. These sequences reflect the *search and test* procedure highlighting that it can be expressed by different process combinations.

Besides the identification of the most frequent patterns, interesting insights were gained by studying sequences that contained specific processes, such as *comment on object features* and *description of mental state*. The newly introduced process of *comments on object features* was frequently succeeded by *hypotheses*. This finding is in line with assumptions about people's problem solving behavior in real live as described by Kirsh (2009). Kirsh (2009) points out that people actively search the environment to find clues to help them solve a problem. Object features such as affordances are used as hints to define their function and suggest which actions can be performed by using them. The result that perception of object features is frequently followed by hypotheses suggests that participants also use object features to generate ideas about functions of objects as well as possible actions. This interpretation is supported by the observation that three-process sequences starting with *comments on object features* → *hypotheses* were frequently followed by *positive evaluation* or *action*. The finding that formulated hypotheses were followed by a positive evaluation suggests that actions might not have been verbalized but their outcome was evaluated subsequently. Interestingly, the rare cases of *dead end* states were either succeeded by *description of mental state* or *hypotheses*. This observation suggests that *dead end states* led to confusion and resignation as expressed in *description of mental state* or it was followed by new ideas as how to proceed as expressed in *hypotheses*.

The presented results highlight that the method was suitable to identify frequently recurring patterns. However, the analysis also revealed that the frequency of sequences, in which each process occurs only once, is only reasonably high to compare different sequences of up to three processes. With sequences containing more than three processes the likelihood increases that each of the processes or all of them occur multiple times. Therefore, it was not possible to identify one single recurring frequent pattern of problem solving processes. This finding can be interpreted in three ways.

On the one hand, participants may not have verbalized all processes and therefore, no single process sequences of the kind *hypothesis* → *action* → *positive evaluation* → *hypothesis* could be identified. This explanation needs to be considered in all analyses of verbal reports because although participants are told to verbalize everything that comes to their minds there will always be thoughts that are not verbalized (for a general discussion see chapter 3). However, given the great number of processes that was elicited in the 50 think aloud protocols, namely 2765 processes, this effect can be expected to be small. On the other hand, it may be possible that participants did not execute each individual process.

The results suggest a third interpretation, namely that the processes do not occur in the schematic way as proposed in the literature (e.g. de Groot 1969; Miller et al. 1970; Schoenfeld 1985; Dörner 1987). The observation that the proposed sequences occurred but only as one option among different sequences suggests that the literature focuses on only one possible combination of processes. This interpretation is supported by findings reported in Wedman et al. (1996) as outlined above. Their study of problem solving processes involved in analogical transfer revealed that processes occurred in various sequences and occasionally processes that were proposed by the theoretical model were missing. Thus the observed sequences of problem solving processes did not represent the proposed theoretical model exactly.

Nonetheless, a comparison between Wedman et al.'s (1996) study and Schoenfeld's (1985) study suggests that although no single generic model can be identified the identification of problem solving processes and their content highlights a structure that allows for general conclusions. These general structures are detailed enough to compare problem solving behavior in different tasks within the same problem area, i.e. analogical reasoning. Schoenfeld (1985) proposed a general model for mathematical problem solving, namely *read – analyze – explore – plan – implement – verify*. Based on this theoretical model, Wedman et al. (1996:60) identified the general sequence of *plan – implement – verify* to describe analogous problem solving in their study. Thus their finding supported the initial general model by testing it in a specific area of mathematical problem solving. This finding stresses that recurring patterns can be identified but for drawing conclusions they need to be generalized, which results in abstractions of the observed processes. Ericsson and Simon (1993) emphasize the importance of this process of abstraction in order to match verbalizations of problem solving steps to the process traces generated by the computer program applying the theoretical general model. However, Wedman et al.'s (1996) analysis further suggests that variance is observed but is not accounted for in the generalized model.

By adapting a similar approach in which variance is accounted for but repetition of processes is accepted, the findings in the present study revealed the following sequence of problem solving processes for unaided object assembly: *hypothesis → action → hypothesis* and *hypothesis → action → positive evaluation*. This generalized sequence supports one part of the structure that was identified at an early state of the analysis (see 6.2). The part that is supported involves the most frequently identified processes, i.e. *hypotheses* and *evaluations*. However, the analysis also revealed that sequences involving less frequent processes, such as *dead end* and *fresh start*, could not be generalized because they occurred in various combinations with other processes.

This means that a trend that was observed in the analysis of eleven protocols could not be observed as clearly in the data of 50 protocols. This finding suggests that there are various possibilities for solving a task and the more verbalizations of structures are analyzed the less frequent one combination will be. Research reported by Gugerty and Rodes (2007) also highlights that there are different strategies that lead to the same observable performance. In their study they analyzed participants' strategies of using a map for determining cardinal directions between two objects. They compared the observed behavior to the performance of the theoretical computational model. Behavioral data was elicited by think aloud protocols during practice trials and in terms of performance time in test trials. Their theoretical model made predictions for three different strategies. The analysis of think aloud protocols highlighted that participants used all three strategies (Gugerty & Rodes 2007:206). The comparison between the model and the observed behavior revealed that different mental strategies result in the same observable behavior.

The findings presented by Gugerty and Rodes (2007) and the results of the present study point out that single case studies, such as those reported by Newell and Simon (1972) or Bégoin-Augereau and Caron-Pargue (2010), are well suited for studying one possible way of solving a task but the comparison between different problem solvers allows to study alternative structures. However, as illustrated in the present analysis, the trade-off is a less coherent and less well-structured representation of the problem solving activity that is difficult to generalize. Therefore, future work would need to identify systematic patterns of variance, e.g. which process occurs multiple times most frequently and in which position, in order to define rules for generalizing sequences that display much variance by preserving the specific character displayed in the verbal protocols.

6.6 Conclusion

Before going into detail it needs to be stressed that the basic finding is the first important step in the analysis presented in this thesis; the identification of problem solving processes in the recorded think aloud protocols verifies the hypothesis that unaided assembly is a task that results in problem solving activity.

More specifically, the categories that are identified to describe the initial sequences of think aloud protocols support one assumption postulated in the *general structure of mental processes in assembly*, namely that goal representations can be formulated on a conceptual level or on the action level. Additionally, the frequency of the categories *action* and *conceptualization* suggests that these are common points of reference when goal representations are formulated. With regard to the motivation for achieving the formulated goal, the great number of references to the task instructions suggests that participants were highly extrinsically motivated. Furthermore, the categories identified in instructions resembled some strategies that are described in the literature. The analysis of the closing phrases in think aloud protocols highlighted that assemblers evaluated the final product in the end as a final step of the assembly task if they provided information about the reason for determining the end of the task. This finding supports the assumption about the final evaluation stage postulated in the *general structure of mental processes in assembly*.

The analysis and description of the verbalized mental and physical actions revealed categories that are commonly assumed to characterize problem solving activities. The high frequency of *hypotheses*, *evaluations*, and *actions* represent the components of the search and test procedure. This observation is strengthened by the finding that the sequence *hypothesis* → *action* → *positive evaluation*, which reflects the search and test procedure, was among the three most frequent three-process sequences. The identification of frequently recurring patterns of problem solving processes stressed that processes occur multiple times. This observation suggests that generalized models, such as the *general structure of mental processes in assembly* as proposed in 6.1.2.2 are helpful for a basic understanding but they cannot adequately represent the actual problem solving behavior that appears less serial structured. In addition to the identification of known problem solving processes, the analysis highlighted three processes that were activated by the problem solving activity and that are not systematically addressed in the literature on problem solving so far.

6.7 Appendix

6.7.1 Script 'find pattern'

```
#!/usr/bin/python
#version 0.5.1
import getopt, re, sys


def usage() :
      print '''Usage:  ./findpattern [OPTIONS] PATTERN INFILE
OUTFILE
Example: ./findpattern -g ABC file1.txt file2.txt

Options:
 -h, --help        Print this help.
 -1, -e, --exact      Search for each occurrence of PATTERN in
each line
              of INFILE. (This is the default.)
 -2, -g, --greedy     Search for each occurrence of PATTERN in
each line
              of INFILE. Each single letter in PATTERN may occur
              multiple times. Matching is greedy. e.g. given the
              pattern "ABC" and the line "AABBCCDDDDAABCCCD" the
              search matches the substrings "AABBCC", "AABCCC".
 -3, -c, --count      Only count the number of occurrences of the
given
              pattern.
   -s, --sortbynumber  When counting, sort by the number of
occurrences
              instead of the alphabetical order of matches.
              (Implies -c)
   -m, --showmatches   Show detailed information about the matches
in
              each line.
Comments:
 Filename '-' stands for StdIn or StdOut.'''


def main() :
      # get options and arguments:
      try :
            opts, args = getopt.getopt(sys.argv[1:], 'h1e2g3csm',
['help', 'exact', 'greedy', 'count', 'sortbynumber',
'showmatches'])
```

```
	except getopt.GetoptError, err :
		# print help information and exit:
		print str(err) + "\n" # will print something like
"option -a not recognized"
		usage()
		sys.exit(64)

	# check wether all arguments are given:
	if len(args) != 3 :
		print "Wrong number of arguments!\n"
		usage()
		sys.exit(64)

	# set defaults:
	count = False
	greedy = False
	sortbynumber = False
	showmatches = False

	# read options:
	for o, a in opts:
		if o in ("-h", "--help") :
			usage()
			sys.exit(0)
		elif o in ("-1", "-e", "--exact") :
			greedy = False
		elif o in ("-2", "-g", "--greedy") :
			greedy = True
		elif o in ("-3", "-c", "--count", "-s", "--
sortbynumber") :
			count = True
			matches_dict = {}
			if o in ("-s", "--sortbynumber") :
				sortbynumber = True
		elif o in ("-m", "--showmatches") :
			showmatches = True
		else :
			assert False, "unhandled option"

	if not(showmatches or count) :
		print "Nothing to do!\n"
		usage()
		sys.exit(64)

	# compile the RegEx-Pattern:
```

```
      patterN = args[0]
      if greedy :
            p = re.compile(pattern.replace('','+')[1:])
      else :
            p = re.compile(pattern)

      # open input and output files:
      if args[1] == '-' :
            infile = sys.stdin
      else :
            try :
                  infile = open(args[1], 'r')
            except IOError :
                  print "Could not open the input file %s" %
args[1]
                  sys.exit(66)
      if args[2] == '-' :
            outfile = sys.stdout
      else :
            try :
                  outfile = open(args[2], 'a')
            except IOError :
                  print "Could not open the output file %s" %
args[2]
                  sys.exit(74)

      # search line by line:
      line = infile.readline().strip()

      # write command to outfile:
      outfile.write("#%s " % sys.argv[0].lstrip('./'))
      for i in sys.argv[1:] :
            outfile.write("%s " % i)
      outfile.write("\n\n")

      while line :
            if line[:1] != '#' :
                  string = line[line.find(':') + 1:]
                  if showmatches :
                        outfile.write(line[:line.find(':') + 1])
                        matches_iter = p.finditer(string)
                        try :
                              firstmatch = matches_iter.next()
                        except StopIteration :
                              sys.exc_clear()
```

```
                        else :
                              outfile.write("%d-%d %s" %
(firstmatch.start() + 1, firstmatch.end(), firstmatch.group()))
                              for match in matches_iter :
                                    outfile.write(", %d-%d %s" %
(match.start() + 1, match.end(), match.group()))
                        outfile.write(";\n")
                  if count :
                        matches_list = p.findall(string)
                        for match in matches_list :
                              if match in matches_dict.keys() :
                                    matches_dict[match] += 1
                              else :
                                    matches_dict[match] = 1
            line = infile.readline().strip()
      if count :
            if showmatches :
                  outfile.write('\n')
            if sortbynumber :
                  matches_dict_sorted =
sorted(matches_dict.items(), key=lambda(k,v):(v,k))
            else :
                  matches_dict_sorted =
sorted(matches_dict.items())
            outfile.write("#Number of occurrences:\n\n")
            for k, v in matches_dict_sorted :
                  outfile.write("%s: %d\n" % (k, v))

      # Close all files.
      infile.close()
      outfile.close()


if __name__ == "__main__":
      main()
```

6.7.2 Local structure

process category/ Condition	description of mental state	comment on object features	action	positive evaluation	negative evaluation	hypo-thesis
underspecified	1.4	-0.7	-0.4	0.5	1.7	-1.3

goal condition						
verbal goal condition	-2.0	0.7	0.7	1.0	0.3	-0.5
verbal and visual goal condition	0.3	0.2	-0.2	-1.5	-2.2	1.9

Table 6.18: Cross-table condition × process category displaying standardized residuals.

6.7.3 Process sequences

sequence of problem solving processes	raw frequency	sequence of problem solving processes	raw frequency
Dead end → Fresh start	1	Fresh start → Comment on object features	1
Dead end → Comment on object features	1	Fresh start → Description of mental state	0
Dead end → Description of mental state	7	Fresh start → Action	3
Dead end → Action	0	Fresh start → Plan for action	0
Dead end → Plan for action	0	Fresh start → Hypothesis	5
Dead end → Hypothesis	7	Fresh start → Negative evaluation	0
Dead end → Negative evaluation	0	Fresh start → Positive evaluation	0
Dead end → Positive evaluation	2	Fresh start → Dead end	0
SUM	19	SUM	9

Table 6.19: Frequency of all possible combinations of the category *dead end* and *fresh start* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	sequence of problem solving processes	raw frequency
Comment on object features → Description of mental state	25	Description of mental state → Action	46
Comment on object features → Action	37	Description of mental state → Plan for action	2
Comment on object features → Plan for action	5	Description of mental state → Hypothesis	88
Comment on object features → Hypothesis	94	Description of mental state → Negative evaluation	32

Comment on object features → Negative evaluation	20	Description of mental state → Positive evaluation	25
Comment on object features → Positive evaluation	15	Description of mental state → Dead end	5
Comment on object features → Dead	2	Description of mental state → Fresh start	0
Comment on object features → Fresh start	0	Description of mental state → Comment on object features	23
SUM	198	SUM	221

Table 6.20: Frequency of all possible combinations of the category *comment on object features* and *description of mental state* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	percentage
Action → Hypothesis → Negative evaluation	16	15.24%
Action → Hypothesis → Positive evaluation	18	17.14%
Action → Hypothesis → Dead end	0	0
Action → Hypothesis → Fresh start	0	0
Action → Hypothesis → Comment on object feature	14	13.33%
Action → Hypothesis → Description of mental state	8	7.62%
Action → Hypothesis → Action	43	40.95%
Action → Hypothesis → Plan for action	4	3.81%
SUM	103	100.00%

Table 6.21: Frequency of all possible combinations of the category *action → hypothesis* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	percentage
Hypothesis → Negative evaluation → Hypothesis	30	40.00%
Hypothesis → Negative evaluation → Positive evaluation	10	13.33%
Hypothesis → Negative evaluation → Dead end	1	1.33%
Hypothesis → Negative evaluation → Fresh start	0	0
Hypothesis → Negative evaluation → Comment on object feature	10	13.33%
Hypothesis → Negative evaluation → Description of mental state	9	12.00%
Hypothesis → Negative evaluation → Action	14	18.67%
Hypothesis → Negative evaluation → Plan for action	1	1.33%

SUM	103	100.00%

Table 6.22: Frequency of all possible combinations of the category *hypothesis* → *negative evaluation* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	percentage
Hypothesis → Action → Hypothesis → Negative evaluation	6	14.63%
Hypothesis → Action → Hypothesis → Positive evaluation	4	9.76%
Hypothesis → Action → Hypothesis → Dead end	0	0
Hypothesis → Action → Hypothesis → Fresh start	0	0
Hypothesis → Action → Hypothesis → Comment on object feature	10	24.39%
Hypothesis → Action → Hypothesis → Description of mental state	3	7.32%
Hypothesis → Action → Hypothesis → Action	17	41.46%
Hypothesis → Action → Hypothesis → Plan for action	1	< 1%
SUM	41	100.00%

Table 6.23: Frequency of all possible combinations of the categories *hypothesis* → *action* → *hypothesis* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	percentage
Action → Hypothesis → Action → Hypothesis	11	28.95%
Action → Hypothesis → Action → Negative evaluation	6	15.79%
Action → Hypothesis → Action → Positive evaluation	11	28.95%
Action → Hypothesis → Action → Dead end	0	0
Action → Hypothesis → Action → Fresh start	1	2.63%
Action → Hypothesis → Action → Comment on object feature	2	5.26%
Action → Hypothesis → Action → Description of mental state	6	15.79%
Action → Hypothesis → Action → Plan for action	1	2.63%
SUM	38	100.00%

Table 6.24: Frequency of all possible combinations of the categories *action* → *hypothesis* → *action* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	percentage
Hypothesis → Action → Positive evaluation → Hypothesis	11	39.29%
Hypothesis → Action → Positive evaluation → Negative evaluation	1	3.57%

Hypothesis → Action → Positive evaluation → Dead end	0	0
Hypothesis → Action → Positive evaluation → Fresh start	0	0
Hypothesis → Action → Positive evaluation → Comment on object feature	4	14.29%
Hypothesis → Action → Positive evaluation → Description of mental state	2	7.14%
Hypothesis → Action → Positive evaluation → Action	9	32.14%
Hypothesis → Action → Positive evaluation → Plan for action	1	3.57%
SUM	28	100.00%

Table 6.25: Frequency of all possible combinations of the categories *hypothesis → action → positive evaluation* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	percentage
Hypothesis → Action → Action → Hypothesis	13	38.24%
Hypothesis → Action → Action → Negative evaluation	3	8.82%
Hypothesis → Action → Action → Positive evaluation	9	26.47%
Hypothesis → Action → Action → Dead end	0	0
Hypothesis → Action → Action → Fresh start	1	2.63%
Hypothesis → Action → Action → Comment on object feature	3	8.82%
Hypothesis → Action → Action → Description of mental state	4	11.76%
Hypothesis → Action → Action → Plan for action	1	2.63%
SUM	34	100.00%

Table 6.26: Frequency of all possible combinations of the categories *hypothesis → action → action* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	percentage
Hypothesis → Action → Hypothesis → Action → Hypothesis	5	31.25%
Hypothesis → Action → Hypothesis → Action → Negative evaluation	3	18.75%
Hypothesis → Action → Hypothesis → Action → Positive evaluation	4	25.00%
Hypothesis → Action → Hypothesis → Action → Dead end	0	0
Hypothesis → Action → Hypothesis → Action → Fresh start	0	0
Hypothesis → Action → Hypothesis → Action → Comment on object feature	1	6.25%

Hypothesis → Action → Hypothesis → Action → Description of mental state	2	12.5%
Hypothesis → Action → Hypothesis → Action → Plan for action	1	6.25%
SUM	16	100.00%

Table 6.27: Frequency of all possible combinations of the categories *hypothesis* → *action* → *hypothesis* → *action* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	percentage
Hypothesis → Action → Action → Hypothesis → Negative evaluation	2	18.18%
Hypothesis → Action → Action → Hypothesis → Positive evaluation	1	9.09%
Hypothesis → Action → Action → Hypothesis → Dead end	0	0
Hypothesis → Action → Action → Hypothesis → Fresh start	0	0
Hypothesis → Action → Action → Hypothesis → Comment on object feature	0	0
Hypothesis → Action → Action → Hypothesis → Description of mental state	1	9.09%
Hypothesis → Action → Action → Hypothesis → Action	6	54.55%
Hypothesis → Action → Action → Hypothesis → Plan for action	1	9.09%
SUM	11	100.00%

Table 6.28: Frequency of all possible combinations of the categories *hypothesis* → *action* → *action* → *hypothesis* with the other categories that have been identified.

sequence of problem solving processes	raw frequency	percentage
Action →Hypothesis → Action → Positive evaluation → Hypothesis	6	85.71%
Action →Hypothesis → Action → Positive evaluation → Negative evaluation	0	0
Action →Hypothesis → Action → Positive evaluation → Dead end	0	0
Action →Hypothesis → Action → Positive evaluation → Fresh start	0	0
Action →Hypothesis → Action → Positive evaluation → Comment on object feature	0	0
Action →Hypothesis → Action → Positive evaluation → Description of mental state	0	0
Action →Hypothesis → Action → Positive evaluation →	1	14.29%

Action		
Action → Hypothesis → Action → Positive evaluation → Plan for action	0	0
SUM	7	100.00%

Table 6.29: Frequency of all possible combinations of the categories *action* → *hypothesis* → *action* → *positive evaluation* with the other categories that have been identified.

7 Features of analysis: verb form in problem solving processes

7.1 State of the Art

> "The gateway through which meanings are brought together and realized in ordinary grammar is the clause; and the clause nucleus is a happening (Process + Medium, in systemic terms). So natural language represents reality as what happens, not as what exists; things are defined as contingencies of the flow." (Halliday 2009:39)

This quote by Halliday (2009b) on language and linguistics summarizes his postulate that language is the construction of meaning[137] and the representation of

[137] Halliday (2009a:60) stresses that "language has the further property that it is a semogenic system: a system that **creates** meaning" (emphasize in the original). Furthermore, Webster (2009) summarizes very cunningly that Halliday understands meaning as a formation "out of an infinite meaning potential for reflecting on the world and interacting with others in it" (Webster 2009:1).

experience. Furthermore, he stresses the dynamic nature of reality, which needs to be represented in a theory of language as well – this view will be outlined in more detail in the presentation of the systemic functional grammar approach to language. In general, this chapter focuses on the representation of the *process*, i.e. the experience, in the clause. Verb phrases encode a variety of aspects, among them are the nature of the process (*verb type*) and interpersonal elements (mood and modality). By the selection of verb mood, the speaker selects his own and the addressee's role in the speech situation. The speaker's judgments and predictions are encoded in the choice of modality (Halliday 2009:103). These features of the verb phrase will be introduced regarding the state of the art starting with the representation of modality in German. Mood is discussed as one aspect of modality in this section. Following this presentation, literature on verb types in German is summarized and discussed. Based on this overview, the guiding questions for the analysis of think aloud protocols in self-assembly and spoken instructions are formulated. Following a detailed description of the procedure for the analysis, the results are presented. The chapter closes with a summary of the main findings that are discussed in the light of the theoretical observations and descriptions that are outlined now.

Erben (1972:62) states that it is the finite verb that encodes the sentence's intention. Therefore, some more words shall be spent on the information that may be encoded in the verb in general before focusing on the literature on German modality and verb type. Boroditsky et al. (2003) subsume that verbs transmit different information for the addressee in the various languages. They summarize that verbs may encode different information about the speaker (the number of speakers, the gender), the action that is being described (is it completed or only partially done), the relevance of the action to the speaker and the addressee, politeness, and temporal aspects. In Japanese, politeness is encoded in verb form, i.e. potentially face threatening acts of asking a favor, permission, or saying out loud one's thoughts are represented by the use of morphemes (for a brief overview see Bateman 1988). In Russian, the verb encodes information on the agent's gender and on the action's stage of completion if the speaker refers to the past. The stage of completion is also marked in the future tense, a fact which Lera Boroditsky (no date) commented on by stating that in Russian a teacher makes strong commitments by telling a student to read his paper tomorrow.[138] In English information about tense is predominant. Hentschel and Weydt (2003:38) list

[138] This examples is given in a talk about the relation between language and thought (Boroditsky, no date; http://fora.tv/2010/10/26/Lera_Boroditsky_How_Language_-Shapes_Thought; 25.05.2012).

aspect as one feature that is shared by Slavic languages. They further remark that German also has aspect marking in its verb system but it is not as fine grained as in Slavic languages (Hentschel & Weydt 2003:46)[139]. Besides aspect, speakers of Turkish include information on the source of knowledge (eyewitness vs. hearsay) in the verb (Boroditsky, no date).

In German, verbs encode information on person, number, gender, mood, and tense (Hentschel & Weydt 2003:4). Hentschel and Weydt (2003) state that verbs are distinguished from other word classes by its information on time. Glinz (1973) provides an overview on the information that is encoded in a German verb (Figure 7.1) illustrating that Hentschel and Weydt's (2003) view is too limited. Glinz (1973) outlines that verb form encodes information on the role of the subject (person)[140] within the speech situation, i.e. is it the person speaking or is the subject the person who is addressed or is he/she being talked about and how many people are taking this role (number). The verb also encodes information on modality (mode), which is bound to tense as will be explained in the following sub-section. Glinz (1973:111) also includes the imperative mode in his overview. He calls it "fordernde Formen" (forms of request).

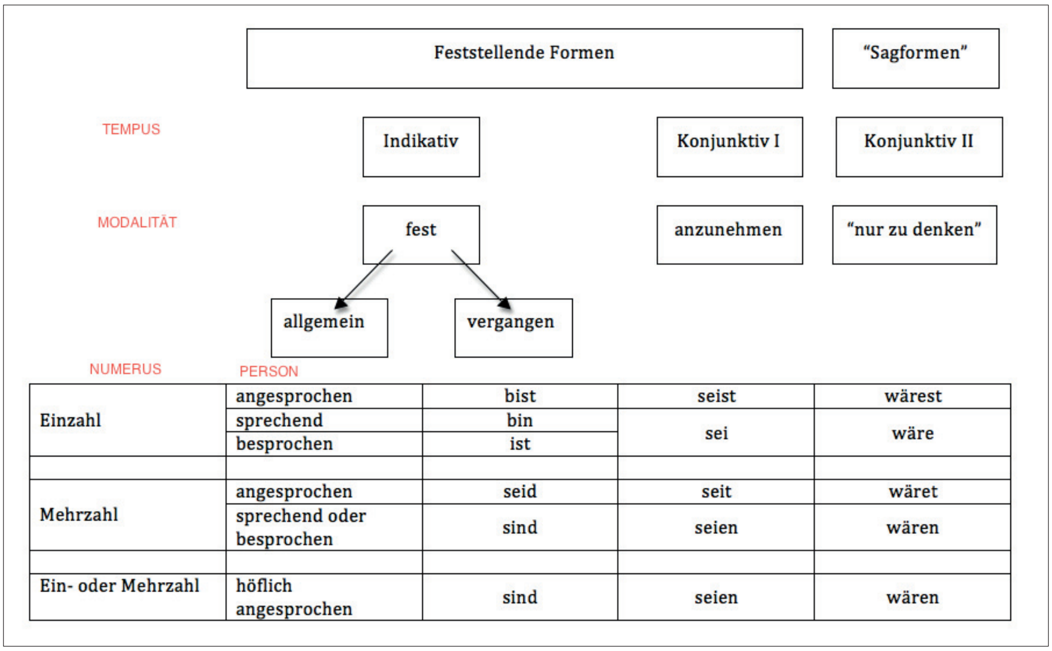


Figure 7.1: Information encoded in the German verb; here exemplified with 'be' (adapted from Glinz 1973:110,111).

[139] Buscha and Helbig (2005:62-67) also describe the different types of aspect that can be found in German. But since aspect is not an important feature of German it will not be discussed in detail in this thesis.

[140] The figure is presented in German because Glinz (1973) specifies the information encoded in German verbs and the analysis in this thesis will focus on German data. The terminology provided in brackets is in English because the English terminology is adapted throughout the theoretical discussion within this chapter.

Engel (2002:77) argues that there is a need to construct temporal relation expressed in a verb by considering tense marking and nonverbal elements. He chooses the following examples to illustrate that verbs encode information beyond the temporal aspect:

(123) Geht ihr mit zur Sitzung? (Präsens, aber Zukunft) (Are you going to the meeting?)

(124) Wie war noch ihr Name? (Präteritum, aber Gegenwart) (How was your name?)

(125) Harald wird wohl im Bett liegen. (Futur, aber Gegenwart) (Harald is probably lying in bed.)

(126) Am 11.9.2001 wird das World Trade Center in New York zerstört. (Präsens, aber Vergangenheit) (On the 11 of September the World Trade Centre was distroyed.)

In the first question, the speaker asks the addressees if they will be going to the meeting with him/her. Even though the event lies in the future, the speaker chooses the present tense. The second example represents a question that is frequently posed when unfamiliar people meet. The speaker is asking for the addressee's name, knowledge that he is lacking at the moment, but he chooses the past tense. This example illustrates how much information is encoded in tense marking. Not knowing anything about the addressee and the speaker, the reader can already infer that the two people have introduced themselves before. It can be assumed that the speaker wants to highlight the moment of recognition by choosing the past tense marking over the present tense. By choosing the future tense in the third example, the speaker signals that she is assuming Harald's current location but that he/she is not certain about it. This information is not included in the future tense marking but rather in the particle 'wohl'.[141] The final example illustrates a case in which a past event is described in the present tense. Based on examples as these, Engel (2002) concludes that verbs encode information on reality, relevance, and tense (see Figure 7.2 for an overview). Therefore, he argues that much information would be lost if the focus was exclusively on the tense aspect.

[141] The presented analysis of Engel's (2002) examples extends what is being said about their meaning potential by Engel himself. This more elaborate analysis shall give the reader a first impression of the detailed information that can be gained by a detailed linguistic analysis which goes beyond tense marking.

	reality	relevance	time
present tense	+	+	not specific
past tense	+	–	past
subjunctive mood[142]	hypothetical	–	not specific
imperative	–	+	future

Figure 7.2: Information that is encoded in the tense marking of the verb (adapted from Engel, 2002:78) (+ information encoded, - information not encoded).

This view was already stated by Erben (1972) when he remarks that

> "eine Aussage (oder Frage), deren Hauptverb im Präsens steht, ist nicht schon durch die Verbfrom in ihrer zeitlichen Geltung eingeschränkt. Erst der Kontext kann solche Einschränkungen hinzufügen." (Erben 1972:87)
> ("A statement (or question), with a main verb in the present tense is not restricted by the verb concerning its temporal dimension. It is the context that adds this restriction.")

In sum, tense is an indicator concerning the relation between time and the content expressed in an utterance. However, this relation is further refined by the context of speaking and additional markers within the sentence, such as modal adverbs, or aspects encoded in the verb, such as modality.

7.1.1 Modality

Modality encodes the speaker's evaluation concerning the reality or possibility to realize the circumstances that are described in the utterance (Halliday 1970; Hentschel & Weydt 2003:114). It may also encode the speaker's "request of a judgment of the listener on the status of what is being said" (Halliday & Matthiessen 2004:143). Languages offer different possibilities to encode modality in a sentence. In English, modality can be expressed by modal verbs, modal adjuncts[143], or a combination of both (Halliday & Matthiessen 2004:147). In German, it can be expressed by modal verbs, verb mood, modal adverbs[144], and

[142] In this category 'Konjunktiv I' and 'Konjunktiv II' have been subsumed.

[143] Halliday's category of modal adjuncts (terminology introduced in Halliday & Matthiessen 2004:147) corresponds to the class of modal adverbs. Modal adverbs are described in more detail for German later. In the following chapter the focus will be on the German modality system since the data basis for this thesis is German. For those who are interested in the modality system in English, Halliday (1970) and Halliday and Matthiessen (2004) are recommended for a thorough account.

[144] Modal adverbs are adverbs that express the speaker's evaluation of the uttered information, e.g. 'certainly' modifies a given information expressing the speaker's degree of certainty (Brinkmann 1971:362).

modifying verbs[145] (Hentschel & Weydt 2003). Brinkmann (1971:362) argues that out of these four, the first two options are the most important ones for expressing modality.

Brinkmann (1971:359) distinguishes two types of communication, namely request and wish from question and statement. Utterances belonging to the first type are uttered with the intention of changing reality, hence it is referred to as the 'realization system' (r-system). By uttering questions and statements, reality is captured and described hence it is called 'information system' (i-system). He distinguishes these two kinds of communication to show that it is not the lexical content of modal verbs and mood that encodes meaning but the type of communication in which it is used. Brinkmann (1971) points out that the same sentence can express very different modalities. The sentence 'Tom should go on holiday' may either be uttered by the boss who would express a wish or request (r-system). Or the speaker reports what someone else has said about Tom in which case it would be a statement (i-system). The same observation holds for the subjunctive mood. It may either express a wish that has not been realized (r-system) or it may be used in a question in which the speaker wants to know if the addressee could have behaved in the intended way (i-system); Figure 7.3 provides an overview on possible realizations of modality in r-type and i-type utterances.

realization system	information system
imperative	indicative
subjunctive[146]	subjunctive[147]
modal verbs	modal verb + become
	modal adverbs

Figure 7.3: Possibilities of realizing modality in German (based on Brinkmann 1971).

The language elicited in the verbal protocols is likely to resemble speech within both systems because participants may utter statements and questions connected

[145] The following German verbs are classified as modifying verbs: 'brauchen' (need), 'lassen' (let), and 'haben' (have). However, Hentschel and Weydt (2003:84) point out that it is difficult to distinguish this class of verbs from other classes based on syntactic and morphological features. The only distinguishing factor is their manner of modifying the action that is expressed by the main verb in the sentence but they do so in very different ways. Additionally, 'need' fills a gap within the system of modal verbs (Hentschel & Weydt 2003:82); a fact that encourages some linguists to classify 'need' as a modal verb.

[146] In German the subjunctive mood is expressed by 'Konjunktiv I' and 'Konjunktiv II'. 'Konjunktiv I' contains the present tense of the finite verb and 'Konjunktiv II' contains the past tense of the finite verb (Hentschel & Weydt 2003:118).

[147] In this system, 'Konjunktiv I' marks indirect speech, and 'Konjunktiv II' expresses that a condition is not fulfilled (Brinkmann 1971:366).

to the perceived situation, i.e. reality. However, they are also likely to express whishes as to how the situation needs to be changed in order to resemble a different state.

7.1.1.1 Modal verbs

As Erben (1972) stated, the finite verb encodes the sentence's intention but these intentions can be modulated by specific accompanying verbs, namely modal verbs. Weinrich (2005) states that modal verbs add information about physical, psychical, and social conditions that need to be incorporated in order to construct the meaning of the respective utterance. Grammatically, modal verbs are defined by their syntactic characteristic of taking an infinite as a verbal complement and their special rules of inflection (Eisenberg 2006:90). The class of modal verbs has a limited number of members. In German there are six core[148] members: 'wollen' (want), 'dürfen' (be allowed), 'sollen' (have to), 'mögen' (like), 'müssen' (must), and 'können' (can)[149]. Eisenberg (2006) divides those core modal verbs into two classes, namely transitive and intransitive. The verbs 'wollen', 'mögen', and 'möchten' are transitive because the subject defines the goal of the action that is being expressed in the main verb. Thus these verbs are bound to the subject's mental processes (Eisenberg 2006:97). The verbs 'müssen', 'können', 'dürfen', and 'sollen' are defined as intransitive because the initiating source of the stated action lies outside the sentence. The subject may define the goal but it is not the initiator. This can be observed on the grammatical level because in sentences with a transitive modal verb the subject needs to be explicit whereas in sentences with intransitive modal verbs the subject may be of a general nature, e.g. 'man' (one) (Eisenberg 2006:97-99).

Based on the distinction between i- and r-system, Brinkmann (1971) argues that modal verbs express preconditions for the realization of something in the r-system whereas they are used to mark the condition for the validity of the information provided in the sentence in the i-system. A further distinction is proposed by

[148] These six verbs are introduced as core modal verbs because some grammarians argue that *need*, *like*, *let*, and *become* should also be classified among the class of modal verbs (for a brief discussion see Eisenberg 2006:91).

[149] The English correspondences to the German modal verbs 'wollen', 'müssen', 'können', and 'sollen' are adapted from Nehlsen (1986) because this work investigates the correspondence between German modal verbs and their English counterparts. Nehlsen illustrates that German modal verbs have different English correspondences depending on the connotation which they have in a specific sentence. The translations which have been chosen express the most common use of the respective modal verb.

Brinkmann (1971) and Engel (2002)[150]: subject-centered use versus speaker-centered use. In *subject-centered use*, modal verbs encode the subject's relation to the uttered information. Engel (2002:92) is more specific by saying that the subject-centered use specifies the possibilities that are given to the subject by a third party. In the *speaker-centered use* modal verbs encode the speaker's attitude towards the uttered information. This distinction is important for the addressee because modal verbs encode the truth-value of a question or information that is distributed. More specifically, modals encode information regarding who is responsible for the action and which prerequisites need to be met (Brinkmann 1971:397).

Moreover, Brinkmann (1971) elaborates on the meaning of modal verbs that are unrelated to the specific situation. He distinguishes between 'wollen', 'dürfen', and 'sollen' which provide information on the driving force responsible for the realization or validity of the sentence's verbal content. The verbs 'müssen' and 'können', on the other hand, provide information on the prerequisites for the realization or validity of the sentence's verbal content. Furthermore, Engel (2002) describes different meaning potentials depending on the *subject-centered* or *speaker-centered* perspective. To provide a schematic overview on the potential meanings of modal verbs, Table 7.1 summarizes the potential meanings described in Brinkmann (1971), Engel (2002), and Weinrich (2006).

modal verb	Brinkmann (1971)	Engel (2002): subject-centered use	Engel (2002): speaker-centered use	Weinrich (2006)
'wollen' (want)	points at the subject as the driving force behind the realization or validity of the sentence's verbal content (362)[151]	expresses a subject's certain intention (93)	expresses a possibility which is based on the claim made by the subject but which is doubted by the speaker (92)	expresses the subject's interest to enforce a statement (303)

[150] In this thesis, Engel's (2002:91-92) terminology will be adapted. Brinkmann (1971) distinguishes between *objective use*, which corresponds to 'subject centered use', and *subjective use*, which corresponds to 'speaker centered use'.

[151] Numbers in brackets refer to page numbers in the respective texts.

'dürfen' (be allowed)	points at the addressee as the driving force behind the realization or validity of the sentence's verbal content (362)[152] and the speaker is content that the addressee will agree to the expressed assumption (400)	expresses the possibility of something to happen because of a granted permission (either by the speaker or a different force) (92)	only used in subjunctive mood in which it expresses an assumption (91)	signals that the forces and obstacles that are acknowledged do not hold in this specific context (302)
'sollen' (have to)	points at an (unknown) force as the driving force behind the realization or validity of the sentence's verbal content (362)	expresses a force created by a third person's demand (93)	expresses a plausibility that is based on statements made by a third person (92)	expresses that the subject's interest is put on the subject by a third party (396)
'müssen' (must)	names compulsory prerequisites for the realization or validity of the sentence's verbal content (362)	expresses a force created by external conditions (92)	expresses a strong, fact-based assumption (92)	signals that specific forces demand the enforcement of the statement
'können' (can)	names sufficient prerequisites for the realization or validity of the sentence's verbal content (362) – the possibilities for a realization are given (394)	expresses the possibility given by the speaker's abilities or based on a granted permission (92)	expresses a vague possibility (91)	external obstacles do not hold (297)
mögen (like)	depending on the mood it may either imply that the prerequisites can be met (indicative) or it points out that the subject is the driving force (subjunctive) (362) – both realizations imply that a wish is being uttered (393)	expresses the subject's wish or willingness (92)	expresses a possibility which is restricted in ways that are specified in the subsequent sentence (91)	expresses the subject's liking

Table 7.1: Overview on possible meanings encoded in modal verbs.

[152] This meaning of 'allowed' clearly encodes the function of modality as requesting a listener's judgment on the information presented in the utterance.

In addition to the six modal verbs that are described in more detail in the table, Engel (2002:92) includes 'werden' (become) as a modal verb. He argues that 'become' expresses the future realization of the sentence's content in the *subject-centered use* whereas it marks an assumption that is believed to be fulfilled in the *speaker-centered use*. Brinkmann (1971:387) acknowledges the high correspondence between 'become' and 'want' in which 'become' expresses the fact that the speaker already imagines the realization of the sentence's content. Additionally, 'become' marks an assumption to be based on experiences (Brinkmann 1971:398). However, he does not go as far as including 'become' into the class of modal verbs and thereby extending the class beyond the six core modal verbs. Another verb that is acknowledged to share some semantic features with modal verbs is 'need'. It has a strong correspondence to 'must' because it is often used to state that something must not be done (Brinkmann 1971:395; Engel 2002:92). Hentschel and Weydt (2003:82) state that there is a meaning difference between 'must not' and 'need not'. Both of them express that something is not necessary but they do so with different emphasize. They argue that the speaker uses 'must not' when he/she wants to stress that there is definitely no necessity to do something.

Given the extra information that is provided by 'braucht nicht' (need not), 'brauchen' (need) has entered the class of modal verbs but it is not part of the core modal verbs. Hence Hentschel and Weydt (2003) classify it as a modifying verb together with 'lassen' (let) and 'haben' (have). Although there are these possible extensions to the class of modal verbs, only those modal verbs that were introduced as being core modal verbs are coded as such in this thesis.

Concerning the general distribution of German modal verbs in language, results obtained by different researchers and in different corpora of German can serve as a reliable baseline. For the subsequent analysis of think aloud protocols it is specifically interesting to look at corpora of spoken German but in order to check whether spoken language differs from written language in this respect, different corpora need to be considered. Based on Welke (1965)[153], Brinkmann (1971:381) postulates three levels of frequency of modal verbs. On the first level there are 'können' and 'müssen'. On the second level there are 'wollen' and 'sollen'. The third

[153] Welke (1965:19) cites Kaeding's results, i.e. 'können' (52,384 times), 'müssen' (30,350 times), 'wollen' (27,834 times), 'sollen' (23,910 times), 'mögen' (14,406 times), and 'dürfen' (9,432 times). Brinkmann (1971:381, footnote 2) notes that the countings presented by Kaeding have been validated by the analysis performed at the Institut of German Language in Mannheim. Unfortunately in neither of the two publications any information can be found about Kaeding's data and the original publication by Kaeding is not to be found in any German library or the internet. Therefore, these numbers could not be validated beyond these quotes.

level contains 'mögen' and 'dürfen'. Brinkmann (1971) notes that the second level is twice as frequent as the third level. In addition to this report, two studies have been selected that provide an overview on different corpora and at different points in time. Wunderlich (1981:13) reports that he analyzed transcripts of spoken language of different discourse types, namely consultations[154], game instructions[155], instructions[156], institutional learning[157], and cooperative interactions[158]. All in all, he analyzed 2661 occurrences of modal verbs. The distribution of these occurrences among the different types of modal verbs is summarized in Table 7.2.

Diewald (1996:7) analyzed the use of modal verbs in written German in the weekly newspaper *Der Spiegel* and spoken German as transcribed in volume III of *Texte gesprochener deutscher Standardsprache*. She reports to have identified 839 modal verbs in the corpus and 391 modal verbs in the sub-corpus of spoken German. The distribution is very much the same in the overall corpus and the spoken sub-corpus (Diewald 1996:9). Since no notable difference is reported between the two corpora and this thesis focuses on spoken language, the results for the spoken sub-corpus are reported in Table 7.2. When considering these results as a basis of comparison, it needs to be kept in mind that these corpora contain spoken language, which is produced with a specific addressee in mind or present.

modal verb	Wunderlich (1981)	Diewald (1996)
können (can)	42.5%	37.60%
müssen (must)	29.4%	28.39%
wollen (want)	12.1%	15.60%
sollen (should)	8.5%	9.97%
dürfen (allowed)	2.6%	3.32%
möchten (want)	1.1%	not reported
mögen (like)	0.1%	5.12%

Table 7.2: Frequency of modal verbs as reported for different corpora of German.

[154] Here he looked at consultations at insurance agencies and at consultations regarding the choice of study (Wunderlich 1981:13).

[155] Those games were either dice games or card games (Wunderlich 1981:13).

[156] Those were instructions regarding first use of technical devices or specific activities (Wunderlich 1981:13).

[157] Wunderlich analyzed learning at German *Volkshochschulen* (comparable to American community colleagues) and seminars at universities (Wunderlich 1981:13).

[158] Wunderlich looked at transcripts of cooperative planning and joint action. (Wunderlich 1981:13)

7.1.1.2 Modal adverbs

Even though modal adverbs are not as frequent as modal verbs and verb tense in the expression of modality, they will be briefly introduced for two reasons. First, adverbs belonging to the category of modal adverbs are often used to distinguish different levels of certainty. However, they are not explicitly called modal adverbs, for example in Ericsson and Simon (1993). Second, Brinkmann (1971:402) observed that speakers show a tendency to combine information and evaluation in one unit. In these cases, the evaluation is encoded in modal adverbs. Brinkmann (1971:362) distinguishes between two characteristics. Modal adverbs can function as an answer (example 5) or they join other parts of the sentence to make a judgment about these parts (example 6).

(127) Did he go abroad? – Certainly.

(128) His plan, which is certainly (possibly) a good proposal, should be thoroughly evaluated.

The characteristics of marking the speaker's personal stance on the information that is provided in a sentence is achieved by three functions of modal adverbs. These functions will be briefly introduced and a list of adverbs provided by Brinkmann (1971:401) is cited.

- Function 1: Modification of the given information

by *emphasizing the certainty of a given information*: gewiß (definitely), bestimmt (certainly), sicher (certain) – in jedem Fall (at any rate), jedenfalls (anyway), unter allen Umständen (by all means) – zweifellos (certainly), unbedingt (absolutely); or

by *removing restraints*: in keinem Fall (under no circumstances), keinesfalls (on no account), keineswegs (by no means), unter (gar) keinen Umständen (under no circumstances); and weakening restraints: kaum (hardly), schwerlich (hardly)

- Function 2: Marking the information

either as a fact: tatsächlich (actually), wirklich[159] (really), natürlich[160] (of course), selbstvertsändlich (of course), bekanntlich[161] (as is generally known), offenbar (apparently), sichtlich (apparently); or

[159] Marks a quest or search for confirmation that is formulated against a voiced doubt.

[160] Signals that information confirms given expectations. Thus it is a clear marker of affirmation.

[161] The adverbs 'bekanntlich' and 'angeblich' hint at the source on which the information is based.

as an assumption: wahrscheinlich (probably), voraussichtlich[162] (expected), vermutlich (presumably), angeblich (allegedly), vielleicht [163] (maybe), wohl [164] (probably), möglicherweise (possibly)

- Function 3: Express the emotional quality of a statement (Brinkmann 1971:402): many adverbs with the suffix "–weise" (e.g. glücklicherweise (luckily), seltsamerweise (curiously))

7.1.1.3 Verb mood

Hentschel and Weydt (2003:115) describe verb mood as a statement about modality that is closely connected to the selected verb form. In German, three verb moods are distinguished, namely imperative, indicative, and subjunctive. The subjunctive has two main forms: 'Konjunktiv I' and 'Konjunktiv II'. On the grammatical level, 'Konjunktiv I' and 'Konjunktiv II' are distinguished by the verb tense of the finite verb. In the 'Konjunktiv I' the finite verb is in the present tense, for example

(129) "er komme, sie sei gekommen" (Hentschel & Weydt 2003:119).

The finite verb in the 'Konjunktiv II' is marked in the past tense

(130) "er wäre gekommen, sie würde kommen" (Hentschel & Weydt 2003:119).

However, even though the two types of subjunctive mood can be distinguished by the verb tense of the finite verb, the tempus function is not important in interpreting the subjunctive mood (Hentschel & Weydt 2003:119) because all moods assign different meaning to the content expressed in the sentence. Furthermore, the frequency of use of both possibilities reveals that the 'Konjunktiv I' is rarely used. If so, it is done in highly conventional contexts, such as in manuals and conventional forms of wishes, which results in it being almost non existent in colloquial speech (Hentschel & Weydt 2003:118). Even though the 'Konjunktiv II' is more frequent in a direct comparison, there is a tendency to substitute the subjunctive forms of the verb with a 'würde'-construction (example (10)) in spoken language (Hentschel & Weydt 2003; Thierhoff 2010).

(131) Wenn sie doch mal vorbei käme! (If she would only be dropping in!) [Konjunktiv II]

(132) Wenn sie doch mal vorbei kommen würde! (If she would only be dropping in!) [infinitive + 'würde']

---

[162] The adverbs 'voraussichtlich' and 'vermutlich' mark assumptions that are based on expectations and speculations.

[163] The adverb 'vielleicht' signals that information is an assumption and thus it is often referred to as 'hedge' in the literature on politeness.

[164] The adverb 'wohl' also signals the tentative character of an utterance.

Acknowledging that there are different forms of expressing the subjunctive mood in German, it is important to look at its meaning more closely. This can be achieved by contrasting its meaning with the meaning encoded by the other mood types. The imperative mood is the basic mode of realization, marking the content of the sentence to be the realization of an action. A request or wish is directed towards the addressee with the assumption that the action will be performed directly after it is uttered. The assumed reaction is the performance of the specified action which lies within the scope of the sentence, i.e. all prerequisites are met. The imperative may have two forms, either the imperative singular ('Sprich!') or the imperative plural ('Sprecht!'). Usually imperatives are used without a subject and there are no tense distinctions (Thieroff 2010:150).

The indicative mood is the basic mode of providing information. It makes statements within the scope that is created by the sentence and marks the content as information (Brinkmann 1971:361). The use of the indicative mood is not bound to reality, i.e. the actual truth-value of the information, but it depends on the speaker's interpretation of the situation (Brinkmann 1971:368). Three kinds of meaning of the indicative mood are distinguished:

- the things and circumstances that are described are given, i.e. assumed to be true (for example scientific insights, perception),
- it marks a goal that still needs to be accomplished, such as in "Ich gebe dir Geld, damit[165] du dich erholen kannst." (I give money to you so you can relax.) (Brinkmann 1971:369)
- it marks a prerequisite that is not yet existent in reality, e.g. "Wenn ich Geld habe, werde ich verreisen." (If I have got money, I will be traveling.) (Brinkmann 1971:369)

In contrast to statements in the imperative and the indicative mood that stay within the scope of the sentence, the subjunctive mood extends it and even goes beyond it. By the use of the 'Konjunktiv I' the speaker extends the scope that is generally assigned to the sentence. This extension is realized by introducing a third person. And the use of the 'Konjunktiv II' makes claims about things that are beyond the extended scope. These different functions of extending the sentence's scope becomes especially clear by looking at their different functions within the i-system. The 'Konjunktiv I' is the mode of indirect speech, thereby introducing a third person. And the 'Konjunktiv II' is used to mark unfulfilled prerequisits, thereby pointing at something which is beyond the current scope (Brinkmann 1971:366).

[165] The different meanings can be distinguished by looking at the conjunctions that are used (for illustration purposes these are in italics here).

Brinkmann (1971) argues against the simplification of calling the indicative mood the mode of reality and contrasting it with the subjunctive mood which is referred to as the mode of possibility as can be found, for example, in Hentschel and Weydt (2003). Brinkmann (1971:369) states that in cases in which both indicative and subjunctive mood are grammatically possible, the indicative mood is not simply the mode of reality but rather the mode of givenness. The speaker determines this givenness within his own 'universe', i.e. his world knowledge, his experience, and his judgment. Brinkmann's (1971) description of verb mood emphasizes that mood choice is highly individual and situation bound.

Based on these considerations, the analysis will distinguish between indicative, subjunctive, and imperative mood. Furthermore, modal verbs and modal adverbs will be coded for. Besides the grammatical category 'tense' verbs will be distinguished based on their semantic meaning, i.e. verb types. This latter approach is central to functional grammar as will be outlined in the following sub-section.

7.1.2 Verb types

7.1.2.1 Verb types in functional grammar

Language in the Systemic Functional Linguistics approach is organized around five features of language, namely genre, register, discourse semantics, lexico-grammar, and graphology/phonology (Martin 2011:103-4). The approach of functional grammar focuses on lexico-grammar. Halliday (1985) was the first scholar to present this concept in a comprehensive book. His studies that started in the 1960s (see Halliday 2009 for an overview of his works including the early ones[166]) were driven by the argument that traditional grammars investigated and described the structure of written texts, thereby ignoring an important mode of language: speaking. He argued that the focus on written texts ignores one fundamental difference between those two modes, namely written language presents the product of speech whereas spoken language presents a process. Based on this thought, he refers to traditional grammar as 'product grammars' (Halliday 1985:xxiii) since they cannot account for the dynamic process character of spoken language. Contrary to traditional grammarians, Halliday (1985) stresses that speaking is the more basic mode of language use. Therefore, Halliday (e.g. 1985)

[166] This book is highly recommended to readers who want to get an overview on the concepts which Halliday has been working on during his long research carreer and how these concepts are described in his numerous publications.

presents the model of functional grammar to describe the basic structure of a process grammar on language in use. He stresses that

> "the internal organization of the linguistic system has itself a functional basis, so that in order to understand the nature of language it is necessary to start from considerations of its use." (Halliday 1970:361)

Ertel and Bloemer (1975) argue for the need to describe grammatical constructions in a process oriented model as well. They start by proposing a new view on affirmative and negative sentences. In his seminal work, Halliday (1985:xxvii) presents a detailed account of functional grammar but he explicitly states that he does not provide a full model of dynamic grammar at this point but rather the first attempts in this direction. Furthermore, he points out that he presents a functional grammar of English but he is convinced that some concepts can be regarded to be universal whereas others will show to be language specific. The 'metafunctional' hypothesis, for example, is considered to be one universal feature of language. This hypothesis states that "in all languages the content systems are organized into ideational, interpersonal and textual components[167]" (Halliday 1985:xxxiv) and "all systemic contrasts in the grammar derive from one or the other of these functions" (Halliday 1970:326). However, the descriptive categories for the individual components are likely to be language particular.

Much work has been done to specify the peculiarities of a systemic-functional approach[168] to language (see Caffarel et al. 2004 for a collection of studies on different languages and Matthiessen 2009, Caffarel 2010 for an overview on SFL). Although numerous theoretical assumptions underlying these functional grammars have been validated by computer-driven text generation (e.g. linear recursion in Bateman 1989; for an overview O'Donell & Bateman 2005), Bateman (2008) stresses that the second part of language use, namely text analysis, does not show satisfactory results when performed by computers. Thus he concludes that as long as this part of language use cannot be modeled, the systemic functional language approach achieved only a limited proportion of the scope of its goal, namely drawing on corpora for validation. Although this on-going debate about

[167] The distinction between those three metafunctions is based on the idea that besides construing experience in language (ideational function, i.e. 'language as reflection') personal and social relationships are enacted in language (interpersonal function, i.e. 'language in action'). Furthermore, a textual metafunction of grammar is represented in the construction of text by "organizing each element as a piece of information and relating it significantly to what has gone before" (Halliday 1979(2002):199). More importantly "the textual function is internal to language and is instrumental to the other two" (Halliday 1970:326).

[168] For a recent reflection on the development of systemic functional linguistics Matthiessen (2009) is highly recommended.

the applicability and validation of functional grammar needs to be acknowledged, the basic idea of a grammar that attempts to capture the dynamic nature of spoken language is uncontested. Given the specific text genre that will be analyzed in this thesis, i.e. think aloud protocols, this approach is specifically suited for the analysis.

Given that the structure of functional grammar is very complex, this sub-section is intended as a brief introduction on the sentence's verbal complex only[169]. More specifically, it will focus on the ideational aspect of grammar and the different types of experiences that are encoded in verbs. After presenting this specific aspect of functional grammar it will be related to German verb class types. It will be argued that the resulting correspondences are valid enough to apply the classify-cations and terminology proposed by functional grammar in the verb analysis in this thesis.

In functional grammar, a sentence is conceptualized as a figure with different constituents, namely participants, a process, and circumstances. To say it more technically, "a figure is a basic fragment of experience that embodies one quantum of change" (Halliday & Matthiessen 2002:128). Experience is encoded in the process of the figure, which in turn is represented in the verb of the sentence. Initially, Halliday (1985) proposed three principal types of processes, namely material, mental, and relational, and three subsidiary process types, namely behavioral, verbal, and existential. These process types encode different category meanings. These category meanings are referred to as 'domains of experience' in Halliday and Matthiessen (2002). In this later work, four primary domains of experience are distinguished, namely *doing* (material process), *sensing* (mental process), *saying* (verbal process), and *being* (relational process). *Sensing* and *saying* share the quality of symbolic processing, which means that those verbs have the power to set up other figures as semiotic realities. Those semiotic realities are encoded in the figure's projections. The process of projection is instantiated on two levels. The first level is the act of saying or sensing itself, i.e. the experience, whereas the second level expresses the content of saying or sensing (example 11) or saying only (example12).

(133) <u>She said/thought</u> (<u>1st level</u>) that he had left (2nd level) (Halliday & Matthiessen 2002:129). (function: proposition)

(134) <u>She asked him</u> (<u>1st level</u>) to leave (2nd level) (Halliday & Matthiessen 2002:129). (function: proposal)

[169] The verbal complex is made up by the Finite [verb] and the Event. Additionally, it may contain an auxiliary (Halliday 1985:175).

These two examples illustrate that different speech functions can be expressed in a sentence. In propositions, information is exchanged in the form of statements or questions. In proposals, goods and services are exchanged either by a command or an offer (Halliday 1985:87). Although structure and function are similar in both domains, there is an important difference with regard to the direction of symbolic processing. *Sensing* projects the subject's interior content, e.g. ideas. These ideas are projected into existence either through cognition or by a wish or need (called desideration). Two types of *sensing* are distinguished based on the product and the activating energy. Cognitive and desiderative processing create ideas. Emotive and perceptive processing, in contrast, are activated by facts. The first type of *sensing* is interpreted as the most central class by Halliday and Matthiessen (2002:144). *Saying* projects exterior content. The differences that are implied by this distinction are summarized in Table 7.3.

parameters of difference	sensing	saying
direction of symbolic processing	interior	exterior
characteristics of the symbolizer[170] on the semantic level	participant engaged in conscious processing thus it is limited to human consciousness	conscious speaker who is not limited to human consciousness allows 'semiotic things' (e.g. documents, institutions) to take this role
characteristics of the receiver on the semantic level	non existent	the figure's addressee
characteristics of the receiver on the grammatical level	receiver is not a grammatical constituent of the mental clause (she wanted → him to leave)	receiver is a grammatical constituent of the verbal clause (she told him → to leave)
directionality	the senser's involvement ranges over the phenomenon the phenomenon has an impact on the senser's consciousness	from sayer to receiver

Table 7.3: Summary of parameters of difference between *saying* and *sensing* (based on Halliday & Matthiessen 2002:129-131).

The remaining domains of *doing* and *being* do not have the characteristic of projection. Both domains contribute the notion of change to figure but they do so

[170] The symbolizer can be understood as the syntactic subject.

with complementary perspectives. Whereas *doing* initiates change in the "thusness of a participant" (Halliday & Matthiessen 2002:132), change in *being* "appears as an achieved or attained result" (Halliday & Matthiessen 2002:132); compare examples 13 and 14:

(135) He washed it. (domain: doing)
(136) It was clean. (domain: being)

Halliday and Matthiessen (2002) include happening in the domain of *doing* because material processes can either refer to an action (doing) or to an event (happening). Furthermore, having is included in the domain of *being*. In order to understand the reason for this classification, the category of relational processes needs a more detailed description. Relational processes have two different modes: attributive ('a is an attribute of x'; example 15) or identifying ('a is the identity of x'; example 16).

(137) Sarah is wise. (Halliday 1985:113) (mode: attributive)
(138) Tom is the leader./The leader is Tom. (Halliday 1985:113) (mode: identifying)

As *being* construes relations between participants, there are three basic subtypes in this domain, namely intensive ('x is a'), circumstantial ('x is at a'), and possessive ('x has a') (Halliday 1985:113). Halliday and Matthiessen (2002) do not give reasons for the combination of *being* and *having* into one domain. However, based on the definition above, it can be argued that having is one realization of the attributive mode, specifically expressing the subtype of possession. The domain of *doing* also contains subtypes, i.e. figures of doing can be transitive or intransitive (Halliday & Matthiessen 2002:148). The verb analysis presented in this thesis will focus on the general distinction between process types; therefore, no further detail is provided on this detailed level. The differences between the two domains on the broadest level are summarized in Table 7.4.

Parameters of difference	doing and happening	being and having
change over time	involves a change over time which may either be	no change over time, i.e. the actualization will be the same at any point in time
energy	input is required either by the participant or an agent outside the figure	no input is required

Table 7.4: Summary on parameters of difference between doing and being (based on Halliday & Matthiessen 2002:132).

Each figure type has a specific realization in functional grammar. The domain of *doing and happening* is realized in material clauses, the domain of *sensing* is realized in mental clauses, the domain of *saying* is realized in verbal clauses, and the domain of *being and having* is realized in relational clauses. These clauses are characterized differently with regard to directionality of the process, nature of participants, unfolding in time, participation, and projection (see Table 7.5 for a summary).

grammatical reactance[171]	mental clause	material clause	relational clause	verbal clause
directionality of process	bidirectional, i.e. on the senser and by the senser[172]	one directional	one directional	not stated[173]
nature of participants	conscious senser; unrestricted phenomenon	unrestricted subject; participants may be phenolmenal or macro-phenomenal[174]	unrestricted subject and unrestricted phenomenon	not stated
unfolding in time	unmarked tense[175]: present tense	unmarked tense: present progressive	unmarked tense: present tense	unmarked tense: present tense
participation	no special pro-verb	have a special pro-verb 'do'	no special pro-verb	not stated
projection	projects ideas and locutions[176]	no projection	no projection	projects ideas and locutions

Table 7.5: Summary of the characteristics of the different types of clauses (based on Halliday & Matthiessen 2002:134-135).

171 The concept of reactance is adapted from Whorf (1956). It means that words become members of a specific category only if they are used and referred to in sentences that belong to this specific class.(Whorf (1956) in Halliday & Matthiessen 2002:27). In this table the specific characteristics of the reactance of different clause types are summarized.

172 "it is a general feature of mental processes that they can be realized in either direction – either the senser, or the phenomenon that is being sensed, can be the Subject, still keeping the clause in the active voice" (Halliday 1985:110).

173 Fewer information is provided on verbal clauses than on mental, material, and relational clauses in Halliday and Matthiessen (2002). This might result from the original categorization (in Halliday 1985) of verbal clauses to be a subsidiary type whereas the others are considered to be principal types.

174 If participants in the material or relational clause are acts then this participant is called marco-phenomenal (example "She remembered <u>him coming down the stairs</u>." (Halliday & Matthiessen 2002:135)). Whereas if the participant in a relational clause refers to a fact it is called meta-phenomenal (example "She remembered <u>that they had been happy in the old house</u>." (Halliday & Matthiessen 2002:135)).

175 The notion of unmarked tense refers to the tense which is conventionally used in the respective verbal clause. This conventional usage results in all other tenses being salient because they are unexpected (for clarification also Halliday 1985:110, table 5(2)).

176 Locution is defined as a construction of wording, a specific phrase (Halliday & Matthiessen 2004:378).

7.1.2.2 Verb types in German grammar books

Engel (2002:77) states that the only thing that can be said about verbs is their function of defining an action within the event described in the sentence. This simple definition leaves a number of aspects underspecified, e.g. the different nature of processes that are described by verbs and the influence of tense marking and mood on the meaning encoded by the verb. Generally, verbs can be classified in three ways, either by morphological, syntactic, or semantic means (Hentschel & Weydt 2003; Helbig & Buscha 2005). Morphological verb classification is based on the observation that the surface appearance of verbs changes differently to express different grammatical meanings, e.g. taking additional words to form the future tense. In general, Hentschel and Weydt (2003:46) conclude that tense building is most important in distinguishing verbs morphologically. Syntactic verb classification is based on the observation that verbs influence the case marking of related words within the sentence. Semantic verb classification focuses on meanings encoded in verbs. Semantic classifications help to systematically organize a potentially indefinite number of semantically different verbs by some characteristics. Furthermore, it may also highlight shared syntactic structures that can be described accordingly (Bredel & Töpler 2007:874). The analysis in this thesis is based on the semantic classification of verb types.

Traditionally, three verb types are distinguished in German, namely performance verbs (Handlungsverben), verbs of state (Zustandsverben)[177], and process verbs (Vorgangsverben) (see Figure 7.4 for a schematic overview). According to Helbig and Buscha (2005:5), this traditional distinction is based on two features:

(139) the verb's inherent semantic characteristic of [+/- static], and

(140) the subject's semantic character of [+/- agency] which may be added to the verbs inherent semantic characteristic.

Performance verbs are dynamic and they co-occur with a strong subjective agent, e.g. 'walk', 'play', and 'fight'. The encoded actions are often goal directed. Frequently, the goal is bound to the verb in object position of the sentence. From a functional linguistic point of view, performance verbs are considered to be the prototypical verb type (Hentschel & Weydt 2003:36)[178].

---

[177] The English translation is based on Palmer (1965 (1974)) who defines *verbs of state* as "verbs which refer not to an activity but to a state of condition" (Palmer 1965(1974):73). This supports the assumption that this class is the English equivalent to the class of *Zustandsverben* by definition.

[178] This view is supported by Eichinger (1989:23) who introduces *performance verbs*, and more specifically *verbs of movement*, as the prototypical verb type.

Process verbs are also dynamic but they lack the subject's agency because they describe an action that is performed on the subject. Hence they are not goal directed and do not imply an object position. This description holds for verbs such as 'fall', 'grow', and 'die'. Verbs of state are static, i.e. not calling for subject agency because they simply describe a change in state. 'Stand', 'lay', and 'be' describe lasting events that are observable in the situation of speaking[179]. The distinction between performance and process verbs is sometimes difficult to make, for example with 'live'[180]. In these cases, the context needs to be considered to draw valid conclusions on the nature of the verb type (Hentschel & Weydt 2003:38).

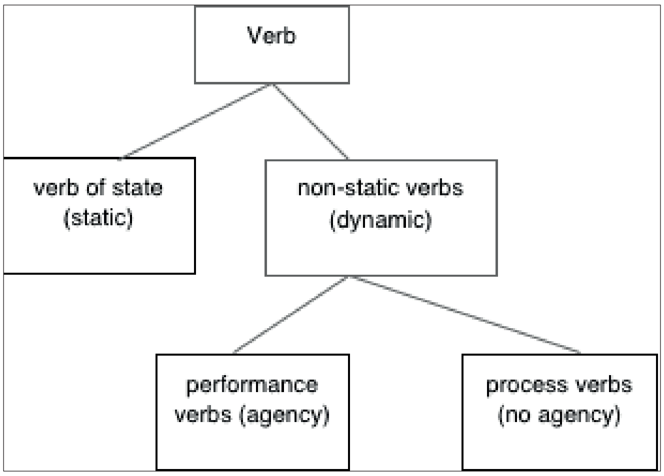


Figure 7.4: Verb types in German (based on Helbig & Buscha 2005:59).

Obviously, these three verb types do not cover the whole range of verbs that are available in natural speech. To account for this observation, Bredel and Töpler (2007:876) mention a verb class called *verba sentiendi* (also referred to as propositional verbs by Bäuerle 1991[181]), which include verbs such as 'feel', 'think', and 'see'. Helbig and Buscha (2005:61)[182] also note the existence of verbs that are

[179] Certainly the communicative situation can also be of the kind that the utterance is written down or only thought of. In favour of text fluency the mode of speaking has been chosen here because interactive communication is assumed to be the basic mode of speaking (e.g. Tomasello 1996).

[180] Hentschel and Weydt (2003:37) use this example to show that some verbs are difficult to categorize. Helbig and Buscha (2005:59), in contrast, list 'live' in the category of state verbs.

[181] This term is used by Bäuerle (1991:709) to describe a class of verbs that provide an attitude on the propositional statement made in the complementary sentence coming with it. For example, in the sentence "Peter assumed that it would be raining tonight." The verb *assume* expresses Peter's personal certainty about tonight's weather.

[182] Interestingly there is no mention about verbs of perception or knowledge in Hentschel and Weydt (2003) or Engel (2002). Unfortunately there is no space here for a thorough discussion on this observation. But it is puzzling that these grammar books leave a large number of verbs unspecified. Nevertheless, there are some publications that focus specifically on a selected group of these verbs (e.g. Latzel 1978).

not covered by the traditional three-type classification and propose to subsume them in one additional category. They propose that this group includes verbs of perception (e.g. 'see', 'hear', 'feel'), verbs of knowledge (e.g. 'think, 'know', 'miss', 'understand'), and verbs of common relations (e.g. 'love', 'hate'). In some respects these verbs behave like performance verbs (e.g. in passive constructions) but in their meaning they resemble state verbs because their subject is a demi-agent. Demi-agents are defined as possessors of perception, knowledge, or relations who are passive with regard to the action described by the verb. Nevertheless, in some cases the subject still behaves as an agent syntactically (Helbig & Buscha 2005:61).

Some linguists refer to a verb type called cognitive verbs (e.g. Bühler 1989) but they neither define the category nor provide references in which this definition can be found. Now it could be hypothesized that there is a class of cognitive verbs in English and the category label has been transferred into German linguistic research without being discussed and specified in more depth. However, the search for a class of cognitive verbs in English grammar books did not yield the expected result. A more detailed search highlighted that verbs that are categorized as private verbs[183] by Palmer (1965 (1974):71) resemble English translations of verbs that are listed as cognitive verbs in Butulussi (1991). The class definitions provided by Palmer (1965 (1974):71) and Butulussi (1991) also match.

Butulussi's (1991) definition is more fine grained by stating that cognitive verbs describe inner, mental states and processes, such as possession, acquisition, loss, or regain of knowledge. Furthermore, she argues that cognitive verbs express assumptions about events.[184] She proposes that German cognitive verbs can be classified with respect to the level of certainty that they express by either expressing the possession of knowledge ('wissen'[185] and 'kennen' (know)) or expressing assumptions ('glauben' (believe), 'vermuten' (suppose), 'ahnen' (guess),

[183] In English the majority of verbs include a temporal aspect. But some verbs do not have a temporal marker. Those are so called *non-progressive verbs*. Palmer (1965(1974)) further divides this class into *private verbs* and *verbs of sensation*. Private verbs "refer to states or activities that the speaker alone is aware of" (Palmer 1965(1974):71). Among them are 'think', 'imagine', 'hope', 'plan', 'forget', and 'believe'. Examples for the category *verbs of sensation* are 'see', 'smell', 'hear', 'taste', and 'feel'. Furthermore, Palmer (1965(1974):73) also lists *verbs of state* as belonging to the class of non-progressive verbs. The definition he gives for this category is equivalent to the definition commonly used to describe the class of *Zustandsverben* in German.

[184] However, it needs to be noted that Butulussi (1991) does not refer to any further publication to verify her classification or the type 'cognitive verb' by itself. She presents the classification as if it was a natural fact that this class exists and that it is referred to as 'cognitive verbs'.

[185] The cited verbs are also in German because the analysis will be performed on German data, hence a list of German cognitive verbs seems reasonable at this point.

'annehmen' (assume), 'zweifeln' (doubt)). Additionally, cognitive verbs can be classified with respect to the nature of the cognitive process. Whereas 'erfahren' (experience), 'kennenlernen' (get to know), and 'erkennen' (realize) signal the acquisition of knowledge, 'vergessen' (forget) expresses the loss of knowledge (Butulussi 1991:3). Based on this more detailed classification of cognitive verbs, it can be argued that verbs expressing processes that are counted among higher-level cognitive functions[186] belong to the class of cognitive verbs. Furthermore, the class of cognitive verbs would include members of the class of verbs of knowledge that is proposed by Helbig and Buscha (2005).

This short review illustrates that a fundamental part of verbs is not classified in traditional accounts of German grammar. In order to compensate for this lack Steiner and Teich (2004), for example, adopted Halliday's category of *mental* verbs for German. In this thesis, the terminology proposed by Helbig and Buscha (2005), that seems to be the most thoroughly defined one for verbs of sensing in German, will be adopted and combined with the concept of *mental* verbs by Halliday.

7.1.2.3 Correspondence between verb types in German and in English from a functional grammar perspective

Before presenting empirical research on verb analysis in think aloud protocols, the domains of experience presented by Halliday and Matthiessen (2002) will be related to traditional German verb classification in order to motivate the approach that will be pursued in this thesis. Combining the characteristics and the mentioned members of verb types listed in traditional German grammar books (see Table 7.6 for an overview), the following correspondences are proposed: the domain of *sensing* expressed in mental clauses subsumes the verb types of verbs of perception, verbs of knowledge, and verbs of relations. Cognitive verbs are already represented in Helbig and Buscha's (2005) category of verbs of knowledge. The domain of *doing and happening*, which is expressed in material clauses, subsumes the verb types of performance verbs and process verbs. Verbs of state correspond to verbs in the domain of *being and having* that are expressed in relational clauses. No correspondence for the domain of *saying* could be found in the German grammar books that have been consulted.

[186] Critical thinking, reasoning, and decision-making are generally referred to as higher order mental process in cognitive science (e.g. http://general-psychology.weebly.com/what-are-the-higher-order-of-mental-processes.html; 08.11.2012).

verb type	characteristics	example of members	corresponding domain in functional grammar
performance verbs (Hentschel & Weydt 2003)	dynamic; strong subjective agency	walk, play, fight	Doing & happening
process verbs (Hentschel & Weydt 2003)	dynamic; no subjective agency	fall, grow, die	Doing & happening
verbs of state (Hentschel & Weydt 2003)	static; no subjective agency	stand, lay, be	Being & having
verbs of perception (Helbig & Buscha 2005)	subject is not an agent but a demi-agent	see, hear, feel	Sensing
verbs of knowledge (Helbig & Buscha 2005)	subject is not an agent but a demi-agent[187]	think, know, miss, understand	Sensing (specifically mental clause of cognition[188])
verbs of common relations (Helbig & Buscha 2005)	subject is not an agent but a demi-agent	love, hate	Sensing (specifically mental clause of emotion[189])
cognitive verbs (Butulussi1991)		know, think, assume	Sensing

Table 7.6: Summary of verb types proposed in traditional German grammar books and the corresponding domain in functional grammar.

Steiner and Teich (2004) outline parts of a functional grammar of German, namely the aspects of mood, transitivity, and theme. Furthermore, Petersen's[190] unpublished manuscript of a functional systemic grammar approach for the German language strengthens the proposed correspondences. In this manuscript Petersen convincingly applies the Sydney functional grammar framework to the German language. The description of the experiential function of language and the different process types in particular support the verb correspondences that have been proposed in Table 7.6. Petersen (unpublished) describes the representation of and distinction between material, relational, verbal, and mental processes in German. Thus the lack of representation of verbal processes in traditional German grammars, as has been pointed out in the previous section, can be filled by

[187] A demi-agent is defined to be the "Wahrnehmungsträger, Erkenntnisträger, Verhältnisträger" (Helbig & Buscha 2005:61). All of these roles are subsumed in the concept of *senser* as proposed by Halliday (1985).

[188] Halliday & Matthiessen (2004:198).

[189] Halliday & Matthiessen (2004:198).

[190] I want to kindly thank Mr. Petersen for providing me with the relevant parts of his manuscript and his interest in our work.

Petersen's (unpublished) category of verbal processes, for which the verb 'sagen' (say) is the prototypical representation. Furthermore, Petersen identifies the four different mental processes (namely cognitive, desiderative, emotive, and perceptive) that have been described for English (Halliday & Matthiessen 2002) for German as well. He further notes that the cognitive type is the most frequent subtype of mental processes whereas desidervative mental processes, such as 'wünschen' (wish), are very rare.

7.1.3 Analysis of verbs and other linguistic markers in verbal reports in problem solving research

A detailed review of studies that investigate the linguistic structure of think aloud protocols recorded during problem solving activities was provided in chapter 3. Therefore, this section highlights the analysis and findings of the structure of the verb phrase in reference to the more detailed accounts provided already.

Being interested in the differences between different problem solving performances on the level of mental processes, Roth (1985; 1987) was among the first researchers who analyzed the grammatical features of think aloud protocols. Drawing on a number of previous studies by other scholars, he analyzed verbal phrase structure with regard to modality, i.e. verb mood and modal verbs. Based on previous findings, Roth (1985) assumed that protocols by successful problem solvers exhibit a different linguistic structure than those reported by unsuccessful problem solvers. These expectations were confirmed as the analysis revealed that unsuccessful problem solvers used modal verbs in the subjunctive mood at a statistically significant higher frequency than successful participants. No difference was observed with regard to modal verbs in the indicative mood (Roth 1985:186). Roth (1985) argued that these findings illustrate that unsuccessful problem solvers feel the need to fill in missing information very quickly as expressed in hypotheses that are marked by modal verbs and the subjunctive mood. He further argued that the lack of subjunctive mood in the protocols by successful participants indicated that those focused on the current task situation. Moreover, he proposed that the indicative mood highlighted their tendency to incorporate new information, which was created in the interaction with the system. These findings were supported by results obtained in a second study by Roth (1987).

Taking a similar approach, Bartl and Dörner (1998) studied participants' problem solving strategies while solving a dynamic system task. Their analysis of linguistic features, i.e. conjunctions, verb tense, and mood, of two recorded think aloud protocols revealed systematic differences between one successful and one unsuccessful participant. The analysis revealed significant differences in the kind of conjunctions and their overall frequency (for more detail see chapter 3). Moreover, Bartl and Dörner (1998) observed differences in verb tense. The participant who did not succeed, used the subjunctive mood only once. This participant used the present tense exclusively. The participant who succeeded, however, used the subjunctive mood nine times as well as past tense ten times and the verb phrase 'I think' (Bartl & Dörner 1998:236). In their discussion they argued that these results highlight that the successful participant speculated about possible moves and incorporated information from the past into the current problem representation. Bartl and Dörner (1998) propose that if these findings are combined with the findings on use of conjunctions, they suggest that participants who are flexible in their thinking, i.e. incorporating new and past information, finding reasons for relations between information, and hypothesizing about possible moves, were more likely to succeed than participants who focus on the current problem state only. This observed difference in thinking partially supports findings previously reported by Roth (1985; 1987). Interestingly, Roth (1985) reported that unsuccessful participants used subjunctive mood more frequently than successful participants; which was the reversed trend to that reported by Bartl and Dörner (1998). As both studies investigated performance on a dynamic systems task and analyzed think aloud protocols, these contradictory findings might be an effect of the case study that was reported by Bartl and Dörner (1998). This highlights the necessity to support findings and effects by analyzing multiple data sets.

Whereas Roth (1985; 1987) and Bartl and Dörner (1998) investigated think aloud protocols on a dynamic system task, another research group focused on the linguistic analysis of think aloud protocols on a logic-based task. Caron, Caron-Pargue, and their collaborators conducted experiments on the tower of Hanoi and the problem of the Chinese rings. Through in-depth linguistic analyses of the recorded think aloud protocols combined with descriptions of the current configuration of pags and rings, they identified linguistic and pragmatic markers in think aloud protocols. They linked those to the organization of the speaker's situational mental model, the function of units in the organization, and changes in the organization. With regard to action, as expressed in the verbal phrase of a sentence, Caron and Caron-Pargue (1987) note that

> "a given action (...) may consist in the achievement of an intended goal, or an auxiliary step to that goal; it may also be the result of an actual computation or the execution of an already learnt pattern." (Caron & Caron-Pargue 1987:173)

They argue that this difference is expressed in the modality of the verbal phrase. Caron and Caron-Pargue (1987) observed that participants pointed to a goal by stating "I have to" or "I want to" (Caron & Caron-Pargue 1987:175). These phrases were later called 'modalities of action', also including "I am able to" (Caron-Pargue & Caron 1991:32). In 1996, Caron disentangled interjections and modal expressions further by stating that "access to LTM [Long Term Memory] is mainly marked by interjections" and "modal expressions (...) can be interpreted as marking, in these protocols, the computation of new goals" (Caron 1996:26). Along these lines, Bégoin-Augereau and Caron-Pargue (2009) studied the linguistic difference between internal, external, and intermediate spaces involved in problem solving. In their final discussion, they argued that their previous work showed that "modal markers suppose a disentanglement from the situation, and imply a gap, difficulties, attempts, retrieval of information for an intended reorganization of the representation" (Bégoin-Augereau & Caron-Pargue 2009:121). They further argued that modal markers highlight "non automatic but critical and intentional goals" whereas connectives mark automatic goals (Bégoin-Augereau & Caron-Pargue 2009:121).

Based on these previous findings, Bégoin-Augereau and Caron-Pargue (2010) argued that modal terms always involve a detachment from the situations. Those are either expressed in goals or in evaluations. Referring to previous work, Bégoin-Augereau and Caron-Pargue (2010) distinguished between strategic goals and automatic goals. They argue that automatic goals do not involve any restructuring of the representation because it is already known to the problem solver; these goals do not contain modal terms. Strategic goals, in contrast, represent restructuring processes that result in planning. Their analysis revealed that strategic goals were marked by modal terms, such as 'could' or 'should' (Bégoin-Augereau & Caron-Pargue 2010:263). If modal terms were identified in evaluative statements, it was at times in which new information was verbalized. Bégoin-Augereau and Caron-Pargue (2010) argued that this finding suggests that the representation is re-organized by a strategic retrieval of information (Bégoin-Augereau & Caron-Pargue 2010:263).

Tenbrink (2008) applies the methodology of Cognitive Discourse Analysis[191] (hereafter CODA) for analyzing the mental processes of one participant engaged in solving the traveling salesman problem. She analyzed three different discourse tasks, namely think aloud protocols, retrospective reports, and instructions to a

[191] The methodology that is applied in this thesis as well is described in detail in chapter 3. Additionally, an overview is to be found in Tenbrink (2010).

friend. The comparison of the verb phrase in these discourse tasks revealed that action verbs in the present tense were frequent in think aloud protocols. Tenbrink (2008:131) argued that this highlighted the participant's focus on the current situation. Building on these findings and further earlier work in spatial problem solving, Tenbrink and Seifert (2011) provide an elaborate overview on the linguistic analysis of written tour plans for holiday trips around Crete (a real island) or Cretopia (a fictional island). They studied the distribution and nature of nouns, verbs, adjectives/adverbs, and temporal markers in written route descriptions. In their analysis the authors were specifically interested in the linguistic representation of the two domains that were involved, i.e. the domain of planning, that was visually present on the map, and the domain of travelling that was a conceptual representation of the real world. They categorized each annotated item as indicators for the conceptual domain of planning (e.g. 'map' (noun), 'helpful' (adjective), and 'plot' (verb)) or travelling activity (e.g. 'trip' (noun), 'boring' (adjective), and '2 days' (temporal markers)) (Tenbrink & Seifert 2011:115). Cases in which this binary categorization was not possible were also indicated. The analysis revealed that modal verbs in subjunctive mood ('could' and 'should') and the discourse particle (e.g. 'in order to') were indicators of the connection between the two domains (Tenbrink & Seifert 2011:116). Moreover, besides encoding domain specificity, linguistic markers, such as nouns, quantifiers, and preposition, were also shown to distinguish different spatial strategies (for an overview see Tenbrink & Seifert 2011:116).

Investigating verbal reports on analogical problem solving, Gralla et al. (2012) also reported that different strategies were represented by a distinct linguistic repertoire with regard to verbs, nouns, temporal order, and markers of generalization ('again')[192] and similarity ('same'). One strategy was marked by the frequent use of the verbs 'connect' and 'draw', referring to the activity of connecting symbols. Another strategy was marked by the verb 'copy' as this strategy involved copying the solution of a previous problem to solve the recent problem.

In their seminal work, Ericsson and Simon (1993) focus on the indicative mood and modal verbs and combine verb tense and other syntactical constructions to distinguish between current-state information, future-state information, and hypothetical information. They argue that current-state information is often expressed

[192] The terminology *marker of generalization* is based on the finding presented in Tenbrink (2010:131) that 'again' marks the process of gradual generalization.

by the present tense and verbs such as ‘be’ and ‘have’[193]. They propose that information that reflects “intentions or goals to attain configurations or attributes not true for the current construction” (Ericsson & Simon 1993:223) is often marked by verbs in the future tense. Additionally, constructions expressing wishes, such as ‘want to have’ or ‘like to have’, are assumed to be common markers of information on the future. Moreover, information can also be marked as hypothetical or conjectural by selecting modal verbs, using modal adverbs (such as ‘maybe’ and ‘perhaps’), verbs marking the mental state of the speaker, e.g. ‘guess’, ‘assume’, and if-constructions.

In sum, the reported results suggest that verb phrases in think aloud protocols contain information on the current situation expressed by present tense marking. Moreover, verbs of *doing and happening* were frequently observed, indicating that participants verbalize their actions rather than plans. If hypothetical states, such as new goals, were verbalized, these were marked by modal verbs. The studies highlighted that verbs signal the connection between different conceptual domains, such as mental and physical, and different ways of approaching problem solving, i.e. closed-systems vs. open-systems into which new information is continuously integrated. However, the reported findings (see Table 7.7 for a summary) were based on verb phrase analyses based on their semantic meaning but not concerning the more general categories of verb typology. Therefore, the analysis presented in this chapter extends these findings in two ways. First, verbs are analyzed with regard to their general type rather than their specific meaning. It is proposed that this approach allows for more general conclusions because fewer categories are compared, i.e. five kinds of verb type. Second, verbs are analyzed with regard to their function as indicators of different problem solving processes. This is achieved by a systematic analysis of verb type patterns with regard to the previously identified processes (see chapter 6). The analysis of verb type is combined with the analysis of verb tense and modality to investigate if the expression of problem solving processes differs in these respects.

[193] As has been shown in the section on verb classes, both of these verbs belong to the class of *being and having* (Halliday & Matthiessen 1999).

Grammatical category	members that are mentioned	indicator of	references
modal verb (subjunctive mood)	wollte (would), könnte (should)	a) connection between two conceptual domains; b) express mental simulation, e.g. hypotheses, and uncertainty	a) Tenbrink & Seifert (2011); b) Roth (1985); Bartl & Dörner (1998); Ericsson & Simon (1993)
modal verb	müssen (must)	dogmatic thinking, i.e. a closed system, and connects actions to the real world	Roth (1985)
modal verb	können (can)	a) non-dogmatic thinking, i.e. open system and connects actions to the real world; b) in the context of goal formulation this marks a strategic goal in which the representation is re-organized and plans are stated	a) Roth (1985); b) Bégoin-Augereau & Caron-Pargue (2010)
modal expressions	'have to', 'able to', 'want to', 'like to'	a) computation of new goals; b) information about intentions and goals that are not yet attained	a) Caron-Pargue & Caron (1991); Caron (1996); b) Ericsson & Simon (1993)
modal adverbs	maybe, perhaps	information that is hypothetical or conjectural	Ericsson & Simon (1993)
epistemic modals[194]	'I think', 'of course'	in an evaluative utterance they express the recognition that a pattern has already been stored (modal in the present or future tense) or the representation is reorganized (modal in the past tense)	Caron-Pargue & Caron (1991)
verb type: sensing	'think', 'assume'	information that is hypothetical or conjectural	Ericsson & Simon (1993); Bartl & Dörner (1998)
verb type:	'I believe'	in the context of evaluation	Bégoin-Augereau

[194] This term is introduced by Caron-Pargue and Caron (1991) with the two examples that are quoted. As these examples belong to different grammatical classes, i.e. verb and adverb, they cannot be grouped with any of the other categories thus the term is quoted here.

sensing		these verbs signal the detachment from the situation	& Caron-Pargue (2010)
verb tense	present tense	focus on the current situation; current state information	Tenbrink (2008); Ericsson & Simon (1993)
verb tense	future tense	future-state information about intentions and goals that are not yet attained	Ericsson & Simon (1993)
verb tense	past tense	information from the past that is reconsidered and incorporated into present considerations	Bartl & Dörner (1998)

Table 7.7: Summary of reported findings on verb analysis in verbal reports.

7.2 Research Questions and expectations

The presented literature overview illustrated that previous research has focused on modal verbs and semantic content of verbs in order to describe the difference in problem solving performance or the reported problem solving behavior itself. In contrast to those studies, the analysis in this thesis focuses on the different problem solving processes and their linguistic representation. Thus the problem solving process, i.e. hypothesis, evaluation, action etc. will be the unit of analysis. Furthermore, the verbs will be annotated on a more general level based on the verb classification proposed by Halliday (1985) and Halliday and Matthiessen (2002). This approach should be suitable to find answers to the following basic research question:

- Research question 1: Is there a systematic difference between problem solving processes with regard to verb type?

As speakers mark certainty about statements by modality, this aspect of the verb phrase is addressed in research question 2:

- Research question 2: Is there a difference between the investigated processes with regard to mood (subjunctive and indicative) and modal adverbs?

Based on the reviewed findings (see Table 7.7) and preliminary analyses (Tenbrink & Gralla 2009; Gralla & Tenbrink 2012), the following expectations are raised. Only those processes which are directly related to the problem solving process will be analyzed, i.e. those processes which are introduced as local processes of problem solving in chapter 6.

It was expected that hypotheses are marked by modal verbs as indicators for the connection between the domains of wooden objects (physically present in the real world) and functional objects as parts of a house or building (conceptual domain of the target object) (Tenbrink & Gralla 2009; Tenbrink & Seifert 2011). Furthermore, modal verbs were reported to mark hypothetical, conjectural and therefore tentative assumptions about information and relations between information (Ericsson & Simon 1993). The tentative and hypothetical character of hypotheses are further expected to be expressed by modal adverbs (e.g. 'maybe', 'perhaps') (Ericsson & Simon 1993) and modal verbs in the subjunctive mood which indicate mental simulation (Roth 1985). Future actions, intentions, and goals, which are typically formulated in hypotheses, are indicated by future tense marking (Ericsson & Simon 1993) and in verbs such as 'have to', 'like to', and 'want to' (Caron-Pargue & Caron 1991; Ericsson & Simon 1993; Caron 1996). Preliminary findings suggest that verbs of *being and having* are frequently used in hypotheses when participants suggest the function of objects (Tenbrink & Gralla 2009) or when they assume which actions are suitable. Verbs of *sensing* (e.g. 'guess') express that the speaker is not certain about the uttered information and refers to himself as the reference point for the proposal which is formulated in the hypothesis (Gralla & Tenbrink 2012). In cases in which different domains are connected, it was expected that verbs of comparison ('use as', 'look like') were used along with other options of mapping between domains (Tenbrink & Gralla 2009:10 and results in chapter 8).

Actions were expected to be represented by action verbs (Tenbrink 2008) and the present tense which signals current state information (Ericsson & Simon 1993; Tenbrink 2008). Tenbrink (2008) does not define her category 'action verbs' but she provides the example 'ich gehe' (go) (Tenbrink 2008:131) and the exemplary think aloud protocol contains the verb 'gehen' (go) in most information units and 'laufen' (run) in one of them. Therefore, it can be assumed that Tenbrink's (2008) class of 'action verbs' corresponds to verbs of *doing and happening* as previously introduced.

Considering the classification of process categories[195], *descriptions of mental state* contain processes in which speakers refer to themselves as the source of information and expresses an opinion regarding the content of what is being said. Those processes are expected to contain epistemic modals, such as 'I think',

[195] Definitions of the respective categories and examples are to be found in chapter 6.

because those have been reported to signal a change in procedure and storage in long-term memory (Caron-Pargue & Caron 1991). The verb 'think' belongs to the class of verbs of *sensing*, more specifically, to the sub-class of cognitive verbs of *sensing*. The use of these verbs in combination with the first person pronoun signals that the speaker focuses on himself/herself for information, i.e. the information source is internal. The source of information may be background knowledge, perceptual information, or memory traces. Additionally, the speaker expresses his/her opinion on what is being said because these verbs signal that the presented conclusions are tentative with regard to their relation to the real world.

Evaluative processes, positive as well as negative, were expected to be marked by the present tense. Given that evaluations refer to the current state, a high frequency of verbs of *being and having* could be expected. The same expectations were raised concerning the category of *features of object parts* since those processes are essentially reports about perceptual information on objects.

7.3 Data Analysis

The first part of the analysis focused on the description of the linguistic representation of problem solving processes. Thus the basic unit of analysis in this part of the chapter were the coded process. A process unit (abbreviated PU) may contain more than one discourse unit (abbreviated DU), as shown in example 17 that represents one hypothesis[196].

(141) (DU 1) okay so <u>sieht</u> das schon aus (DU 2) als <u>würde</u> die hier oben <u>sein</u> (DU 3) deswegen <u>wird</u> das wohl hinten oben unterm Dach <u>sein</u> (okay this already looks / as if this is up here / therefore this will be above in the back underneath the roof)

In each PU all verbs were annotated with regard to form, i.e. tense, mood, and verb category. Tense might either be present, past, or future. Mood was coded as indicative, subjunctive, or imperative. Five verb types were distinguished, namely *being and having*, *doing and happening*, *sensing*, *saying*, and modal. The distinction between the first four categories was based on Halliday and Matthiessen (2002). Furthermore, instances of ellipsis were coded. The application of this scheme results in the following annotation (see Figure 7.5):

[196] The concept of 'process units' is introduced in chapter 3.

A	B	C	D	E	F
TI	utterance	verb lemm	verb tense	verb category	Ausdru
8	und dann sind hier noch mal solche Schrägen	sein	Präsens	existenzielle Prozesse (des Seins)	
8	<- die werden auch wohl erst / erst beim ersten Geschoss zu kommen	kommen	Futur	materielle Prozesse (des Tuns & Gesche	

Figure 7.5: Screenshot of part of the annotation table for one *hypothesis* ("existenzielle Prozesse (des Seins)" correspond to *being and having*).

For a better understanding some examples are provided for all annotated verb categories[197]:

- *being and having:* aussehen (look like), Bedeutung haben (have meaning), gehören (belong), halten (hold), passen (fit), sein (be), stehen (stand), ...
- *doing and happening*: anfangen (start), angucken (look at), anordnen (arrange), auspacken (unpack), bauen (build), benutzen (use), einbauen (build), funktionieren (work), gucken (look for[198]), legen (put), machen (do), tuen (do), verkeilen (fasten), verrutschen (shift), versuchen (try), zusammenfallen (crash), ...
- *sensing:* annehmen (assume), denken (think), erinnern (remember), sich fragen (ask oneself), glauben (believe), *gucken* (look around), hoffen (hope), nachdenken (think), rauskriegen (find out), sehen (see), überlegen (think), ...
- *saying*: sagen (say)
- *modal:* dürfen (allowed), können (can), möchten (want)[199], müssen (must), sollen (should), wollen (want)

The analysis categories were defined in an iterative process. For most verbs it was clear which category they belong to. However, in some cases it was more difficult. Among the more difficult to define ones were 'gucken' (look), 'passen' (fit), and 'funktionieren' (work). Those three cases will be discussed in more detail in turn.

'Gucken' (look) was used in an active sense in some cases i.e. look for an object or try something (example 18). This figure has features of a material clause.

(142) Ich **gucke** mal, ob das passt. (I will see if this fits.)

In other cases, 'gucken' was used in the sense of 'looking around' and seeing something without an intention. If 'gucken' was used with the first connotation, it

---

[197] The first four classes are figure types whereas the class of modal verbs is one expression of modality.

[198] More specifically in the sense of 'search'.

[199] As outlined section 1.2 of this chapter, these are the core modal verbs in German.

was classified as *doing and happening* and if it was used with the second connotation, it was classified as *sensing*.

The verb 'passen' (fit) was frequently used (examples 19 and 20 are typical), e.g.

(143) das **passt** nicht **zusammen** (this does not fit) [26]
(144) jetzt dürfte es aber **passen** (now it should fit) [16]

In both examples the speaker evaluates if the object parts fit together. It is not an action of fitting objects together but rather the state that is evaluated or debated. It is evaluated if an arrangement or object has the characteristic of fitting together. Based on its configuration, the figure is classified as a relational clause thus 'passen' has been categorized as a verb of *being and having*.

The verb 'funktionieren' (work) was also difficult to categorize. The following two examples illustrate its typical usage in which 'funktionieren' encodes an evaluative judgment about the action that has been performed. The figure that is created by the verb determines the structure of a material clause more specifically it encodes a qualitative change[200]. Based on this argumentation, 'funktionieren' was classified as a verb of *doing and happening*.

(145) vielleicht funktioniert es, dass es so hält (maybe it works in a way that this holds)
(146) (DU1) wenn ich einfach nur das versuche erstmal einzubauen/ (DU2) nee das funktioniert, glaube ich, nicht (when I first try to fit this in first/ no, this does not work, I think) [29]

By annotating modal verbs and verb mood two ways of expressing modality are captured. In order to capture all possible expressions of modality, all occurrences of modal adverbs need to be annotated as well. The identification and annotation was based on the list presented by Brinkmann (1971:401). All occurrences have been coded for the respective function that has been assigned to them by Brinkmann (1971). As previously outlined, he identified three functions of modal adverbs, namely modification, marking information as facts or assumptions, and highlighting the emotional quality of what is being said. These functions are coded as

- *function 1 (certainty): highlighting that what is being said is certain*
- *function 1 (prerequisite)*: signaling that prerequisites do not hold anymore
- *function 2 (fact):* signaling that a piece of information is a fact

[200] Please recall the categorization in Halliday and Matthiessen (2002:132) in which they specify that existential change involves creation and destruction. 'Funktionieren' comments on this existential change.

- *function 2 (assumption)*: signaling that an information is an assumption
- *function 3:* encoding emotional quality

7.4 Results

The results describing the form of the verb phrase are presented with regard to the distribution between problem solving processes. In the first part of the analysis, results on the verb class of the main verb, tense, and modality are presented. Modality is investigated in its representation by mood of the main verb, in modal verbs, and modal adverbs. Modal verbs may either fill the role of the main verb or fill one part of a modal verb phrase (e.g. 'wie ich das machen könnte' (how I could go about doing this)); both cases will be looked at separately. The second part of the analysis focuses on modal adverbs.

7.4.1 Form of the verb phrase

In the following graphs the distribution of verb type within process categories is described first. Then the distribution of modality is presented. The analysis of the verb phrase will be ended by the presentation of the results on tense in different problem solving processes.

Overall, 2789 verb phrases were identified and coded. Regarding process types, the majority (1022 phrases, 36.6%) of these were *hypotheses*, followed by *evaluations* (482 cases, 17.3%; *positive*: 236 phrases and *negative*: 246 phrases), *actions* (467 phrases, 16.7%), and *descriptions of mental state* (441 phrases, 15.8%). Less frequent were *features of object parts* (241 phrases, 8.64%) and least frequent were *plans for action* (90 phrases, 3.2%), *dead end* (39 phrases, 1.4%), and *fresh start* (7 phrases, 0.3%). This distribution within verb phrases mirrors the findings on process types in general (see results in chapter 6).

7.4.1.1 Verb type

The majority of verbs were coded as verbs of *being and having* (1351 cases, 48.4%). The second most frequent type were verbs of *doing and happening* (950 cases, 34.1%). Verbs of *sensing* were less frequent (405 cases, 14.5%). Verbs of *saying* were notably least frequent (24 cases, 0.8%). Verbs have been skipped in only 59 cases (2.1%).

A chi-square test revealed highly significant differences in the distribution of verb type between process types; $L\chi^2$ (24, N = 2730) = 828.84, p = .00. The general trend is displayed in Table 7.8 (p.270). The graph visualizes that verbs of *being and having* were most frequent in *hypotheses* (53.0%, SD = 16.7)[201], *features of object parts* (64.0%, SD = 34.2), and *evaluations* (negative (57.0%, SD = 37.5) as well as positive (63.4%, SD = 34.2)). These differences were significant at the p < .001 level for *hypotheses* (z = 4.8), *features of object parts* (z = 6.5), and *positive evaluation* (z = 5.1) and at the p < .05 level for *negative evaluations* (z = 2.4)[202].

Verbs of the category *being and having* were more frequent in *hypotheses* (53.0%) than in *actions* (10.8%) or *plans for action* (7.3%). The processes in which verbs of *being and having* were most frequent are related to mental activities (ideas on what things represent and how they are arranged) and descriptions as well as evaluations of what is being perceived.

Verbs belonging to the type of *doing and happening*, on the other hand, were most frequently observed in *actions* (77.6%, SD = 18.0) and *plans for action* (33.3%, SD = 43.6). In both cases, the difference was highly significant at the p < .001 level, *action* (z = 16.6) and *plan for action* (z = 6.4). Furthermore, verbs of *doing and happening* were the only ones that were used in *fresh starts*. As suggested by the names of these process categories, they are related to action, i.e. interaction with and manipulation of the physically present world.

Verbs of the category *sensing*, which project the speaker's mental world into an utterance, were most frequently used in *descriptions of the speaker's mental state* (32.7%, SD = 25.2). The frequency of verbs of *sensing* in this category was significantly different from the frequency in all other processes at the p < .001 level with z = 10.3.

Verbs of *saying* were rarely observed in the data (0.7%, SD = 5.8 overall processes). Those rare cases were observed in *hypotheses* and *description of mental state*.

---

[201] Raw frequencies and percentage are reported in Table 7.12 and Table 7.13 in the appendix.

[202] Standardized residuals for all process categories × verb type are reported in Table 7.14 in the appendix.

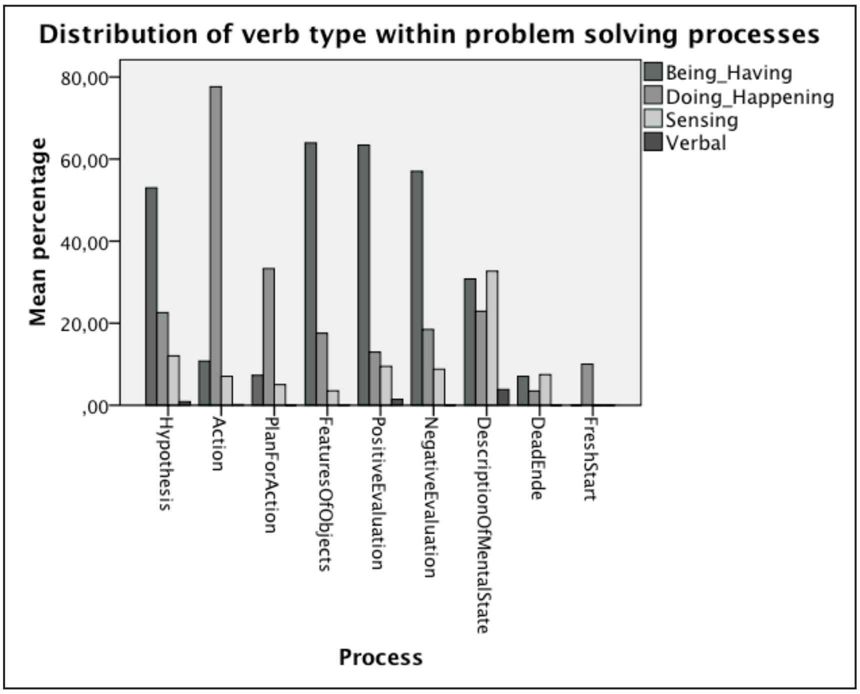


Table 7.8: Distribution of verb type within problem solving processes (by mean percentage).

7.4.1.2 Modality

7.4.1.2.1 Verb mood

The second grammatical feature of the verb phrase that was investigated was the expression of modality. The distribution of indicative and subjunctive mood, the imperative, and modal verbs is displayed in Figure 7.6. The graph clearly shows that the indicative mood was used in the majority of cases. Nevertheless, highly significant differences could be observed between process categories; $L\chi^2$ (16, N = 2718) = 153.01, p = .000. The high frequency of indicative mood in *hypotheses* contributed to this main effect highlighting that indicative mood was significantly more frequent than expected (M = 18.7, SD = 12.6[203], z = -2.0, p < .01[204]).

Furthermore, the distribution of the subjunctive mood contributed to the statistic effect. The statistical analysis revealed that the subjunctive mood was significantly more frequent in *hypotheses* (M = 3.1, SD = 3.5; z = 7.3), *actions* (M = 0.2, SD = 0.4; z = -4.5), *comment on object features* (M = 0.14, SD = 0.41; z = -2.7), *description of mental state* (M = 0.4, SD = 0.9; z = 2.6), and *negative evaluations* (M = 0.2, SD = 0.5; z = 2.5) than in the remaining process categories.

[203] Raw frequencies for all process categories and verb moods are reported in Table 7.15 in the appendix.

[204] Standardized residuals for all verb mood × process categories are reported in Table 7.16 in the appendix.

Imperatives were infrequently used. If they occurred, they were more frequent in *actions* (M = 0.1, SD = 0.4; z = 3.2) and *fresh starts* (M = 0.02, SD = 0.1; z = 5.8) than in all other process categories.

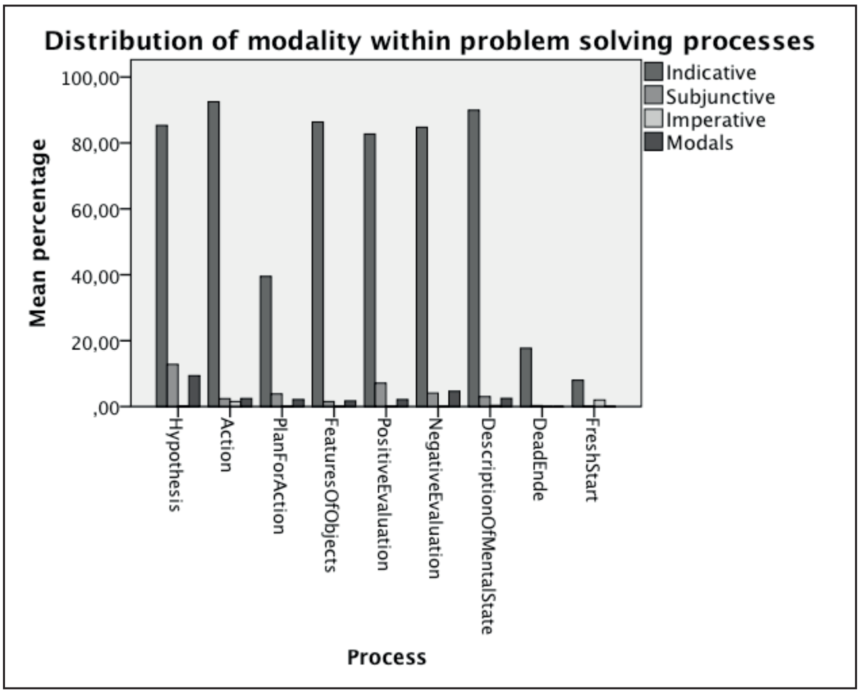


Figure 7.6: Distribution of modality within problem solving processes (by mean percentage).

7.4.1.2.2 Modal verbs

Besides verb mood, modal verbs can be used to take a stance on sentence meaning. Overall, 458 modal verbs have been identified. These were few cases as compared to the overall number of verbs (2.8%, SD = 9.6) (see Table 7.13 in the appendix).

Modal verbs were most frequently observed in *hypotheses* (M = 9.4, SD = 10.7) (see Figure 7.7). Modal verbs can fill two kinds of verb roles in the sentence, i.e. main verb position or as part of the verb phrase. The distribution of these two options illustrated that modal verbs were most often used in verb phrases (see Figure 7.7). Only in *hypotheses*, the two options were observed at a comparable frequency (M = 3.2, SD = 3.5 for verb phrase position and M = 2.2, SD = 3.2 for main verb position[205]). A high frequency of modal verbs that were embedded in a verb phrase was also observed in *descriptions of mental state* (1.10 cases, SD = 1.3).

[205] All results are reported in Table 7.19 and visualized in Figure 7.13, Figure 7.14 in the appendix.

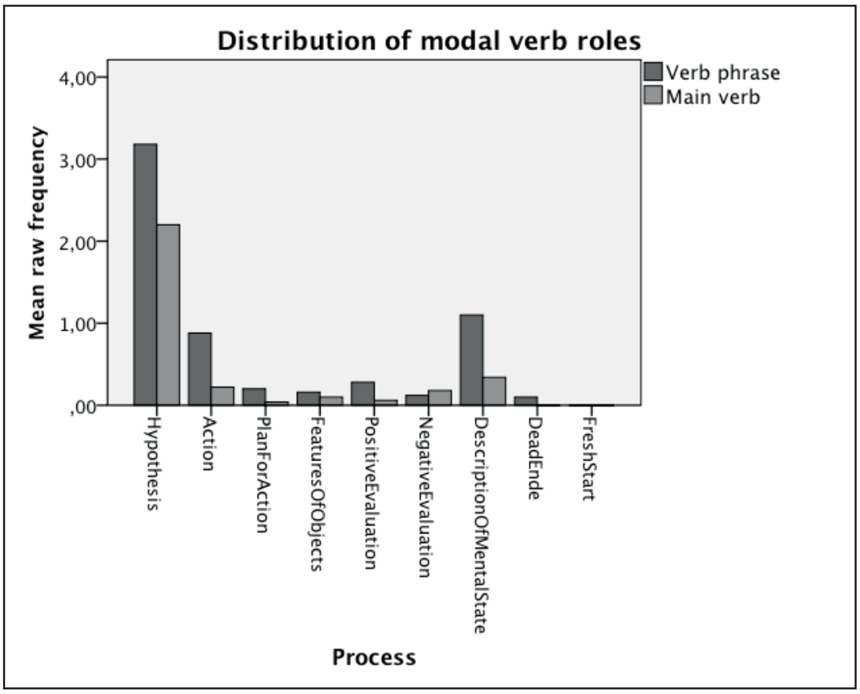


Figure 7.7: Distribution of modal verb roles within problem solving processes (by mean raw frequency).

A more fine-grained analysis of the type of modal verbs that were used adds a further aspect to the understanding of the different nature of the described problem solving processes. A chi-square test revealed highly significant differences of modal verb type between process categories, $L\chi^2$ (35, N = 458) = 114.07, p = .000[206].

The frequency analysis highlighted that 'können' (can) was the most frequent modal verb in all process categories except for the process of *description of mental state* (see Table 7.9). 'Sollen' (should) was more frequently used in *description of mental state* (23.6%, SD = 5.4; z = 4.1)[207] and *action* (2.0%, SD = 1.0; z = -2.5) than in all other process categories. Furthermore, 'können' (can) was observed significantly more often in *positive evaluations* (25.0%, SD = 6.1; z = 2.8) than expected. The modal verb 'müssen' (must) was more frequently used in *hypotheses* (40.6%, SD = 4.8; z = 2.9) and *description of mental state* (9.3%, SD = 3.7; z = -2.9), whereas it was used less frequently in *positive evaluations* (1.00%, SD = 1.00; z = -2.1) than in the other process categories.

Given that 458 modal verbs were identified, the results can be compared to the findings reported in Diewald (1996). This comparison highlighted that the general trend that 'können' and 'müssen' are most frequent was also observed in the

[206] Likelihood ratios are reported because the reported results vary markedly from the expected results.

[207] All standardized residuals are reported in Table 7.20 in the appendix.

elicited self-assembly data. There was only one category, namely *description of mental state*, in which this general trend did not show (see Table 7.9 for an overview).

modal verb	hypothesis	action	description of mental state	features of object parts	positive evaluation	negative evaluation
können (can)	41.46 % (5.00)	33.71 % (6.25)	20.30 % (4.97)	10.67 % (4.31)	25.00 % (6.10)	19.00 % (5.51)
müssen (must)	40.65 % (4.79)	19.96 % (4.81)	9.30 % (3.65)	4.00 % (2.80)	1.00 % (1.00)	7.00 % (3.50)
wollen (want)	0.90 % (0.63)	5.33 % (2.91)	5.71 % (2.89)	0 %	0 %	0 %
sollen (should)	10.92 % (2.75)	2.00 % (1.00)	23.56 % (5.40)	3.33 % (2.83)	2.00 % (2.00)	0 %
dürfen (allowed)	1.68 % (0.91)	1.00 % (1.00)	1.00 % (1.00)	0 %	0 %	0 %
möchten (like)	0.40 % (0.40)	1.00 % (1.00)	2.67 % (2.10)	0 %	0 %	0 %

Table 7.9: Frequency of modal verbs reported for the six most frequent process categories (mean of percentage with standard deviation in brackets) – 'mögen' (like) has not been coded as a modal verb in the analysis because it does not count as one of the core modal verbs.

Furthermore, it could be observed that 'wollen' was notably less frequent in the self-assembly data than in Diewald's (1996) and Wunderlich's (1981) analysis. Based on this comparison, a normal distribution[208] for modal verbs in *hypotheses* and *actions* and a diverging distribution for *description of mental state* can be assumed.

In order to test whether the observed trend holds for both syntactical constructions, modal verb type frequencies were analyzed for each possible construction separately, too. This part of the analysis of modal verbs was restricted to *hypothesis* and *description of mental state* since those two categories were observed to contain most modal verbs[209].

[208] Note that this term does not refer to the statistical phenomenon of normal distribution but rather expresses that the distribution is normal with regard to the reported frequency for spoken German (see section 1 of this chapter for more detail).

[209] For those interested in the distribution within the remaining processes see Figure 7.13 and Figure 7.14 in the appendix.

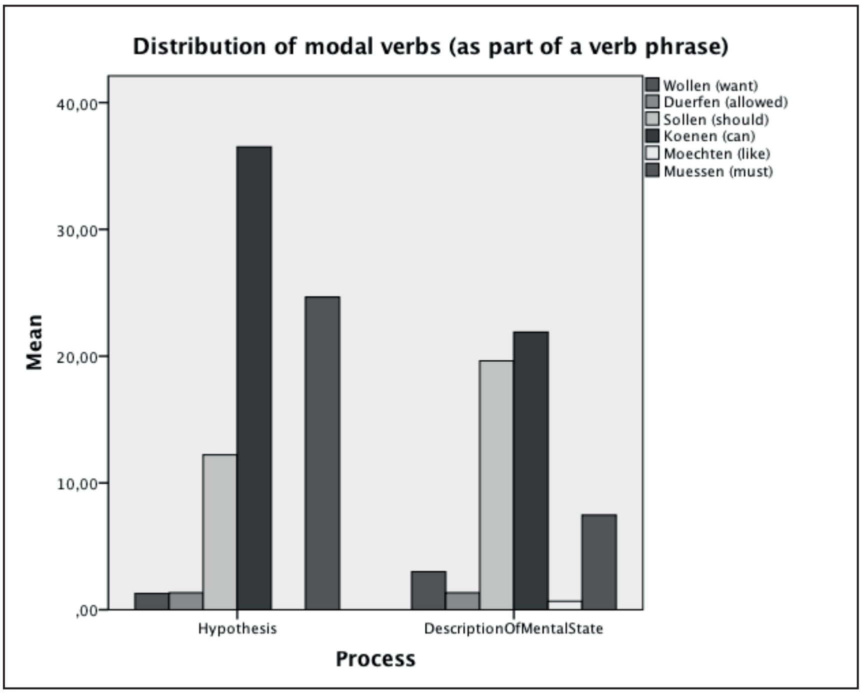


Figure 7.8: Distribution of modal verbs that are part of a larger verb phrase within *hypothesis* and *description of mental state* (by mean percentage).

First, the distribution of modal verb types in verb constructions was analyzed. The analysis revealed that 'können' (can) was the most frequently used modal verb in *hypotheses* (36.5%, SD = 38.8), followed by 'müssen' (must) (24.7%, SD = 31.2), and 'sollen' (should) (12.2%, SD = 25.2) (see Figure 7.8). 'Können' was also most frequently used in *descriptions of mental state* (21.9%, SD = 38.5). However, in this category 'sollen' (19.6%, SD = 37.1) was the second most frequently used form. In both process categories, 'möchten' (like) was the least frequently used modal verb type[210].

The distribution for modal verbs that fill the role of main verbs was different from the previously described pattern (see Table 7.10). The modal verb 'müssen' (must) was most frequently used in hypotheses (41.4%, SD = 44.8) and 'können' (can) less frequently (22.9%, SD = 37.6). Overall, modal verbs in main verb position were less frequent in *description of mental state*. Considering the distribution of the different modal verb types, a different picture emerged in this category than in *hypothesis*. In *description of mental state*, 'sollen' was most frequent (11.0%, SD = 30.8) and 'können' (6.0%, SD = 24.0) was also the second most frequent modal verb[211].

[210] All results are reported in Table 7.21 in the appendix.

[211] All results are reported in Table 7.22 in the appendix.

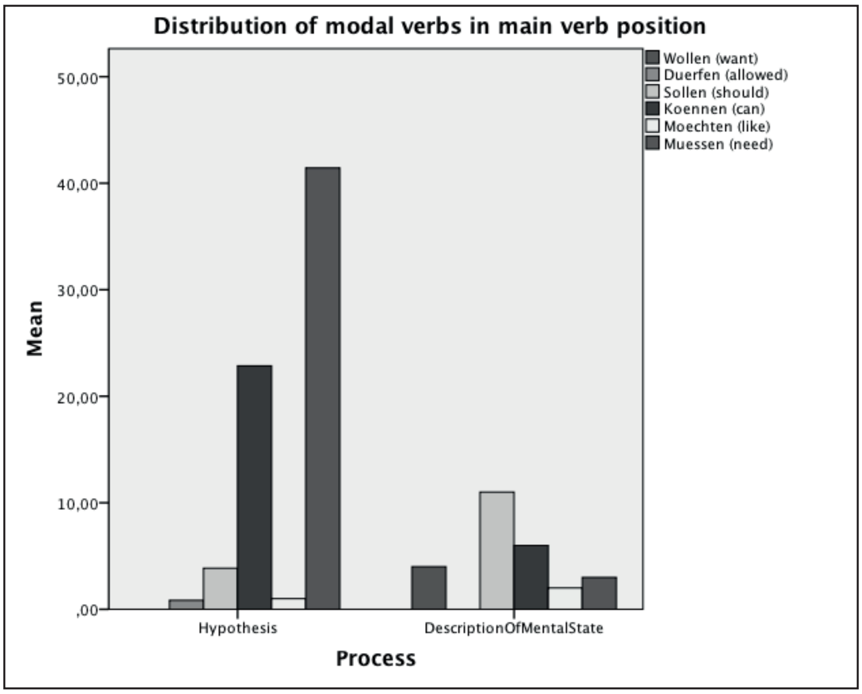


Table 7.10: Distribution of modal verbs in main verb position in *hypothesis* and *description of mental state* (by mean percentage).

7.4.1.3 Tense

The third grammatical feature that was coded for was tense. The distribution of the dependent variable 'tense' with its features (present, past, future) is displayed in Figure 7.9. The graph clearly shows that participants tended to use verbs in present tense. The frequency of present tense was different between *plan for action* (34.5%, SD = 43.0) and *hypotheses* (89.3%, SD = 11.6), *action* (92.6%, SD = 12.8), *positive evaluation* (85.3%, SD = 30.6), and *negative evaluation* (85.3%, SD = 30.1) (all results are summarized in Table 7.18 in the appendix). If past tense was used, those cases were observed in *descriptions of mental state* (6.5%, SD = 14.9). The future tense was sometimes used in stating *plans for action* (6.7%, SD = 19.1)[212].

[212] Raw frequencies are also reported in Table 7.17 in the appendix.

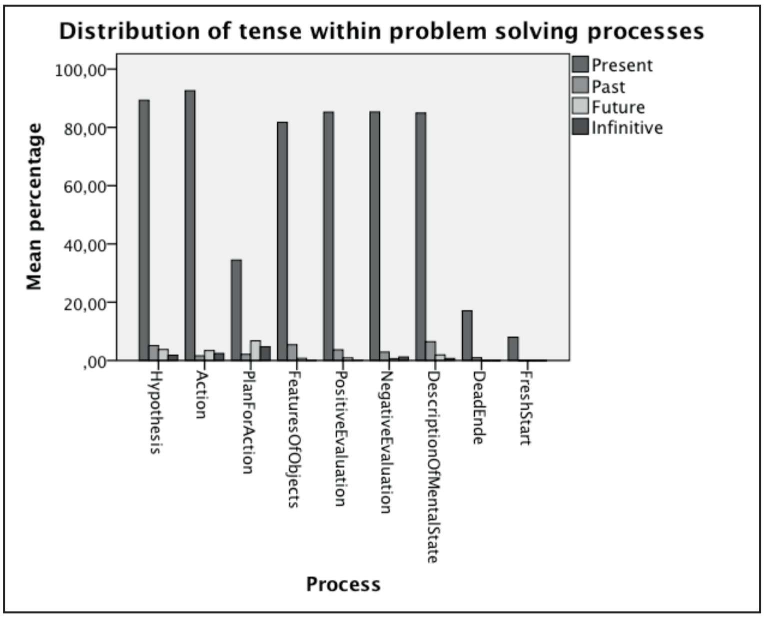


Figure 7.9: Distribution of tense within problem solving processes (by mean percentage).

7.4.2 Modal adverbs

7.4.2.1 Results

As outlined in the introduction of this chapter, modality can be expressed in different ways. Besides mood and modal verbs, modal adverbs serve as indicators of modality. In this part of the analysis the results for the annotation of modal adverbs are presented.

All in all 254 modal adverbs were identified. However, not all modal adverbs that were proposed by Brinkmann (1971) were observed in the think aloud protocols. The following adverbs were identified with differing frequencies of usage (see Table 7.11 for raw frequency): 'bestimmt' and 'sicher' (certainly), 'jedenfalls' (anyway), 'unbedingt' (necessarily), 'tatsächlich' (really), 'natürlich' (naturally), 'wahrscheinlich' (probably), 'vermutlich' (presumably), 'vielleicht' (maybe), 'wohl' (probably). Three participants did not use any modal adverbs in their protocols.

The frequency analysis revealed that 'vielleicht' (maybe) was the most frequently used modal adverb and it was used by the majority of participants (70%). The modal adverb 'wahrscheinlich' (probably) was the second most frequent adverb, being used in 42% of all protocols. The third most frequent adverb, namely 'bestimmt' (certainly), was used to an almost comparable amount of times in protocols but by only 1/3 of the participants. No adverbs encoding emotional quality were identified in the data.

To conclude, modal adverbs that signal the tentative state of an utterance, i.e. function 2, were most frequently used in think aloud protocols as compared to modal adverbs that signal certainty, i.e. function 1.

modal adverb	frequency (overall)	number of participants who used it	function assigned by Brinkmann (1971:401)
vielleicht	127 (50.00%)	35 (70.00%)	2 (assumption)
wahrscheinlich	50 (19.69%)	21 (42.00%)	2 (assumption)
bestimmt	37 (14.57%)	15 (30.00%)	1 (certainty)
wohl	17 (6.69%)	12 (24.00%)	2 (assumption)
sicher	13 (5.12%)	9 (18.00%)	1 (certainty)
jedenfalls	5 (2.00%)	4 (8.00%)	1 (certainty)
vermutlich	2 (< 1.00%)	2 (4.00%)	2 (assumption)
unbedingt	1 (< 1.00%)	1 (2.00%)	1 (certainty)
tatsächlich	1 (< 1.00%)	1 (2.00%)	2 (fact)
natürlich	1 (< 1.00%)	1 (2.00%)	2 (fact)
sum	254 (100%)		

Table 7.11: Summary of occurrences of modal adverbs in think aloud protocols ordered by frequency from most to least frequent (raw numbers with percentage in brackets).

After presenting this general overview, the distribution of modal adverbs will be considered from two perspectives, i.e. with regard to the different conditions and the process categories that were distinguished in the analysis of this chapter. No differences of distribution between conditions were observed. The distribution reflects the general trend that was described above (see Table 7.23 in the appendix). Nevertheless, it is interesting that the two occurrences of modal adverbs belonging to the category 1, i.e. signaling that a fact is being stated, were used by two participants with few prior knowledge (i.e. the *underspecified* and *verbal goal* condition) (see Figure 7.10).

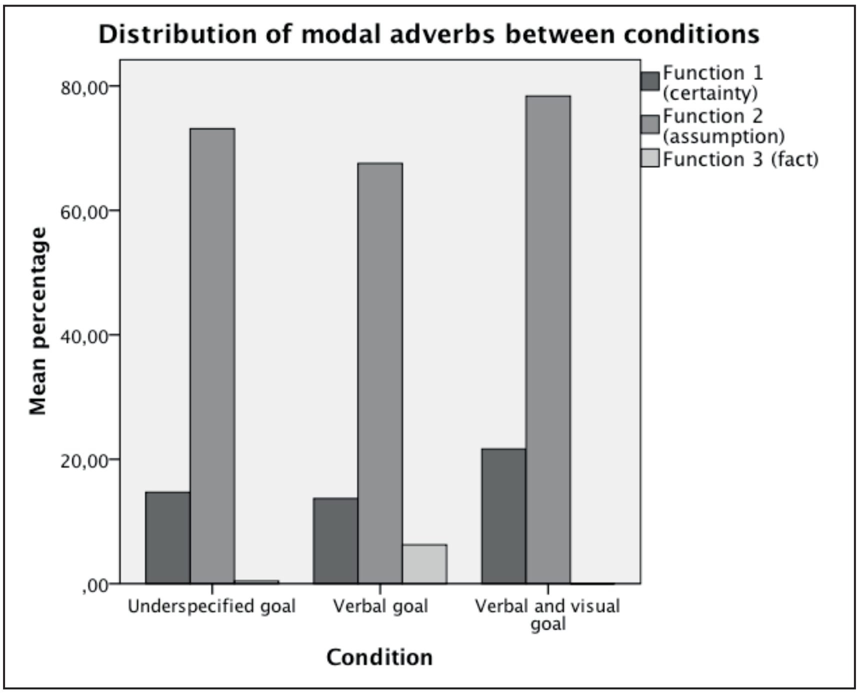


Figure 7.10: Distribution of modal adverbs (categorized according to their function) between conditions (based on mean percentage).

The distribution of modal adverbs between process categories highlighted that markers signaling the tentative character of an utterance were most frequent in all process categories (see Figure 7.11). The only process that was different from this general pattern was the category of *description of mental state* in which modal adverbs signaling certainty were used as frequently as modal adverbs signaling tentativeness (see Table 7.24 in the appendix). The raw frequency of modal adverbs was very low in this category; thus, no conclusions can be drawn but the emerging trend should be noted nevertheless.

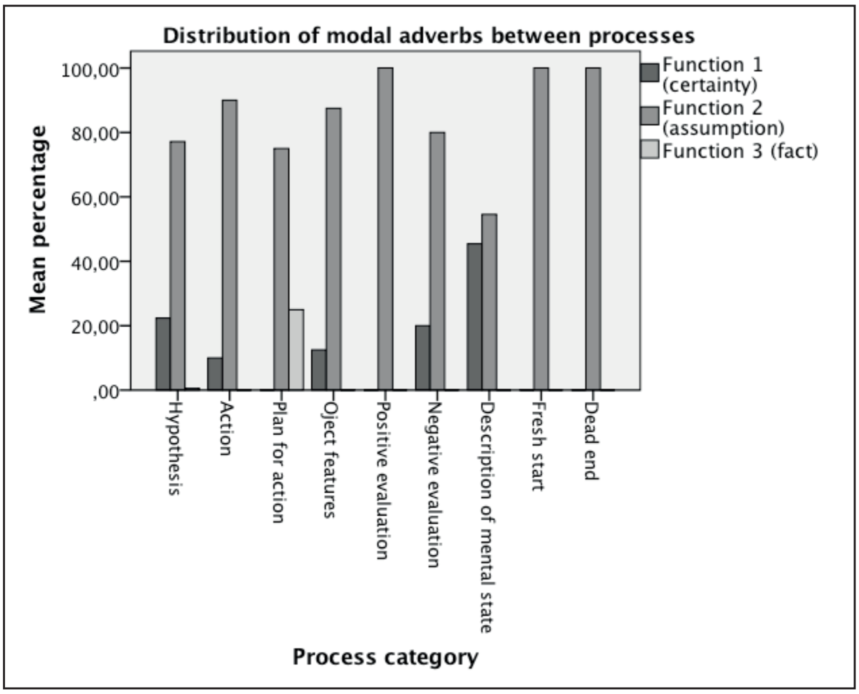


Figure 7.11: Distribution of modal adverbs (categorized according to their function) between process categories (based on percentage overall).

7.4.2.2 Summary

Summarizing and combining all results, the following description of verb phrase form emerges for the respective process categories:

- *Hypothesis*: verbs of *being and having* in indicative mood (85.3%) or subjunctive mood (12.8%) and present tense; modal verbs 'können' (can) (36.5%), 'müssen' (must) (24.7%), and 'sollen' (should) (12.2%); modal adverbs signaling tentativeness ('vielleicht', 'wahrscheinlich')
- *Action*: verbs of *doing and happening* in indicative mood and present tense
- *Plan for action*: verbs of *doing and happening* in indicative mood and present tense (34.5%) or future tense (6.7%)
- *Features of objects*: verbs of *being and having* in indicative mood and present tense
- *Positive evaluation*: verbs of *being and having* in indicative mood and present tense
- *Negative* evaluation: verbs of *being and having* in indicative mood and present tense
- *Description of mental state*: verbs of *sensing* in indicative mood and present tense; modal verbs 'können' (can) (21.9%) and 'sollen' (should) (19.6%)
- *Dead end*: verbs of *sensing* (7.5%) and *being and having* (7.1%) in indicative mood and present tense
- *Fresh start*: verbs of *doing and happening* in indicative mood and present tense

7.5 Discussion

Research question 1 investigated whether there were systematic differences between problem solving processes with regard to verb type. Research question 2 studied the representation of modality in the identified verb phrases. A suggestive picture about the linguistic representation of each process category can only be constructed by combining verb type and modality. Therefore, these two research questions are discussed together.

In order to engage in a solid discussion, the overall frequency of the different processes needs to be taken into account. As outlined in the first paragraph of the results section, the majority of processes were *hypotheses*, *evaluations* and *actions*. *Descriptions of mental state* were more frequent than *features of object parts*. *Plans for action*, *dead end* states, and *fresh start* were rarely verbalized. Since the aim of this study was to investigate the linguistic nature of these processes, all of them will be discussed in turn but the validity of the conclusions is higher for a greater data sample than for a very small one as in the cases of the three last mentioned processes. This being said the discussion is guided by the order of frequency, starting with the most frequent one.

The results revealed that *hypotheses* contained verbs of *being and having* at a very high frequency indicating persistence over time without any active engagement by the speaker. If plans were formulated, verbs of *doing and happening* are used as in "das Dach kommt oben rauf" (the roof goes on top). The low frequency of future tense marking suggests that future goals were rarely mentioned. As expected, present tense, which is the unmarked tense of both verb types (Halliday & Matthiessen 2002:135), was clearly the most frequently used tense. Furthermore, *hypotheses* were either expressed with certainty as in „das **ist** unten“ ("this is below") or signal a tentative character as in „<u>eigentlich</u> **müsste** hier, <u>glaube ich</u>, eine Ecke hin“ ("actually there needs to be a corner part here I think"). Since the indicative mood was the most frequently found verb mood in *hypotheses*, it can be concluded that most hypotheses were uttered with certainty. Nevertheless, there were examples in which tentativeness was expressed by various means. In the example above, the modal verb 'müssen' (must) signals that the speaker's idea, emphasized by the phrase "I think", was externally constraint by requirements. The tentative character of 'could' further stressed by the use of modal adverbs, such as 'vielleicht' (maybe) and 'wahrscheinlich' (probably).

As outlined in the introductory section of this chapter, the distribution of modal verbs in the elicited think aloud protocols was compared to that described for spoken German in the literature in order to discuss the observed distribution and implications of modal verb type. The comparison between the elicited data and that reported in the literature highlighted that the self-assembly findings on the distribution of the two syntactic possibilities of modal verb position reflect Wunderlich's (1981) findings. He reported that most modal verbs were used within a verb phrase and this was also observed in the elicited data. Nevertheless, he did not distinguish between modal verbs in main verb position and modal verbs within a verb phrase in his frequency analysis. Diewald (1996) did not address these two syntactic options at all. Thus the results of the present study are compared to Wunderlich's (1981) findings. The comparison revealed the same distribution for all process categories except for *description of mental state*. Based on this comparison, it can be assumed that the identified patterns of modal verb distribution reflect the general pattern in German.

The analysis of the distribution of modal verbs highlighted that *hypotheses* contained forms of 'können' (can) at the same frequency as forms of 'müssen' (must). Based on Engel's (2002) description of subject-related use[213] of modal verbs, 'können' signals that the speaker has the possibility to act upon his/her intentions because of his/her abilities, i.e. the action or idea is subject driven. The modal verbs 'sollen' and 'müssen' signal that intentions and ideas are driven by external forces. In the case of 'sollen' these external forces are expectations raised by a third party. In the case of 'müssen' requirements influence the speaker's intentions. The finding that *hypotheses* were marked by the use of 'können' and 'müssen' suggests that there were two influencing forces with regard to the formulation of hypotheses, i.e. subject-driven possibilities and requirement-based prerequisites for expressed ideas and intentions. In some cases, the tentative character of *hypotheses* was stressed by the use of modal adverbs. This process category contained the highest frequency of modal adverbs and the majority of them belonged to the functional category marking an assumption, namely 'vielleicht' (maybe) and 'wahrscheinlich' (probably).

Halliday and Matthiessen (2002:131) argue that verbs of *doing and happening* encode dynamic processes involving input of energy that leads to a change of the actualization through time. Given the finding that *actions* contain verbs of *doing and happening* at the highest frequency within its process category and between all

[213] His descriptions of subject-related usage have been summarized in sub-section 1 of this chapter.

process categories, it can be argued that the category *action* refers to the execution of actions. Furthermore, verbs in the category *action* were mostly marked by present tense signaling the situatedness of the performed action. The finding that most verbs were used in indicative mood further adds to the picture of certainty and performance. If uncertainty was expressed, it was done by modal verbs. In *actions* 'können' was more frequent than 'müssen'. This suggests that the subject's abilities influenced the actions that were performed more than requirements that needed to be fulfilled. As expected, *actions* could be distinguished from *plans for action* by tense marking. If future marking was observed at all, it was in *plan for action*. Nevertheless, the unmarked[214] present tense was most frequent in this category as well.

A third process category that was frequently observed in the data but not described in the literature on problem solving so far was *description of mental state*. This category could be distinguished from the previously mentioned processes very clearly. First of all, this process category contained the highest frequency of verbs of *sensing*. These are verbs that encode mental activities, such as 'think', 'hope', 'assume', 'believe', and verb phrases which signal that the speaker turns to himself/herself for answers as in 'ask myself'. Verbs of *sensing* encode the projection of interior content and ideas into 2nd order semiotic reality (Halliday & Matthiessen 2002:128). Therefore, this category contains the speaker's projected thoughts that become part of the present context through their verbalization. The observation that those verbs were mainly used in the present tense stresses the focus on the present context and task. Additionally, the distribution of modal verb type frequency in this process category was a little different from that reported for German in the literature. Whereas 'können' is the most frequent modal verb in the literature, 'müssen' was most frequent in *description of mental state*. This finding suggests that requirements were more important in internal consultations than the speaker's abilities. The distribution of modal adverbs was also different in this category than the one observed in all other process categories. The analysis highlighted that modal adverbs signaled certainty and tentativeness at the same frequency. This suggests that speakers either described their mental state with certainty (see example (147)) or expressed uncertainty about the present state of the assembly process (see example (148)).

[214] The term 'unmarked' refers to the observation that the present tense is the most frequently used tense with verbs of *doing and happening* (Halliday & Matthiessen 2002:135).

(147) „Ich bin mir relativ sicher.“ (“I am almost certain.”)

(148) „... dann vielleicht / wo können die hin hinkommen ...“ (“... then maybe/ where can those be placed ...”)

The linguistic structure of evaluation processes did not differ whichever conclusions were drawn, i.e. either positive or negative. In both categories verbs of *being and having* were most frequent and they were used in the unmarked present tense. This finding confirms the expectation that participants evaluated the stages and outcomes of their actions at the moment of observation and not in retrospect.

Considering the findings for all identified process categories, it can be concluded that systematically different patterns were identified. However, considering the frequency of modal verb types, it was observed that requirements were of greater importance to the assemblers than expectations. This finding can be explained by the fact that in two conditions participants were provided with specific requirements regarding the goal object. Furthermore, these results mirror the findings of the content-based analysis on the use of task specific constraints which revealed that participants verbalized physical constraints and adhered to them in consecutive actions (see chapter 9). Expectations that are expressed by ‘sollen’ (should) are externally presented to the speaker (Engel 2002). In the self-assembly experiment no expectations regarding time or performance were formulated in the instructions; this context may explain why ‘sollen’ was only rarely observed.

In sum, the analysis of verb form revealed a systematic difference of verb type between the identified problem solving processes. The results highlighted that *hypotheses* differ from *actions*, and both of these differ from *descriptions of mental state*. Furthermore, characteristic features of think aloud protocols were identified in the frequency of verbs of *saying* and the frequency of subjunctive mood. Therefore, it can be concluded that the goal of identifying the linguistic structure of problem solving processes has been achieved.

Moreover, the combination of these results highlights specific features of verb phrase form with regard to the task at hand. First, instances of verbs of *saying* were rarely identified in the protocols. This finding can be interpreted as a specific feature of think aloud protocols in which the communicative intention is very low if it is at the speaker’s mind at all. In the self-assembly scenario participants were not encouraged to communicate; thus they do not need to use reported speech or introduce sentences with phrases such as ‘I would say’. Therefore, verbs of *saying* are not needed to structure discourse coherently. Second, the present tense is the most frequently used verb mood. This result needs to be interpreted against the backdrop of the task specific instructions and conditions. In the self-assembly task participants were not given any specific constraints as to how the objects can be placed but most importantly they were free to disassemble the object at any time. Thus there was no need for mental simulation as to where to put a specific object and in which order they need to be assembled.

One indicator for mental simulation is the use of subjunctive mood because it signals that the reported event is assumed (e.g. Glinz 1973). Furthermore, modal verbs encode the simulation character of what is being said because they signal that something is not done yet but will be (indicative mood) or shall be (subjunctive mood) done in the future. In order to assess if the frequency of markers of mental simulation is notably low in the self-assembly data, it was compared to another data set of think aloud protocols which were elicited while performing a task in which planning (one kind of mental simulation) was a necessary prerequisite for success. This data was elicited by Sven Brüssow, Daniel Holt, and Joachin Funke at the university of Heidelberg[215]. The researchers asked participants to think aloud while solving a specific scheduling task called "Plan a day" (abbreviated PAD) (for more details on the design see Brüssow et al. (in preparation)). In a collaborative work with Sven Brüssow[216] I have been looking for the best way of capturing and visualizing the differences in the amount of mental simulation expressed in the protocols. The visualization of a correspondence analysis seems to be the most informative representation (see Figure 7.12). The analysis revealed more indicators for simulation processes in think aloud protocols collected in PAD than in "Unaided Object Assembly" (abbreviated UOA). As outlined above, the majority of verbs that were used in UOA were main verbs in indicative mood. There were only few instances of subjunctive mood and modal verbs. Verb forms used in PAD, in contrast, showed a greater variability with regard to type as well as mood. The search for clusters that were representative for each task revealed that PAD could be described in these terms (see Figure 7.12). However, the graph highlighted clearly that modal verbs (in indicative as well as subjunctive mood) were more frequent in PAD than in UOA and the subjunctive mood was generally more frequent in PAD. If these findings are related to the theoretical assumptions about indicators of simulation processes, a clear picture emerges; participants who were planning the errands for a day engaged in more simulation processes than participants who assembled the dollhouse for

[215] I thank all three researchers for offering us the great opportunity to work with them on this data set. I want to specially thank Sven Brüssow with whom I devised and ran the analysis.

[216] This is ongoing work and a first publication together with Daniel Holt, Thora Tenbrink, and Joachim Funke referred to as Brüssow et al. (in preparation).

themselves. Furthermore, it can be observed that participants engaged in the assembly task used the subjunctive mood frequently with modal verbs which implies that they combine both markers of simulation processes in one verb form thereby strengthening the simulation character.

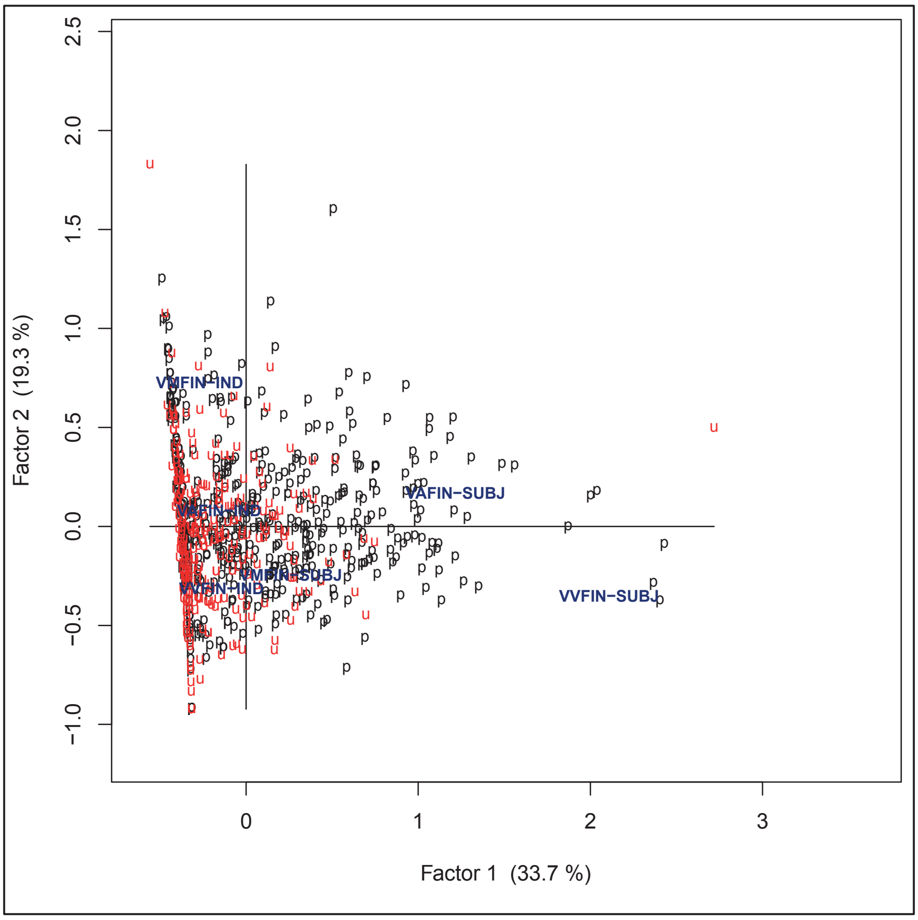


Figure 7.12: Correspondence analysis of the frequencies of verbs in the indicative and subjunctive mood in verbal protocols for PAD (p) and UOA (u); statistical analysis and graphical representation by Sven Brüssow (Verb labels reflect the respective STTS tag and mood information (VMFIN-IND modal verb, finite, indicative; VAFIN-IND auxiliary, finite, indicative; VVFIN-IND main verb, finite, indicative; VMFIN-SUBJ modal verb, finite, subjunctive; VAFIN-SUBJ auxiliary, finite, subjunctive; VVFIN-SUBJ main verb, finite, subjunctive)).

To sum up, in addition to differences in the distribution of verb type between problem solving processes the analysis of verb mood and modal verbs suggests an influence of the task on observed mental planning activity. The differences between the analyzed task of self-assembly in a largely unconstrained context and a problem solving task in which planning was a prerequisite for success suggest that the assembly encouraged situated, action oriented behavior rather than mental simulation.

7.6 Appendix

7.6.1 Results Verb type, mood, and tense

process	being & having	doing & happening	sensing	saying	ellipsis	modals
hypothesis	12.24 (9.71)	5.12 (4.26)	2.42 (2.48)	0.18 (0.48)	0.48 (0.99)	2.20 (3.19)
action	1.12 (1.44)	7.26 (5.83)	0.70 (1.20)	0.02 (0.14)	0.24 (0.69)	0.22 (0.55)
plan for action	0.28 (0.78)	1.32 (2.63)	0.18 (0.44)	0.00	0.02 (0.14)	0.04 (0.20)
features of object parts	3.66 (4.05)	0.70 (1.22)	0.30 (0.79)	0.02 (0.14)	0.14 (0.61)	0.10 (0.30)
positive evaluation	3.32 (3.05)	0.68 (1.11)	0.60 (0.95)	0.06 (0.24)	0.06 (0.31)	0.06 (0.31)
negative evaluation	2.92 (2.97)	1.20 (1.78)	0.72 (1.41)	0.00	0.08 (0.34)	0.18 (0.39)
description of mental state	3.16 (3.23)	2.38 (2.57)	2.94 (2.99)	0.20 (0.45)	0.14 (0.50)	0.34 (0.72)
dead end	0.32 (0.94)	0.22 (0.82)	0.24 (0.72)	0.00	0.00	0.00
fresh start	0.00	0.12 (0.39)	0.00	0.00	0.02 (0.14)	0.00
overall	3.00 (5.34)	2.11 (3.65)	0.90 (1.80)	0.05 (0.25)	0.13 (0.52)	0.35 (1.30)

Table 7.12: Results of distribution of verb classes with regard to problem solving processes; mean of raw frequencies (standard deviation in brackets).

process	being & having	doing & happening	sensing	saying	ellipsis	modals
hypothesis	52.97 (16.74)	22.59 (13.35)	12.05 (12.18)	0.87 (3.31)	2.16 (4.37)	9.36 (10.69)
action	10.77 (13.79)	77.62 (18.02)	7.07 (11.47)	0.17 (1.18)	1.92 (6.17)	2.46 (6.34)
plan for action	7.29 (21.29)	33.29 (43.61)	5.02 (16.63)	0.00	0.29 (2.02)	2.12 (14.15)
features of object parts	63.96 (38.02)	17.54 (28.65)	3.52 (8.94)	0.63 (0.44)	3.17 (14.93)	1.75 (5.98)
positive evaluation	63.42 (34.20)	12.95 (23.69)	9.50 (17.28)	1.44 (7.38)	0.55 (3.01)	2.14 (9.92)
negative evaluation	57.00 (37.49)	18.45 (25.41)	8.79 (16.74)	0.00	1.09 (4.56)	4.67 (16.14)
description of mental state	30.78 (24.49)	22.93 (19.70)	32.69 (25.24)	3.86 (15.02)	1.23 (4.30)	2.51 (5.40)

dead end	7.06 (19.34)	3.46 (11.05)	7.48 (22.10)	0.00	0.00	0.00
fresh start	0.00	10.00 (30.30)	0.00	0.00	2.00 (14.14)	0.00
overall	32.58 (35.91)	24.31 (35.56)	9.57 (18.26)	0.71 (5.78)	1.38 (7.68)	2.78 (9.60)

Table 7.13: Results of distribution of verb classes with regard to problem solving processes; mean of percent (standard deviation in brackets).

process category/ Verb type	being & having	doing & happening	sensing	saying
hypothesis	4.8	-5.2	-2.5	0
action	-11.1	16.6	-3.9	-1.5
plan for action	-4.5	6.4	-1.1	-0.9
features of object parts	6.5	-5.0	-3.3	-0.7
positive evaluation	5.1	-5.1	-0.6	0.7
negative evaluation	2.4	-2.6	0	-1.5
description of mental state	-3.9	-2.6	10.3	3.2
dead end	-0.6	-0.6	2.7	-0.6
fresh start	-1.7	2.8	-0.9	-0.2

Table 7.14: Cross table with standardized residuals for verb mood × process category.

process	indicative	subjunctive	imperative	infinitive
hypothesis	18.66 (12.57)	3.14 (3.45)	0.02 (0.14)	0.36 (0.78)
action	8.70 (7.16)	0.16 (0.37)	0.12 (0.36)	0.20 (0.53)
plan for action	1.44 (2.82)	0.08 (0.27)	0.00	0.30 (0.76)
features of object parts	4.64 (4.92)	0.14 (0.41)	0.00	0.04 (0.28)
positive evaluation	4.28 (3.65)	0.44 (0.84)	0.00	0.02 (0.14)
negative evaluation	4.80 (4.87)	0.16 (0.51)	0.00	0.08 (0.27)
description of mental state	8.42 (7.18)	0.38 (0.90)	0.06 (0.31)	0.10 (0.30)

dead end	0.76 (2.05)	0.20 (0.14)	0.00	0.00
fresh start	0.10 (0.36)	0.00	0.02 (0.14)	0.00
overall	5.76 (8.11)	0.50 (1.56)	0.02 (0.18)	0.12 (0.45)

Table 7.15: Results of distribution of mood with regard to problem solving processes; mean of raw frequencies (standard deviation in brackets).

process category/Verb type	indicative	subjunctive	imperative
hypothesis	-2.0	7.3	-1.5
action	1.1	-4.5	3.2
plan for action	0.3	0.8	-0.6
features of object parts	0.8	-2.7	-1.0
positive evaluation	-0.2	0.9	-1.0
negative evaluation	0.8	-2.5	-1.0
description of mental state	0.7	-2.6	0.9
dead end	0.4	-1.2	-0.4
fresh start	-0.2	-0.7	5.8

Table 7.16: Cross table with standardized residuals for verb mood × process category.

process	present	past	future
hypothesis	19.54 (13.82)	1.36 (2.74)	0.82 (1.08)
action	8.62 (6.97)	0.18 (0.56)	0.38 (1.24)
plan for action	1.08 (1.82)	0.04 (0.20)	0.40 (0.20)
features of object parts	4.54 (4.95)	0.14 (0.35)	0.06 (0.24)
positive evaluation	4.46 (3.92)	0.22 (0.51)	0.04 (0.20)
negative evaluation	7.86 (6.89)	0.16 (0.37)	0.10 (0.36)
description of mental state	8.42 (7.18)	0.70 (1.39)	0.26 (0.49)
dead end	0.74 (2.03)	0.04 (0.20)	0.00
fresh start	0.10 (0.36)	0.00	0.00
overall	5.74 (8.42)	0.32 (1.14)	0.23 (0.75)

Table 7.17: Results of distribution of tense with regard to problem solving processes; mean of raw frequencies (standard deviation in brackets).

process	present	past	future
hypothesis	89.30 (11.64)	5.12 (8.47)	3.77 (5.91)
action	92.59 (12.79)	1.06 (5.29)	3.39 (8.60)
plan for action	34.46 (42.98)	2.12 (14.15)	6.74 (19.08)
features of object parts	81.71 (36.61)	5.42 (20.10)	0.75 (3.46)
positive evaluation	85.25 (30.57)	3.67 (9.82)	0.90 (4.48)
negative evaluation	85.27 (30.11)	2.92 (7.87)	0.63 (2.49)
description of mental state	84.90 (27.24)	6.47 (14.88)	1.94 (4.10)
dead end	17.05 (37.05)	0.95 (5.09)	0.00
fresh start	8.00 (27.40)	0.00	0.00
overall	64.28 (43.96)	3.14 (11.20)	2.01 (7.88)

Table 7.18: Results of distribution of tense with regard to problem solving processes; mean of percent (standard deviation in brackets).

7.6.2 Distribution of modal verbs within process categories

process category	within verb phrase	main verb
hypothesis	3.18 (3.52)	2.20 (3.19)
action	0.88 (1.27)	0.22 (0.55)
plan for action	0.20 (0.53)	0.04 (0.20)
features of object parts	0.16 (0.51)	0.10 (0.20)
positive evaluation	0.28 (0.54)	0.06 (0.24)
negative evaluation	0.12 (0.44)	0.18 (0.39)
description of mental state	1.10 (1.34)	0.34 (0.72)
dead end	0.10 (0.42)	0
fresh start	0	0.
overall	0.67 (1.67)	0.35 (1.30)

Table 7.19: Distribution of modal verbs between process categories depending on the modal verb position; mean raw frequency with standard deviation in brackets.

process category/ modal verb type	wollen (want)	dürfen (allowed)	sollen (should)	können (can)	müssen (must)	mögen (like)
hypothesis	-1.3	-0.1	-1.0	-1.4	2.9	-1.3
action	1.6	0.2	-2.5	1.3	-0.4	0.3
plan for action	1.0	-0.4	0.2	-1.0	-0.2	4.6
features of object parts	-0.7	-0.4	0.8	1.0	-1.2	-0.4
positive evaluation	-0.7	-0.5	-0.9	2.8	-2.1	-0.5
negative evaluation	-0.7	-0.5	-1.5	1.8	-0.6	-0.4
description of mental state	1.7	0.9	4.1	-0.6	-2.9	1.1
dead end	-0.4	-0.3	2.7	-0.1	-1.3	-0.3

Table 7.20: Standardized residuals of modal verb type × process category.

modal verb type	hypothesis	description of mental state
wollen (want)	1.28 (6.35)	3.00 (15.68)
dürfen (allowed)	1.33 (6.60)	1.33 (9.43)
sollen (should)	12.22 (25.23)	19.63 (37.09)
können (can)	36.50 (38.82)	21.90 (38.52)
möchten (like)	0	0.67 (4.71)
müssen (must)	24.66 (31.17)	7.47 (22.44)

Table 7.21: Distribution of modal verb type (in verb phrase constructions) in *hypotheses* and *description of mental state*; mean percentage with standard deviation in brackets.

modal verb type	hypothesis	description of mental state
wollen (want)	0	4.00 (19.79)
dürfen (allowed)	0.85 (4.86)	0
sollen (should)	3.86 (11.85)	11.00 (30.79)
können (can)	22.85 (37.63)	6.00 (23.99)
möchten (like)	1.00 (7.07)	2.00 (14.14)
müssen (must)	41.43 (44.82)	3.00 (15.68)

Table 7.22: Distribution of modal verb type (in main verb position) in *hypotheses* and *description of mental state*; mean percentage with standard deviation in brackets.

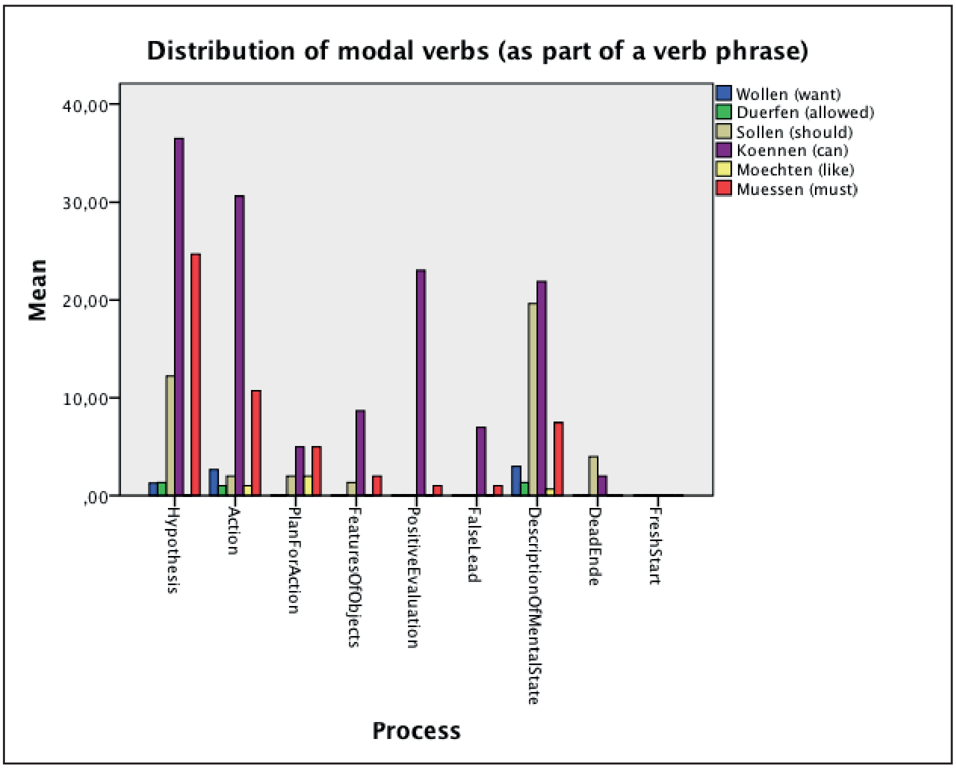


Figure 7.13: Distribution of modal verbs that are part of a larger verb phrase within problem solving processes (by mean percentage) (false lead (a term adapted from Palmer (1977)) refers to negative evaluations).

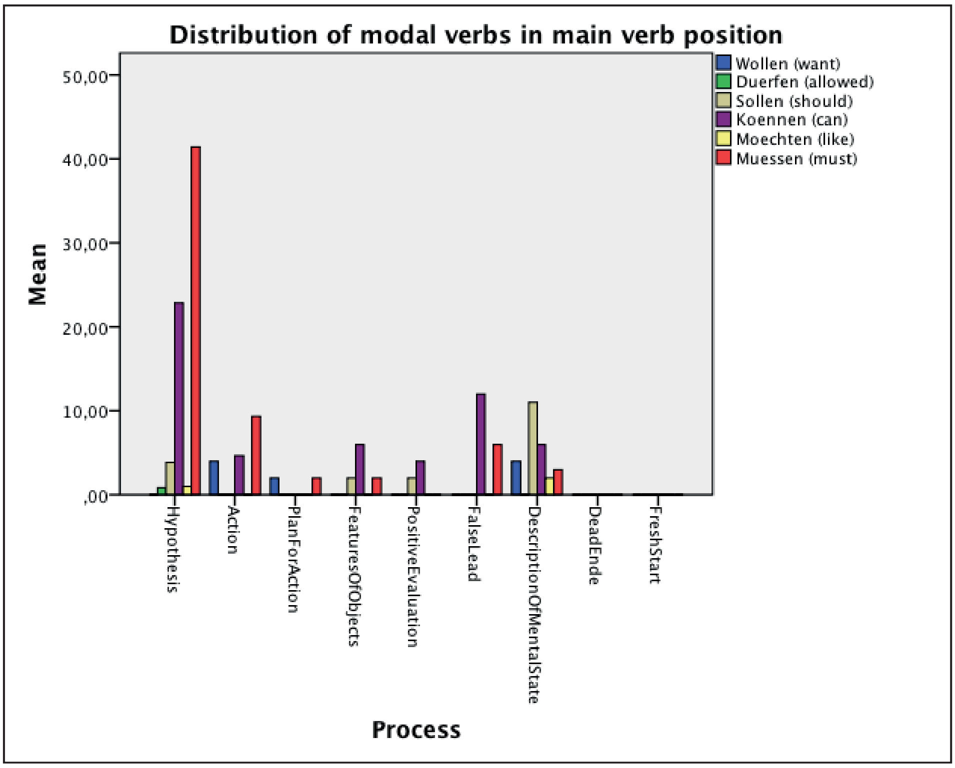


Figure 7.14: Distribution of modal verbs in main verb position within problem solving processes (by mean percentage) (false lead (a term adapted from Palmer (1977)) refers to negative evaluations).

7.6.3 Results modal adverbs

condition	function 1 (certainty)	function 2 (assumption)	function 2 (fact)
underspecified goal	14.70% (4.45)	73.12% (7.89)	0.42% (0.42)
verbal goal	13.69% (5.00)	67.56% (9.60)	6.25% (6.25)
verbal and visual goal	21.64% (7.57)	78.36% (7.57)	0

Table 7.23: Results distribution of modal adverbs (categorized according to their function) between conditions (mean percentage with standard deviation in brackets).

process category	function 1 (certainty)	function 2 (assumption)	function 2 (fact)
hypothesis	42 (22.34%)	145 (77.13%)	1 (0.53%)
action	1 (10.00%)	9 (90.00%)	0
plan for action	0	3 (75.00%)	1 (25.00%)
features of object parts	1 (12.50%)	7 (87.50%)	0
positive evaluation	0	2 (100.00%)	0
negative evaluation	2 (20.00%)	8 (80.00%)	0
description of mental state	5 (54.45%)	6 (54.55%)	0
fresh start	0	2 (100.00%)	0
dead end	0	1 (100.00%)	0

Table 7.24: Results distribution of modal adverbs (categorized according to their function) between process categories (raw frequency with percentage in brackets).

8 Features of analysis: object reference

8.1 State of the Art

8.1.1 Traditional accounts: references in communicative settings

> "People create symbols in order to use them as instruments in acts of communication – and these communicative acts always involve another person as recipient." (Tomasello 1996:230)

Referential form can be very different depending on context (e.g. Laucht 1982; Ungerer & Schmid 2006), the speaker's memory (e.g. Horton & Gerrig 2005a), and the speaker's ability to conceptualize the object or event (Mangold-Allwinn et al. 1995). As an example consider the situation in which we want our friend to hand us the milk. We may either use the proper noun, i.e. *milk*, or we may refer to it by an exophoric reference.

(1) Please pass me the milk.
(2) Please pass me that milk that you bought this morning.
(3) Please pass it to me.

Although the noun is identical, there is a difference between examples (1) and (2) regarding the information provided about the referent. In example (1) the noun phrase contains enough semantic information to identify the referent. This also holds for example (2) but in this case the reference is more specific pointing out an object that has been encountered before. This information is encoded in the demonstrative *that*. In example (3) the referential expression *it* does not carry sufficient semantic meaning. It is rather an act of verbal pointing which may be visually supported by a pointing gesture.

As this example shows and Tomasello (1996) rightly pointed out, naming is in almost all cases embedded in an act of referring for an addressee hence his/her needs need to be considered by the speaker as well. This fact reveals the fundamental difference between the presented ways of pointing out an entity. The use of deictic expressions is only successful if the desired object is salient, i.e. cognitively or visually accessible to the interlocutor. If the speaker is uncertain about the listener's current focus of attention, the use of a proper noun can be more successful. However, in order to use a proper noun its semantic information has to be known to both speaker and listener. Von Stutterheim et al. (1993) showed that specific nominal references (i.e. subordinate terms) are only helpful for the identification of a referent if object parts closely resemble generally known features of the associated concept. In their study, participants were told to instruct someone else on the assembly of a robot. If there is no clear resemblance between the object and the concept, references to objects are based on physical properties. They argue that referring to a specific concept in cases in which resemblance is not salient does not help in solving the task and involves additional work for the addressee (von Stutterheim et al. 1993:118). Additionally, Clark and colleagues found that instructor and addressee engage in negotiation processes until a shared referential term is established if no conventional referring expression exists (e.g. Clark 2005).

It is generally assumed that speakers tailor the content and linguistic form of a reference inferring the addressee's needs, a phenomenon called *audience design* (e.g. Clark 1996; Clark & Krych 2004; Arnold 2008). Methodologically, many studies have been conducted in an interactive setting such as the referential *communication paradigm* (e.g. Krauss & Weinheimer 1966; Hermann & Deutsch 1976; Clark & Wilkes-Gibbs 1986; Horton & Gerrig 2002, 2005a; Fukumura & van

Gompel 2012; Matthews et al. 2012). Different strategies for tailoring an utterance according to the listener's needs as inferred by the speaker have been described in the literature. The following two sections will focus on two of them, namely salience of objects that is expressed in linguistic form and the conceptualization of a given object as expressed in nominal specificity.

As will be shown in the literature review most studies that investigate referential form and referential change over time focus on interactive settings, i.e. those in which an addressee is present or imagined. Nevertheless, there are also a few studies that point out that referential form may also be influenced by the speaker's own mental capacities. These theoretical observations will be reviewed regarding the assumed integration of the addressee's needs. Based on the literature overview, different degrees of interactivity and engagement between speaker and addressee will be distinguished. Then research questions and expectations will be outlined. These research questions investigate the general pattern of referential form as observed in think aloud protocols (experiment 1[217]) and compares it to referential forms observed in instructions (experiment 2). Furthermore, the influence of amount of prior information on referential form in both experiments will be investigated. Following these theoretical sections the coding procedure as well as the annotation scheme will be presented followed by a detailed overview of the results. The chapter will close with a general discussion of the results.

8.1.1.1 Cognitive statuses of referential expressions

Prince (1981) states, that "information packaging reflects the sender's hypotheses about the receiver's assumptions and beliefs and strategies" (Prince 1981:224). Thus, she views text as a "set of instructions from a speaker to a hearer on how to construct a particular discourse model"[218] (Prince 1981:235). She distinguishes three types of givenness that are expressed in referential form, namely Givenness$_P$, Givenness$_S$, and Givenness$_K$. *Givenness$_P$* is based on predictability or recoverability in which the predictions are based on the receiver's knowledge about sentence structure (Prince 1981:226). Here Prince (1981) refers to the different sentence positions of information that is assumed as given and information that is newly introduced. *Givenness$_S$* is based on saliency that is evoked by an entity being in the receiver's consciousness while listening to an utterance. This kind of givenness is expressed by the use of pronouns, including deictics, and syntactically by clause

[217] For a general introduction to the experimental studies see chapter 4.

[218] This model contains entities, attributes, and links between entities. According to Prince (1981), discourse entities are discourse-model objects and they are represented by NPs (Prince 1981:235). But note, that not all NPs are automatically discourse entities.

subjects (see example 4). *Givenness$_K$*, which is based on shared knowledge, signals that the speaker assumes the hearer to know the referent or at least that he/she is able to infer it (Prince 1981:230) (see example (5)).

(4) "We got some beer out of the trunk. **The beer** was warm." (Prince 1981:229; emphasize in the original).

(5) "We got some beer out of the trunk and **it** was warm." (Prince 1981:231; emphasize in the original).

In her theory, Prince (1981) considers syntactical constructions as well as nominal phrases and her givenness classes are defined by their source of information. Gundel et al. (1993) draw upon this theory but extend Prince's (1981) thoughts by proposing a more fine-grained distinction between the different states of givenness and without focusing on the source of information. In their work, they analyzed the distribution of nominal and pronominal references in English, Japanese, Mandarin Chinese, Russian, and Spanish. Based on their findings, they conclude that "different forms of determiners and pronominal forms signal different cognitive statuses" (Gundel et al. 1993:274). Cognitive statuses include information about the location in memory (short- vs. long-term memory) and the attention state. By selecting a specific referential expression, the speaker signals which cognitive status he/she assumes for a referent to be met. Gundel et al. (1993) propose six different *cognitive statuses* (see Table 8.1). These statuses are hierarchically ordered, thus the use of a more restricted status entails the assumption that all lower statuses (to the right in the table) are met. For the addressee these different referential forms provide processing signals and clues about the assumed discourse status of a referent (Gundel et al. 1993:275).

cognitive status	in focus	activated	familiar	uniquely identifiable	referential	type identification
referring expression	*it*	*that* *this* *this* N	*that* N	*the* N	indefinite *this* N	*a* N
Example[219]	it (280)	that dog (279)	that dog next door (278)	the dog next door (277)	this dog next door (277)	a dog (276)

Table 8.1: Givenness hierarchy from most restricted (far left) to least restricted (far right) (Gundel et al. 1993:375).

[219] These examples are taken from Gundel et al. (1993); the numbers in parentheses provide the page numbers of the respective example.

A noun phrase with an indefinite article is the least restricted form because the addressee is assumed to have sufficient semantic knowledge to access a general representation of the nominal expression, i.e. the type, of the entity in question (*type identification*). If the speaker wants to point out a specific object, he/she needs to signal this by using the indefinite *this* with a noun (*referential*). This identification of a specific object is necessary for the appropriate use of all definite expressions (Gundel et al. 1993:276). If the speaker wants to signal that a referent is already represented in the speaker's memory, either because of previous mention in discourse or because the necessary information is sufficiently encoded in the nominal alone, he/she uses the definite article *the* (*uniquely identifiable*). If a very specific referent is assumed to be familiar to the addressee because of a prior representation in long-term memory, this is expressed by the use of personal pronouns and definite demonstratives (*familiar*).

The use of pronominal forms, demonstrative *that*, stressed personal pronouns, and the definite demonstrative determiner *this*, signals that a referent is assumed to be held in short-term memory (*activated*). These "activated representations might have been retrieved from long-term memory, or they may arise from the immediate linguistic or extralinguistic context" (Gundel et al. 1993:278). By the choice of an unstressed pronominal, clitics, or zero pronominals the speaker signals that he/she assumes the referent to be in short-term memory and at the current center of attention (*in focus*) because these forms do not provide sufficient semantic information about the referent by themselves. In general, Gundel et al. (1993) observed that "entities in focus (…) include at least the topic of the preceding utterance as well as any still-relevant higher-order topics" and "membership of the in focus group is partially determined by linguistic form" (Gundel et al. 1993:279) and partially by pragmatic factors. Prince's (1981) familiarity scale can be considered to be one of these pragmatic factors. Furthermore, Fukumura & van Gompel (2012:1292) provide a literature overview of factors that have been shown to foster pronoun usage besides frequent mentioning of the referent in prior discourse or more recent linguistic context. Among those are the referent's animacy, visual context, and similarity between discourse entities. Another concept that is important with respect to this most restrictive cognitive status is Garrod and Sanford's (1982) distinction between *implicit* and *explicit focus*. An entity is in *implicit focus* when it has been mentioned in the discourse whereas it is in *explicit focus* when it is directly relevant to something mentioned and it is presented in the situational scenario.

8.1.1.2 Nominal references – Towards a German Givenness Hierarchy

Gundel et al. (1993) based their givenness hierarchy on the English language thus it was necessary to define a similar categorization for German. Since, to the best of my knowledge, no such hierarchy was proposed for the German language, works that focused on parts of the hierarchy are reviewed.

Bosch and Umbach (2007:50) found that personal pronouns most frequently refer to discourse topics[220] that are most expectable. If expectable means that the referents in this discourse topic are most salient, it seems reasonable to equate 'it' as signaling 'in focus' with German personal pronouns. German demonstrative pronouns were found to mainly relate to discourse-new referents (Bosch & Umbach 2007). Petrova and Solf (2010) presented the German equivalents for the statuses 'in focus' and 'activated'. They highlight that unbound demonstrative pronouns can encode information of both of these two cognitive statuses in German. Based on this finding, it can be argued that unbound demonstrative pronouns signal the status 'activated' but nothing can be said about it being in explicit focus in German. But given the specific experimental context of our study and the rule formulated for the coding process, we can be quite certain that objects are in focus when they are referred to by unbound demonstratives. The status of 'familiarity' is expressed by a demonstrative pronoun used within a nominal phrase. This status can be distinguished from 'uniquely identifiable' in which a definite article is used with a noun phrase. The two statuses of 'referential' and 'type identifiable' cannot be distinguished in German because they are both encoded in nominal phrases with an indefinite article. This is supported by Gundel et al.'s (2010) acknowledgement that

> "not all statuses are encoded by separate lexical items in every language. For example, few languages explicitly encode the distinction between the two lowest statuses, 'referential' and 'type-identifiable'." (Gundel et al. 2010:1773)

Based on this literature a reduced Givenness Hierarchy for German is proposed in Table 8.2. This Givenness Hierarchy will be referred to in the discussion.

[220] Bosch and Umbach (2007:50) define discourse topics to be referents that have been introduced into the discourse before, i.e. they are discourse-old.

cognitive status	in focus	activated	familiar	uniquely identifiable	referential
German expression	Personal-pronomen (personal pronoun), unge-bundenes Demons-trativ-pronomen (demon-strative pronoun)	ungebun-denes Demonstrativ-pronomen (demon-strative pronoun)	gebundenes Demonstrativ-pronomen (demon-strative pronoun)	bestimmter Artikel + NP (definite article + np)	unbestimmter Artikel + NP (indefinite article + np)

Table 8.2: Cognitive statuses German.

8.1.1.3 Nominal specificity

8.1.1.3.1 Definition of nominal specificity

Malt et al. (1999) argue that there are two kinds of categorization of objects, i.e. *knowing* an object versus *naming* an object. *Knowing* an object implies that an object is mentally encoded in terms of 'recognition categories' that are based on perception and representation of objects (Malt et al. 1999:259). *Naming* an object, implies that a name in terms of linguistic categories is chosen for an object. This is only possible if the object as well as its linguistic label is stored as a mental representation in long-term memory. Since naming an object is motivated by the intention to communicate the concept of an object, Malt et al. (1999:260) argue that "linguistic categories (...) may be more complex because of the various mechanisms that influence their composition". According to the authors, the complexity of linguistic categories is based on the process of referring to known object that are similar and the use of conventional terminology. Malt et al. (1999:260) point out that current research investigates the categories that are chosen to describe objects rather than the "process of perceiving, representing, and communicating" about objects. They strongly suggest that future research should study the latter process systematically as well.

Malt et al. (1999) investigated naming and knowing of objects in non-interactive settings. Adopting this approach, Malt et al. (2003) compare their findings regarding the difference between *knowing* an object and *naming* an object to findings on *knowing* and *naming* an action. Again, their study is based on cross-linguistic differences but they were more specifically interested in assessing the influence of language on thought. This investigation is consistent with their differentiation between *knowing* and *naming* as they expect that 'recognition categories' are not influenced by language whereas 'linguistic categories' most naturally are. The findings of the more recent study confirm their previous findings, highlighting that thinking about an object or event in terms of 'recognition categories' is similar across languages whereas 'linguistic categories' differ.

As outlined above, there are different levels of *naming* objects, i.e. refer to them on a descriptive level as opposed to referring to them functionally. It could be argued that references on the descriptive level are linguistic representations of 'recognition categories' and are thus different from 'linguistic categories' that include conventional names etc. If this is assumed as a working hypothesis, it is interesting to study how participants express these two levels of conceptualization linguistically and how they are connected.

The theories that were presented so far investigated references in interactive communicative settings or with an addressee in mind but findings obtained in these studies can be adopted to analyze referential expressions in a setting in which only minimal communicative intention is assumed. Hermann and Deutsch (1974:19) also stress that the process of *naming* is a cognitive process in which the speaker chooses a verbal sign from a repertoire of signs in order to classify an object, person, or action. Thus, they argue that the analysis of object reference allows conclusions about the features that have been classified as relevant for the identification (Hermann & Deutsch 1974:171). Based on the presented literature review, it can be concluded that the nominal phrase carries much information about the level of conceptualization as well as about the cognitive status of a referent.

Mangold-Allwinn et al. (1995) summarize the following features of a noun phrase that encode its specificity:

- definite vs. indefinite article,
- specific modifiers, and
- the lexical specificity of the noun and the object class[221] that it denotes (Mangold-Allwinn et al. 1995:13).

Its taxonomic level can define the specificity of a noun. The following nouns, for example, demonstrate increasing lexical specificity: building, house, semi-detached house. According to Rosch et al.'s (1976) classification, 'building' is a member of the *superordinate level*, 'house' belongs to the *basic level* category, and 'semi-detached' house is a member of the *subordinate level*.

[221] Object class is defined as the amount of objects that can be regarded to be the same (Mangold-Allwinn et al. 1995:116).

Ungerer and Schmid (2006:49) introduce the concept of *cognitive representations* to represent the interaction of a real world object to the mental image of an object. The mental concept of an object is referred to as *cognitive categories*. The authors propose that all cognitive representations that have been encountered are stored in memory as *cognitive models*. These cognitive models are activated when a perceived event including objects is categorized. Thus Ungerer and Schmid (2006:51) argue that one of the characteristics of models is their omnipresence along with their incompleteness. Besides these general characteristics, Ungerer and Schmid (2006) propose that there are differences between cognitive models as stored by individuals because they make different experiences. These different experiences are assumed to be personal as well as collective experiences, i.e. cultural background. It is proposed that cultural models are shared by members of the same social group and represent specific abstractions, i.e. prototypes of events and objects (Ungerer & Schmid 2006:51-53). Based on these assumptions, Ungerer and Schmid (2006) point out that cognitive categories mainly contain prototypical examples as they have

> "the largest number of attributes in common with other members of the category and the smallest number of attributes which also occur with members of neighbouring categories." (Ungerer & Schmid 2006:32)

This idea is already expressed by Brown (1958) in his theoretical observation that

> "the name of a thing, the one that tells what it 'really' is, is the name that constitutes the referent as it needs to be constituted for most purposes. The other names represent possible reconceptualizations useful for one or another purpose." (Brown 1958:17)

The second sentence stresses the influence of context on the conceptualization of an object or concept. Ungerer and Schmid (2006:46) state more specifically that context influences the structure of the individual items with regard to their position in terms of prototypical examples or marginal bad examples for this specific category. The concept of cognitive categories is of interest to the theoretical framework of this thesis because it aims at investigating the linguistic representation of mental processes in general and sepcifically the concept of conceptualization in this chapter. Therefore, Rosch et al.'s (1976) taxonomy of the three levels will be reviewed in reference to Ungerer and Schmid (2006).

Superordinate categories function as collectors of categories highlighting salient attributes shared by all members because they do not contain distinctive features but a large set of categories (Ungerer & Schmid 2006:84). *Basic level* terms encompass the greatest, presorted bundle of shared attributes of specific members. Most of the time, the *basic level* term is associated with the prototypical representation of the term (Ungerer & Schmid 2006:75). Thus, categories can be best distinguished on this level and *basic level* terms are very frequently used in everyday language in order to minimize cognitive effort (Ungerer & Schmid 2006:71). Tversky and Hemenway (1984) argue, that members of the basic level category can be distinguished by their structural components, such as parts. Thus, they propose that "the informativeness of the basic level may originate from the availability of inference from structure to function at that level" (Tversky & Hemenway 1984:169). *Subordinate* categories share parts and are separated by other attributes (Tversky & Hemenway 1984). Those specific attributes are only shared by items of the category in question, and not by any other category which belongs to the same basic level category (Ungerer & Schmid 2006:81). *Subordinate* terms are the second most frequent category in natural language (Ungerer & Schmid 2006:79).

Reviewing numerous studies on reference production by other authors and reporting their own work conducted with various collaborators over the years, Mangold-Allwinn et al. (1995) conclude that nominal specificity is not only a feature of audience design. It is rather influenced by visual perception, expertise and prior-knowledge as well as by aspects of the communicative situation. It was observed that participants who had seen an object for a short period of time produced more basic level terms to refer to them than participants who were allowed to inspect the object for a longer time. With more time for inspection, participants used more subordinate terms. The authors conclude that participants can only perceive the shape of an object during brief investigation whereas they can pay attention to details if they are allowed to spend more time on inspection (Mangold-Allwinn et al. 1995:135). This finding supports the argument that object class specificity can only be as specific as the representation of the object (Mangold-Allwinn et al. 1995:133).

With regard to expertise it is assumed that experts have a richer and more differentiated cognitive representation of the objects that they regularly interact with, which shows in a high frequency of subordinate level terms (Mangold-Allwinn et al. 1995:138). Kiefer et al. (1993), for example, investigated the influence of the nature of prior-knowledge on reference production. They asked students (role: speakers) to either instruct another student (role: addressee) in an assembly task or to describe an assembled object. The speaker's prior-knowledge was varied between two conditions. Half of the participants were shown a video of the assembly process and they were given the time to assemble the object themselves (*action-oriented prior knowledge*)[222]. The other half of the participants was shown a picture of the assembled object (*static prior knowledge*). Results show that participants with *action-oriented prior knowledge* produce more subordinate level terms than participants with static prior knowledge (Kiefer et al. 1993:22). Mangold-Allwinn et al. (1995) argue that participants who interacted with the objects and assembled it themselves created a richer representation of the different parts and the assembly structure. Thus they were able to produce more subordinate nouns than participants with a static representation.

These findings can also be interpreted in relation to Rieser's (1996) concept of *representational metonymy*. Representational metonymy describes the combination between a verbal phrase and a representational relation in order to classify nominal references with regard to their specificity to the goal domain (Rieser 1996:2). The basic idea is that speakers describe object parts in relation to another well-known object in order to facilitate identification of the object at hand, i.e. they re-conceptualize the objects at hand by projecting the features and functions of another object onto them. The concept of *representational metonymy* involves two different levels of conceptualization. The first level is the *description level* that is based on the physical appearance of an object and the conventionally known way to describe it (e.g. "Fünfträger" denotes a bar with five holes (Rieser 1996:15)). The second level, which is called *representational metonymy*, contains the conceptualized description of an object, i.e. the object is described according to its function in the conceptualized target object (e.g. propeller, airplane). Information on the first level is unspecific and perception based. References on the second level are constructed based on world-knowledge, which makes it context specific. Following this line of argumentation, the use of representational metonymy reflects the speaker's contextual representation based on which specific objects are interpreted. It may well be that Kiefer et al. (1993) observed this phenomenon. Participants who were given the time to inspect the object may have conceptualized the object in relation to objects that they know and thus they described the objects by more specific terms on the subordinate level.

Besides investigating the influence of prior-knowledge, Mangold-Allwinn et al. (1995:36) also studied the influence of different aspects of the communicative situation, i.e. communicative goal, addressee, and discourse context, on referential

[222] This terminology is not mentioned in the experimental description by Kiefer et al. (1993:19) but is later introduced in Mangold-Allwinn et al. (1995:141).

form. In their study, discourse context[223] is defined as the position of a reference within the discourse. Based on von Stutterheim et al. (1993), Mangold-Allwinn et al. (1995) distinguish between three discourse times: *introduction*, *referential maintenance*, and *reintroduction*. Furthermore, they varied the communicative goals, either instructing a person to assemble an object or describing an object to a person. They find that people tend to use definite articles in introductions during instructions and indefinite articles in the description scenario. With regard to nominal specificity the results revealed that participants tended to provide as much information as possible in introductory phrases, i.e. subordinate terms were most frequent in this category. In subsequent references, the same trend showed regardless of the communicative goal: pronouns were frequent in referential maintenance and use of definite articles in cases of reintroduction.

Wachsmuth and Jung (1996) investigated reference production in an interactive assembly task. They point out that names and classifications of objects change over time within the assembly process because

> "internal representations are constructed, maintained and restructured along with the changes that occur when assemblies are constructed from a variety of multi-function construction objects." (Wachsmuth & Jung 1996:348)

They refer to two kinds of concepts that drive this dynamic conceptualization process, namely long-term and short-term concepts. *Long-term concepts* contain general information on multi-functional objects such as fixation points or shape and the potential roles of these objects. Additionally, long-term concepts contain information on actions involved in assembly processes and background-knowledge on specific objects such as airplanes (Wachsmuth & Jung 1996:354). *Short-term concepts* hold "temporal conceptualizations of individual entities (objects and aggregates) in the current task environment" (Wachsmuth & Jung 1996:357). The concept of short-term concepts is needed to account for the creation of new concepts when aggregates are built or objects are assigned a function within the goal assembly. This observation of dynamic conceptualization, its representation in language, and the change of referential behavior over time stresses the

[223] Note that this terminology does not seem to be a good choice because the concept 'context' has been used differently in all other studies. In the presented studies 'context' was defined by factors such as participants, communicative intentions, time pressure, written or spoken mode etc. Miller (1996:4) defines context as "the part of the situation (or field) that is used to determine meaning in general, and in particular is used to resolve potential ambiguities of meaning". Von Stutterheim et al. (1993) call this tracing of references over repeated mentioning 'referential movement' (von Stutterheim et al. 1993:107). In this thesis, the terminology 'referential process' will be used to refer to referential change and continuity in repeated mentioning.

importance of analyzing object references throughout the assembly process and not restricting it to initial references.

8.1.1.3.2 Explicit assignment of function to objects[224]

Concerning the function of representational metonymy, Rieser (1995; 1996) alone as well as with colleagues (e.g. Meyer-Fujara & Rieser 2005) studied conceptualization of objects as expressed in referential forms chosen in an interactive setting. In these experiments instructor and assembler could not see each other, thus only verbal communication was possible, but they shared the same workspace. These dialogues highlighted that participants used representational metonymies to facilitate reference to the parts at hand and already assembled aggregates. Two different strategies were identified. First, instructors used metonymies for aggregates which could not be easily described by referring to the individual parts and which had a close resemblance to a real-world object, e.g. ‘propeller’ when referring to two joint bars (Rieser 1996:7). Second, representational metonymy was used when the fixing point of the object part at hand needed to be defined. In this case, the intrinsic orientation of the real-world object was used to indicate the orientation of the aggregate(s) to be worked on. So far the focus has been on the nominal representation of the different layers (e.g. von Stutterheim et al. 1993) or the interactional motivation for function assignment (Rieser 1997) if the phenomenon of conceptual layers in construction tasks was studied.

Instead of analyzing instructions that are collected in an interactive set-up, Daniel and Tversky (2012) analyzed written instructions to a generic addressee on the assembly of a TV stand. In their study, participants assembled the TV stand themselves before writing the instruction. Their analysis revealed different strategies of presenting objects in the introductory phase of the assembly. Some instructors started their instruction by giving a list of objects. Unfortunately, the authors do not specify if the objects are referred to on the descriptive level or with functional terminology. A second strategy for starting the instruction is reported, namely providing information on the task or the goal object. To illustrate the strategy of providing information on the task they give the following example: “the two square-shaped pieces of wood are the sides of the stand” (Daniel & Tversky 2012:306). This example illustrates that at least one participant assigned function at this early state of the instruction by relating the descriptive nominal phrase *the*

[224] This part of the introduction together with the respective parts of the analysis and the discussion were shortened for a paper that is accepted as a talk at the annual meeting of the Cognitive Science Society 2013 in Berlin. The paper with the title ”‘This is a wall’ – Assigning function to objects” is joint work with Thora Tenbrink.

two square-shaped pieces to its function within the TV stand, namely *the sides of the stand*. Although the studies reported by Rieser (1995; 1996) as well as the one reported by Daniel and Tversky (2012) suggest that explicit function assignment happens in instructions it is not addressed in the extensive body of research on the design of instructions, i.e. manuals. This may be the result of the observation that text is accompanied by visuals in the majority of manuals (Maes & Lenting 1999), which makes explicit function assignment unnecessary.

Tenbrink and Seifert (2011) were interested in the linguistic representation of conceptual mapping in a spatial problem solving task analyzing written tour plans for holiday trips. Their study involved two domains of conceptualization. The first domain was that of the road map that was presented to the participants, i.e. the map that was physically present. The second domain was that of the real but unknown holiday destination that was represented by the map. Tenbrink and Seifert (2011) analyzed the linguistic structure of the collected verbal reports in depth focusing on the distribution and nature of nouns, verbs, adjectives, adverbs, and temporal markers. The verb analysis highlighted that 'wollte' (want) and 'könnte' (could) were frequently used to mark mapping processes between the two domains. According to Engel (2002:93), these two modal verbs express intentions to act but with different implications. Whereas 'wollen' (want) expresses the subject's intention, 'können' (can) expresses the possibility given that the subject is granted the permission to act in the described way (Engel 2002:92). The subjunctive mood that was described by Tenbrink and Seifert (2011) adds a tentative character to the statements. Furthermore, their analysis revealed that discourse markers introducing a purpose (i.e. 'in order to') and the particle 'als' (as) also signaled "mapping from plan to purpose" (Tenbrink & Seifert 2011:116).

This review illustrates that although different authors provide examples of function assignment it is not systematically investigated so far, except for those findings reported in Tenbrink and Seifert (2011). Parts of the analysis in this thesis aim at filling this gap.

8.1.2 Different degrees of interactivity and engagement between speaker and addressee

8.1.2.1 Influence of the speaker's cognitive state on audience design in reference production

As outlined above, there are numerous studies supporting the assumption that referential expressions are tailored for the addressee. This is also mirrored in the fact that Gundel et al.'s (1993) proposal of cognitive status is reviewed and discussed in various fields of linguistics (e.g. psycholinguistics (Cornish 2001; Wittek & Tomasello 2005), traditional linguistics (Halmari 1994; Epstein 2011), and sociolinguistics (Meyerhoff 2009)). In 1983, however, Levelt already suggested that

> "the speaker solves the listener's continuation problem[225] by solving his own continuation problem, rather than by keeping a running model of the listener's state of knowledge." (Levelt 1983:100)

This question arose after the analysis of descriptions on visual patterns verbalized for a generic addressee[226]. Levelt (1983) studied the kinds of repair mechanisms, i.e. hesitation markers and interruptions, that are produced in spontaneous speech in order to learn more about the language production system. His data suggests that there is a structural relation between repair and original utterance similar to the structural relation between question and answer[227]. This finding supports the theory that speakers monitor their own speech as if they were their own listeners. Hence, Levelt (1983:97) proposes that "controlling one's own speech is like attending to somebody else's talk", which allows the speaker to apply the same parsing and repair procedures that he/she uses in speaking to others. Based on these theoretical and empirical observations, Levelt (1983) argues further that

> "in listening to somebody else, one normally matches it to the current discourse model, in order to modify or extend the latter. In listening to oneself the matching is with the intended message, and the criterion is identity of intention." (Levelt: 1983:97)

Although he presents evidence for the self-monitoring theory, which he calls 'the perceptual theory of monitoring and repair'[228], Levelt (1983) concludes that he cannot provide a final answer to the question if self- or other-monitoring is predominant at his time of writing.

---

[225] Presumably, Levelt (1983) refers explicitly to the continuation problem because the data that he investigates in his paper has been elicited in order to investigate the question on how speakers organize information for speaking (for more detail see Levelt 1981).

[226] The addressee was generic in the sense that the study design did not involve any physically co-present addressees but rather asked the participants to describe the visual patterns in a way that the listener could draw them by listening to their recorded descriptions. However the listener was specific with regard to background knowledge because the speaker was informed that the listener had some background knowledge about these visual patterns in general.

[227] The following example illustrates this hypothesis very nicely: *"With his sister he talked frequently, uh with his mother he talked frequently"* (Levelt 1983:98). Levelt argues that the repair is the answer to the question *With whom did he talk frequently?*. In order to answer the question properly the preposition *with* is needed and it is also needed to form a correct sentence after the repair signal *uh*.

[228] Levelt (1983) points out that the theory "is perceptual only in that the same parser is involved in understanding an interlocutor's speech and in deriving the message from one's own inner speech" (Levelt 1983:97).

Recent studies support the claim that the speaker's cognitive state is the predominant influence on referential behavior. Horton & Gerrig (2005a), for example, conclude that "speakers will show evidence of audience design to the extent that suitable memory representations become accessible within an appropriate time course" (Horton & Gerrig 2005a:141). This conclusion is based on their findings that "directors[229] mostly used information to the extent that it was readily accessible to them" (Horton & Gerrig 2005a:140). More specifically, Horton and Gerrig (2005b) argue that individuals, who are engaged in a conversation function as cues for retrieving memories that are associated with these people. Given that utterances are produced within a short period of time only those memories that are highly and consistently activated are retrieved (Horton & Gerrig 2005b:10). The speaker assumes that the retrieved information is shared by himself/herself and the interlocutor because he/she associates these memories with this specific individual. This assumption leads to the proposal that common ground is only one potential constraint on message formation besides factors such as lexical and contextual frequency (Horton & Gerrig 2005b:28). To support their claims Horton and Gerrig supplemented experimental studies (e.g. Horton & Gerrig 2005a) with corpus studies of actual telephone conversations[230] (Horton & Gerrig 2005b).

Horton and Gerrig's (2005a;b) proposal is supported by Arnold (2008) who also investigated the influence of production-internal and addressee-oriented processes in audience design. She concludes that referential behavior might seem addressee designed but it may as well result from production-internal processes. As examples of such processes, she mentions subject position for accessible referents in sentences. Even though this may be a strategic move to facilitate understanding for the listener this observation can also be explained by the need to "facilitate production by postponing more difficult words and phrases" (Arnold 2008:508). She also notes that "speaker-internal processing effects are often evident in situations of mental effort" (Arnold 2008:520). As an example, she mentions disfluent speech when referring to 'difficult' objects, such as new or unfamiliar objects without conventional names (Arnold 2008:509). As further evidence Arnold (2008) outlines that speakers intend to produce reference phrases, which

[229] Horton & Gerrig (2005a) asked participants (role: director) to instruct participants (role: matcher) to arrange arrays of 16 cards. Four cards showed the same kind (e.g. flower) but different types of objects. Each director instructed two different matchers. By choosing this design the authors wanted to investigate how directors keep track of the established references and how they make use of this knowledge.

[230] In their study, Horton and Gerrig (2005b:5) investigated a subset of the CallHome American English corpus.

make the intended referent most salient against all other candidates. But the production process is constrained by the speaker's ability to detect all possible competitors within the time that he has for preparing the utterance (Arnold 2008:520). So she concludes, in contrast to Gundel et al. (1993), that referential accessibility is not just a result of discourse status, but rather the result of "both linguistic and nonlinguistic constraints on the attention of speakers and their interlocutors" (Arnold 2008:510).

In this line of thought, Fukumura and van Gompel (2012) present an experiment that is specifically designed to investigate whether speakers base the produced referential form on their own discourse model or if they consider the addressee's discourse model. Speaker and addressee were engaged in an arrangement task in which a pictured scene needed to be represented. In the beginning of each trial the addressee arranged two LEGO toy characters, e.g. a pirate and a mermaid, as presented in a picture. Then speaker and addressee were provided with some information about the pictured scene. Next new information about the scene was provided which initiated a new arrangement of the scene. In order to investigate which discourse model was referred to by the speaker Fukumura and van Gompel (2012) distinguished between two conditions. In one condition the new information was shared between speaker and addressee (*shared context*) and in the other condition it was presented exclusively to the speaker (*privileged context*). Additionally, the referent of the second sentence was varied, i.e. it was either the toy character that needed to be moved or the one that remained in its initial position. Fukumura and van Gompel (2012) expected that speakers who take the addressee's discourse model into account produce more pronouns when the previous information was about the character to be moved in the *shared context* condition than in the *privileged context* condition. If speakers mainly focus on their own discourse model they were expected to produce more pronouns when the previous information was about the character to be moved regardless of the kind of context that was provided. Their data revealed that speaker referred to their own discourse model more frequently than to the addressee's discourse model. Thus, Fukumura and van Gompel (2012) argue that

> "speakers do not routinely use the addressee's discourse model when choosing between pronouns and definite noun phrases, but instead choose referring expressions depending on how accessible the referent is in their own discourse model." (Fukumura & van Gompel 2012:1306)

This argument is very strong and it might be too strong since it allows for the reading that referents that are highly accessible in the speaker's discourse model are always referred to by pronouns. However, these pronominal references can be expected to contain too few information for an addressee to distinguish the referent against other competing objects. This reading would result in great difficulties in referential identification tasks until the addressee asks for clarification. Based on the literature review as a whole, it seems more likely that speakers rely on their own discourse model as well as on the one which they assume for their addressee. Furthermore, the studies show that it is difficult to assess to which degree and under which circumstances one discourse model is more influential regarding referential choice.

8.1.2.2 Referential form in self-oriented speech

Now it could be argued that it is not important to know if a phenomenon is strategically designed for the listener or a product of speaker-internal processes because, in general, it has been observed that listeners use cues such as sentence position (Arnold 2008) or hesitations (Maclay & Osgood 1959; Levelt 1983) to draw conclusions about referents and other discourse entities. Even though this debate may not be of great importance to study the effect of audience design on the flow of communicative situations, it is a very valuable contribution to the question, which referential forms people may produce when they have little communicative intentions.

So far, the focus was on experimental studies that were either conducted within the referential communication paradigm or studies that investigated referential behavior with regard to a distant listener (Levelt 1983). Maes et al. (2004), for example, investigated referential form in written instructions for a distant reader. They observed a general trend of shortening in consecutive references. However, differently than in spoken instructions, they found that instructors had a tendency to be more specific than necessary. They conclude that this may be an effect of the specific context in which instructions for use are consulted. Given the need to perform an action and then consult the text again, more information may help readers to keep track of the necessary objects. These findings seem to reveal a high level of speaker consideration with regard to their addressee's needs.

Summarizing the previously presented studies, the interaction between speaker and listener can be understood as a continuum with full engagement and interactivity on the one end of the continuum and no engagement on the other end. Writing or speaking for a distant addressee can be located in between these two poles because there is no interactivity but there is mental engagement with regard to assumed knowledge, expertise, goal of reading etc. (see Table 8.3 for an overview).

forms of talk	degree of interaction and engagement with another person as addressee	research methods	exemplary studies
dialogue	full interaction and engagement	referential communication paradigm	Krauss & Weinheimer (1966); Clark & Wilkes-Gibbs (1986)
written text, recorded spoken text	no interaction, full engagement	text analysis of transcriptions or written texts (e.g. instructions, letters)	Levelt (1983); Maes et al. (2004)
self-talk, monologue	no interaction and engagement	Discourse analysis of transcriptions of think aloud data or spontaneous speech[231]	Ericsson & Simon (1993)

Table 8.3: Overview of forms of talk and its features with regard to speaker/addressee interactivity and engagement.

The proposed continuum ties in with a proposal by Horton (2009). Based on experimental findings (Horton & Gerrig 2005) and different theoretical positions that are postulated on referential behavior, he proposes the following continuum of degrees of speakers' considerations of their addressees (see Figure 8.1).

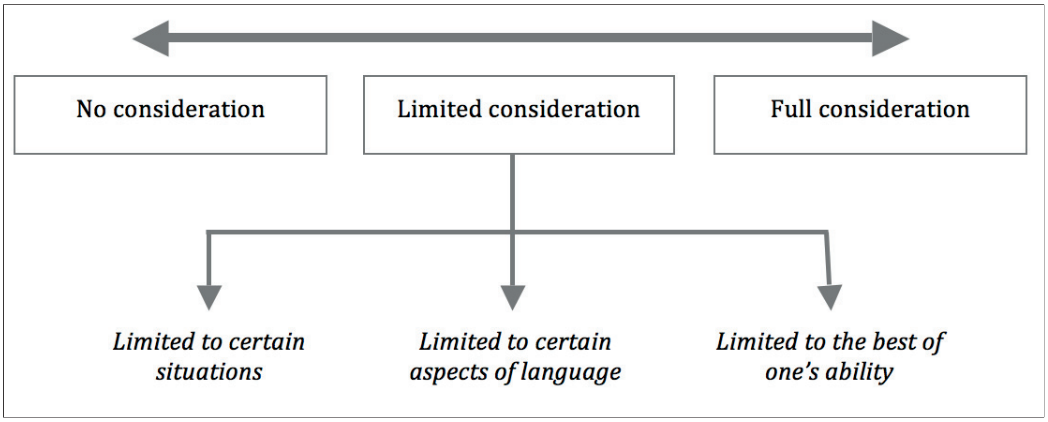


Figure 8.1: A continuum of possibilities describing the extent to which speakers might consider their addressees, including ways in which full consideration might be limited (Horton 2009).

[231] It needs to be pointed out that self-talk is a form of spontaneous speech but spontaneous speech does not necessarily need to be self-talk. This later notion is incorrectly suggested by Clark and Wilkes-Gibbs (1986:36) in their review of literature in which there was no interaction between speaker and listener. Maclay and Osgood's (1959) study is listed as studying monologues but they study transcriptions of a university conference. This speech can be classified as spontaneous in that it is not planned but it is not a monologue because it is directly addressed at the audience and with the expectation for concurrent feedback.

Much has been said about theories that propose that the speaker considers the addressee. Horton (2009) explicitly mentions audience design referring to his own work. Additionally, some literature has been presented which proposes a limited view on the ways in which a speaker can actually consider his/her addressee. But unfortunately, Horton (2009) does not go into detail about the theories that have paid special attention to situations in which speakers are assumed not to consider their addressee's needs. After substantial literature review it needs to be concluded that to the best of my knowledge there are no studies on referential expressions produced in self-oriented speech[232]. Kibrik (2011:62) postulates that "when referring, speakers use various referential devices – nominal expressions that link to concepts in the cognitive representation". Thus, it can be argued that the analysis of referential form produced in think aloud protocols reveals the currently held concepts of a referent.

Assigning meaning to an object – collectively or alone – can be understood as an externalized procedure of conceptualization. In line of this thought previous studies in linguistics have already investigated speaker's conceptualization of objects by presenting participants with pictures of objects and asking them to name those (Rosch et al. 1976; Schmid 1993) or to perform typicality judgment tests (Labov 1973). However, although these studies provide insights on the concepts that are associated with objects, they do not provide any information on the dynamic nature of object conceptualization. So far, the dynamic nature has been studied and described in dialogue only (e.g. Pickering & Garrod 2006; Steels & Loetzsch 2009). The process of referring to an object by the same referential term is called *alignment* (e.g. Pickering & Garrod 2004; Pickering & Garrod 2006). In cases in which no conventional names for an object exists a new reference is negotiated between partners (e.g. Clark & Wilkes-Gibbs 1986).

The data analysis that will be presented in detail in the following sections sheds light on the process of conceptualization of unfamiliar objects that need to be integrated into the structure of a dollhouse. Furthermore, it is proposed that a detailed analysis of multiple references to the same object combined with the analysis of nominal specificity provides a tool to study the speaker's mental conceptualization process. This proposal is based on the assumption that participants who think aloud do not tailor their speech for an addressee. Although there is no way to rule out the possibility that speakers have an addressee in mind,

[232] The phenomenon of talking to oneself has been termed differently: e.g. private speech (Vygotsky 1934) or self-talk (Goffman 1981). In this thesis the terminology 'self-oriented speech' will be used because the orientation is explicitly highlighted. Furthermore, the other terms carry specific associations, e.g. private speech is associated with child speech.

there are a number of reasons that suggest that the intention to communicate is very low in the study on unaided object assembly. First, it can be assumed that the given task reduces the motivation to communicate because it does not pose any explicit communicative task. Second, literature that investigates the influence of speakers' internal processes on reference production, suggests that audience design is only part of reference production. Third, participants are expected to engage in the assembly task so much that speaking out loud accompanies assembling as a subconscious act. The practice session on think aloud before the experiment is assumed to acquaint participants with the task and thus reduce the necessity to concentrate on verbalizing thoughts. Rather than consciously formulating speech, it is expected that think aloud protocols reflect an observation, as reported by Levelt (1983). He suggests that

> "the speaker will try and interpret his own speech in the context of what was previously said by himself (…). He may thus become aware of ambiguity, vagueness, indeterminacy of reference, incoherence etc." (Levelt 1983:97)

Based on these theoretical assumptions, the following sub-section focuses on research questions investigating different aspects of referential behavior in think aloud protocols. The presentation of research questions and expectations will be followed by a description of the steps taken in the analysis including the presentation of the devised annotation scheme (section 8.3). Applying this analytical background the results of referential behavior in self-oriented speech (first experiment: Unaided Object Assembly) and in a setting with communicative intentions but without interactivity (second experiment: Instruction) will be presented in section 8.4. The findings of both experiments will be contrasted and discussed in the final section of this chapter.

8.2 Research questions and expectations

A number of questions arise from the presented literature. After stating the questions for an overview, the expectations regarding each of them are presented in the following paragraphs.

- Research question 1: How do people refer to objects that are not prototypical in nature? How is referential behavior influenced by the two independent variables, i.e. amount of prior information and communicative intention?

- Research question 2: How does reference change over time? How is referential change influenced by the two independent variables?
- Research question 3: At which point in time do participants conceptualize the objects within the goal domain?
- Research question 4: How is explicit function assignment linguistically expressed?

As outlined in detail in chapter 4, there are three conditions in the first experiment on unaided object assembly (hereafter UOA) that vary regarding the provided amount of prior information about the goal state (independent variable: prior information). To recall, in the first condition underspecified information about the goal object is provided, i.e. a sensible object (*underspecified goal condition*). In the second condition verbal information about the goal object is included in the instruction to the participants, i.e. the participants are told to assemble a two-story dollhouse (*verbal goal condition*). In the third condition this verbal information is supplemented by visual information, i.e. participants are shown a picture of the ready assembled dollhouse for 30 seconds (*verbal and visual goal condition*). This experiment does not involve an addressee and no explicit motivation to communicate.

This is different in the second study, referred to as *instructions on object assembly* (hereafter Instructions). In this experiment a distant listener is introduced (for more detail see chapter 4). This difference between experiments introduces the independent variable communicative intention. As will be illustrated in much detail in the following sub-section the dependent variable will be reference form. More specifically, references are analyzed with regard to form (reduced forms vs. nominal phrases and definiteness), nominal specificity, and the referential process.

8.2.1 Research question 1

Research question 1 investigates how people refer to objects that are not typical in nature, i.e. objects that do not suggest a function and that do not have a conventional name. The experiment involves more non-typical objects than conventional ones. The only object that can be easily recognized as a conventional object is the roof-like object because it is red and has a triangular shape (see Figure 8.2).[233] The other objects that were provided to the participants did not suggest any

[233] The upright position might have already cued participants to conceptualize this object part as a roof. It could be assumed that the conceptualization process might be different if it lies on the table in a more unnatural position. This difference was not addressed in the

prototypical function or name (for examples see Figure 8.3 and Figure 8.4[234]). Only the bevels hint at their functional property within a possible structure because they need to be placed beneath the roof to stabilize it. However, in order to draw this conclusion the participant needs to have an idea of possible structures already. Based on Norman's (2002) concept of cultural constraints[235] and von Stutterheim et al.'s (1993) observation on the use of functional terminology, it is assumed that people will use domain specific nominal references, i.e. roof, frequently when referring to the object displayed in Figure 8.2. Assuming that there are no remembered functional terms for referring to the other objects it is expected that participants use *domain unspecific* terms, such as piece and thing, to refer to them.

Beyond this general trend, the independent variables *communicative intention* and the amount of *prior information* are expected to influence the dependent variable nominal specificity. Miller (1966:6) proposes that "instructions provide a context that makes an experimental situation meaningful". As the goal concept 'two-story dollhouse' is explicitly mentioned in the instructions for participants in the *verbal goal condition* it is expected that those participants use more *domain specific* references than participants who do not know anything about the goal object. The picture shown to participants in the *verbal and visual goal condition* provides a specific external model that can be used in subsequent assembly. Miller (1966:7) points out that such a context "can enable a person to restrict the range of possible alternatives". Therefore, it is expected that participants who are given much prior information, i.e. who saw the picture, use more specific nominal references than participants who do not know anything about the nature of the goal object.

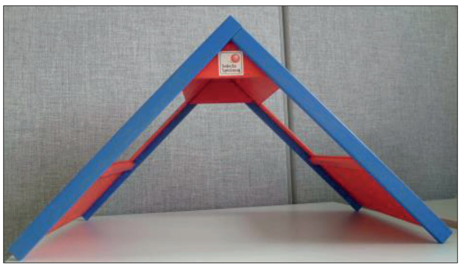


Figure 8.2: Roof piece from the dollhouse. The picture presents it in upright position as it was presented to the participants in the experiment.

present experiments. Nevertheless, one participant tried to use it as a wall but changed this idea later on.

[234] Pictures of all objects are provided in chapter 4.

[235] Norman's (2002) classification of constraints is introduced in chapter 2.

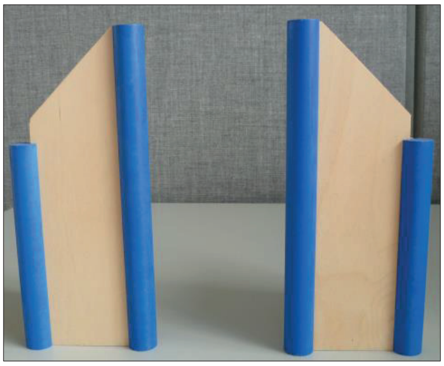


Figure 8.3: Non-prototypical objects that belong in the second story.

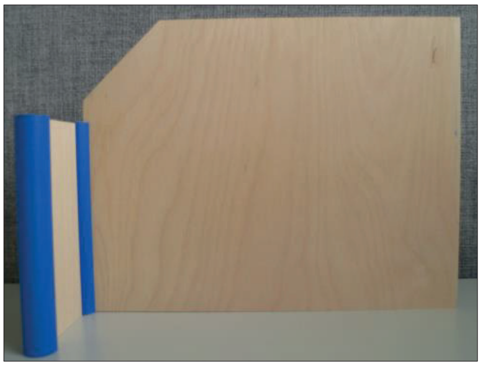


Figure 8.4: Non-prototypical object that belongs in the second story.

8.2.2 Research Question 2

Research question 2 addresses the change of reference over time, referred to as *referential process*. Wachsmuth and Jung (1996) described the dynamic nature of conceptualization in an assembly tasks. Therefore, changes in nominal references are expected to be found as the same object is referred to throughout the assembly. In the literature strategies of shortening in repeated references (Krauss & Weinheimer 1966; Gundel et al. 1993; Maes et al. 2004) as well as syntactical (Bock 1986; Levelt 1983) and lexical (Fukumura & van Gompel 2012) entrainment have been observed in different discourse tasks, such as describing a scene for an addressee or instructing an addressee. Although none of these studies investigated think aloud reports, those are general strategies and thus they are expected to be observed in the analyzed data as well. Generally, subsequent references are assumed to be reduced, i.e. higher frequency of pronouns. It is expected that objects are referred to by a *domain specific* noun at some time within the assembly. If the subsequent reference to that object is not shortened, it is expected that the same specific noun is reused.

The way in which references change is assumed to differ depending on the participant's prior information. Participants who are provided with much prior information (*verbal and visual goal* condition) are expected to use many pronominal forms because they are familiar with the objects right from the start based on the presented external model. Those participants are assumed to remember the overall structure. This knowledge allows them to start incorporating the given object parts much earlier than participants who need to construct the overall structure first. This tendency is expected to show in the use of definite noun phrases[236] as well. The following observations are expected; if participants are provided with much prior information they use definite noun phrases with specific nouns earlier within the assembly than participants who are provided with verbal prior information only. If participants are provided with no specific information (*underspecified goal* condition), they are expected to use *domain specific* nominal references only rarely and late in the assembly process. This difference is expected because those participants need to construct the goal structure first before labeling object parts functionally.

Instructors are assumed to have gained detailed knowledge about the structure of the goal object in the assembly phase[237]. Previous studies showed that instructors who assembled the goal object themselves incorporated the individual parts in the goal structure and label them accordingly more often than instructors who did not assemble the object themselves (von Stutterheim et al. 1993; Kiefer et al. 1993). Therefore, instructors are assumed to be specific in their references by using *domain specific* terminology and definite noun phrases. The referential form is expected to change throughout the instruction. Studies that investigated referential form of introductions of objects in instruction texts revealed that instructors are very specific (Maes et al 2004). According to von Stutterheim et al. (1993:109), instructors assign specific meaning to the objects that are important in the manipulation process at this early state of object conceptualization. Based on these findings, it is expected that instructors will be specific about the object in the introductory phase, i.e. in first reference to the object. Furthermore, it is assumed that instructors will repeatedly use the same references or pronouns after introduction because they assume the objects to be in focus and common ground.

Differences between the two studies are expected to be observed regarding the dependent variable of communicative intention. Participants assembling the object for themselves are expected to use more pronouns than instructors. Given that participants who are engaged in assembly depend on their own discourse model without any need to consider an addressee's discourse model they can theoretically refer to every object by using a pronoun. Instructors need to be certain that their addressee can identify the object in question and thus they need to be specific. A pronominal reference would not be unambiguous given that different objects are in the set of available objects twice.

[236] On the use of definite noun phrases in instructions see von Stutterheim et al. (1993:115).

[237] Fore more detail on the assembly phase and specifically the concept of action-oriented prior knowledge see chapter 4.

8.2.3 Research question 3

Based on the assumed differences in references as outlined above, research question 3, focuses on the identification of the time at which participants thinking aloud use *domain specific* references within the referential process. In the investigation of this research question two different times need to be distinguished, namely *reference process time* and *assembly/instruction time*. The former refers to the number of references (1^{st} – 4^{th}) and the later brings into picture the actual time within the assembly or instruction process. The technicalities of the time scale of the later time will be outlined in the next section on data analysis.

The use of *domain specific* terminology is assumed to highlight the speaker's conceptualization of the specific object into the goal domain. The analysis of think aloud protocols allows to investigate the time of conceptualization. It is assumed that participants conceptualized objects functionally within the goal domain when they refer to them by *domain specific* nouns. Participants who are provided with specific verbal and visual prior information are expected to conceptualize the given object in first or second reference. Participants who are given unspecific prior information (*underspecified goal condition*) will conceptualize the object parts late in the referential process. And participants who are provided with a verbal cue about the goal object (*verbal goal condition*) are expected to be in between those two extremes.

Instructors are expected to conceptualize the objects when they assemble the dollhouse for themselves silently, i.e. in the explorative assembly phase[238]. Thus the time at which instructors use *domain specific* nouns highlights the addressee's familiarity to an object as assumed by the instructor.

8.2.4 Research question 4

In addition to implicit function assignment, i.e. use of a *domain specific* noun from one reference to the next one, participants are also expected to assign function explicitly. Explicit function assignment is defined as instances in which explicit mapping between the descriptive domain of the given object parts and the goal domain is expressed. In a content-based analysis all discourse units that contain a *domain unspecific* object reference as well as a *domain specific* noun were extracted. Examples (6), (7), and (8) exemplify the different ways in which function is assigned to the given objects in the protocols:

[238] For more detail on the different phases involved in the second experiment see chapter 4.

(6) „also **das** werden dann die **Rückwände** für hinten sein“ (“**these** will be the **walls** in the back”).

(7) „das **rote Baustück** nen bisschen wie n **Dach eines Hauses** aussieht“ (“this **red building part** looks a bit like a **roof of a house**”).

(8) „ähm **das** können ja nur **Seitenwände** sein“ (“uhm **these** can only be **walls** that go on the sides”).

Based on first analyses of some mapping phrases (as reported in Tenbrink & Gralla 2009:9-10) a *mapping phrase* can be described as consisting of the reference to a physically present object **x** that is assigned the functional term **y** by a *relational term*. **X** may either be referred to by a deictic expression (examples (6) and (8)) or by a nominal phrase (‘rote Baustück’ in example (7)). The *domain specific* term (**y**) belongs to the semantic filed ‘house’, in the previous examples the participants referred to two kinds of walls and the roof.

The relational term has different forms signaling different states of certainty. Either function is directly assigned by stating that x is y (examples (6) and (8)) or by using comparisons x looks like y (example (7)). In the first kind of mapping the relation is established by the verb ‘sein’ (be). Halliday (1985:113) calls this type of relational clause intensive. It expresses a high level of certainty thus it is called *direct mapping*. The second category of mapping between domains includes different realizations, namely by comparison (examples (7) and (9)), by verbs of *doing and happening* (examples (10), (11), and (13)), or by verbs of *being and having* (example (12)).

(9) „**es** scheint schonmal sowas wie **eine Wand** zu sein“ (this seems to be something like a wall).

(10) „ich muss **das** wohl *als* **Seitenwand** benutzen“ (I probably need to use this as a side wall).

(11) „und **es** *als* **Boden** nehme“ (and take it as the floor).

(12) „und zwar fungieren **sie** dann *als* extra **Miniräume**“ (thus they function as extra tiny rooms).

(13) „**das** stell ich einfach so *als* **Wand**“ (I put this as a wall).

The following verbs are to be found in the category of representational mapping: take as x (example(11)), use as x (example (10)), function as x (example (12)), and put as x (example (13)). Most of these examples contain the preposition ‘als’ (as) which specifies that something (here source entity) represents something else (here domain specific function). This assignment of function is marked by tentativeness because the speaker determines that x is used as y; this implies that x represents y but x is not y. This difference means that x has properties shared with y thus it can function as y but also as something different and it is the speaker who defines the nature of x in the specific context. In order to stress the difference between the two categories of domain mapping this kind is called *representational mapping*.

Preliminary findings (reported in Tenbrink & Gralla 2009:9-10) further suggest that the tentative character of assignment expressed by comparisons can be stressed by using modal verbs in the subjunctive mood to express low certainty ('müsste' (might), 'könnte' (could), 'sollte' (should)). Modal verbs can also be used to stress the certainty of an assignment, for example by using 'müssen' (must) in the indicative mood. Based on findings reported in Tenbrink and Seifert (2011) who identified 'wollte' (want) and 'könnte' (could) as markers of domain mapping it is expected that these modal verbs will be frequently used in mapping phrases. In contrast to Tenbrink and Seifert's (2011) findings, it is expected that the subjunctive mood is not as prominent in the assembly data because all actions can be redone, i.e. all arrangements of objects can be disassembled and reassembled in a different way. Additionally, instructors are expected to assign function by mapping phrases expressing high certainty because they assembled the dollhouse at least two times[239] and thus they are expected to know the function of the individual object parts.

8.3 Data Analysis

Table 8.4 combines the research questions with the analytical tools that will be proposed in this sub-section. This overview of the features of analysis will be theoretically introduced, motivated, and practically outlined in this sub-section.

research question	features of analysis
1.1 How do people refer to objects that are not typical in nature?	nominal specificity
1.2 How is this referential behavior influenced by amount of prior information and communicative intention?	referential form (reduced vs. full noun phrase, nominal specificity)
2.1 How does reference change over time?	referential process
2.2 How is referential change influenced by amount of prior information and communicative intention?	referential process
3. At which point in time do people conceptualize the objects within the goal domain?	nominal specificity combined with the referential process

[239] For more detail on the procedure see chapter 4.

4. How is explicit function assignment linguistically expressed?	mapping phrases

Table 8.4: Summary research questions and features of analysis.

8.3.1 Reference coding

In order to identify the referents of all referential phrases the video recordings of the assembly task were annotated for the first experiment (UOA). For the second experiment (Instructions) the skype recordings were annotated[240]. To collect unconstrained language data experiments were not designed as a referential task and people were not instructed to explicitly refer to all objects or display them when talking about them. Thus the annotation procedure needs to account for the fact that some participants may not refer to all objects or some objects were very rarely referred to by all participants.

The annotations of the video recordings are based on the assumption that attention to an object is the prerequisite for naming it because during the process of attending to an object it is categorized and conceptualized. Brinck (2001) points out that this act of 'attention-focusing' has an intentional character. She argues that "attention is attracted by objects that have an informational (…) impact on the subject" (Brinck 2003:288). Brinck (2003) distinguishes two kinds of motivations for attention, i.e. it is either goal-driven or stimulus-driven. In goal-driven attention, people are actively searching for an object directing their attention to the object when it appears. In stimulus-driven attention the object initiates the focus of attention. Clark (2003) describes that the intention to focus an addressee's attention is displayed by acts of pointing and placing objects in interactions. A study that recorded participants' eye movements during performance of ritualized and novel actions highlighted that in performance of routine tasks, such as brewing tea, people focus on task relevant objects fast without a long orientation phase. Interestingly, Funke and Sperring (2006) report that people did not fixate the hands but the objects that were grasped when performing actions.

This review suggests that there are three ways that signal that an object is in the current focus of attention, i.e. pointing to an object, moving an object, or looking at it for some time. In the assembly task, participants are expected to move and place objects when they focus their attention on them. Additionally, they may look at an object for some time before turning to another object if they did not find a location where they wanted to move it to.

[240] For more details on the procedure in each of the experiments see chapter 4.

Based on this review, it was defined that people focus on an object when they hold it and direct their gaze towards it, or when they are engaged in an action with it, such as lifting or placing it. Since no eye movements were recorded in these studies the focus was on peoples' head movements. Following Funke and Sperring's (2006) report it was assumed that if participants moved their heads towards the object and thereby facing it, people were looking at the object in their hands. This approach has been supported post-hoc by a taxonomy of 'indexical referential behaviour' that Brinck (2012) proposed recently. It consists of four categories, namely individual (visual) attention, individual manual action, shared attentional frame, and ostensive/explicit communication. The category individual manual action is interesting in the context of object assembly because it consists of reaching for an object, grasping/holding an object, and placing an object.

The coding was tested by myself and then performed manually by one trained research assistant. First, each object was assigned a label. There are always two objects that look alike but that can be unambiguously identified by their position and orientation in the goal structure (see Figure 8.5 and Figure 8.6). The annotator needed to keep track of all 13 objects and mentally rotate them in cases in which their label was not obvious right away. While watching the videos the objects that were referred to were marked in the transcripts of the think aloud protocols. The annotator put down the object ID and highlighted the referential phrase.

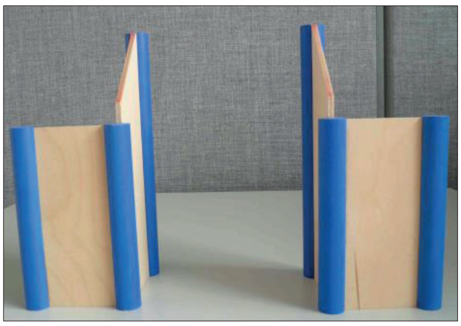


Figure 8.5: Corner pieces on the second floor.

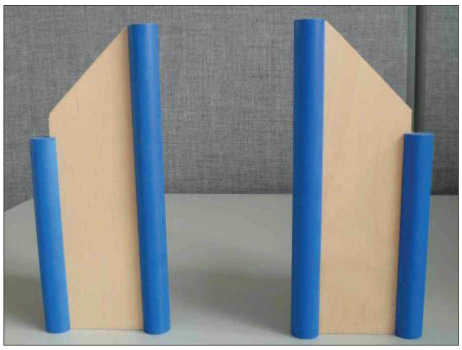


Figure 8.6: Corner pieces on the second floor they way they would be placed on the board.

Besides the technical problems that are outlined in detail in chapter 4 the video was not recorded or incomplete for three participants. Thus, 47 protocols could be coded in the UOA experiment. The coded protocols were almost equally distributed between conditions, i.e. 17 protocols in the underspecified goal condition, and 15 protocols each in the verbal goal condition and verbal and visual goal condition.

As outlined in detail in chapter 4, a comparable number of protocols could be coded in the instruction experiment, namely 16 instructions. The coding was facilitated when participants used the possibility of visual support. Six instructors made use of the visual support and assembled the house for the addressee while speaking, i.e. those participants oriented the house towards the video camera. Four instructors assembled the house for themselves while speaking, i.e. they oriented the rear end of the house towards the camera. In these cases, coding was difficult if participants did not display the objects before placing them. The third strategy was to ignore the camera during the assembly process, i.e. by turning the back on the camera, and showing the ready assembled house to the camera at the end. One participant gave verbal instructions but did not assemble the object in parallel at all. Table 8.5 provides an overview of the number of protocols that could be coded and the objects that were referred to at least once and which could be coded unambiguously[241].

condition	number of coded protocols	number of potential references[242]	number of coded references	% of all possible references
underspecified goal condition	17	221	170	76.9%
verbal goal condition	15	195	154	79.0%
verbal and visual goal condition	15	195	173	88.7%
instructions	16	208	195	93.8%
overall	63	819	692	84.5%

Table 8.5: Overview of protocols and coded references in both experiments.

The figures in Table 8.5 highlight that most references could be coded unambiguously in the instruction experiment. The figures are different for think aloud protocols (UOA). Least references could be annotated unambiguously in the *underspecified goal* condition. On average 81.3% of all references could be coded unambiguously in the experiment on unaided object assembly. This study is explorative in that it collected unconstrained language as opposed to data in referential tasks as reported by Fukumura and van Gompel (2012) for example. Thus the obtained coding percentage cannot be accessed against previous studies

[241] This overview is limited to first mentions because, as will be outlined in the results section, the majority of participants tended to refer to an object only once or twice. Thus first mention will provide an accurate count of all coded objects.

[242] The number of possible object references is calculated by the formula 13 (objects)×the number of protocols in the respective condition.

but with an average of more than 80%, it seems reliably high. The high number of coded references in the instruction scenario already indicates that producing references for an addressee fosters reference to each object. As described in the literature, instructors presented objects to their addressees in order to direct their attention to the object in question. This visual presentation of objects helped the annotator in coding the objects.

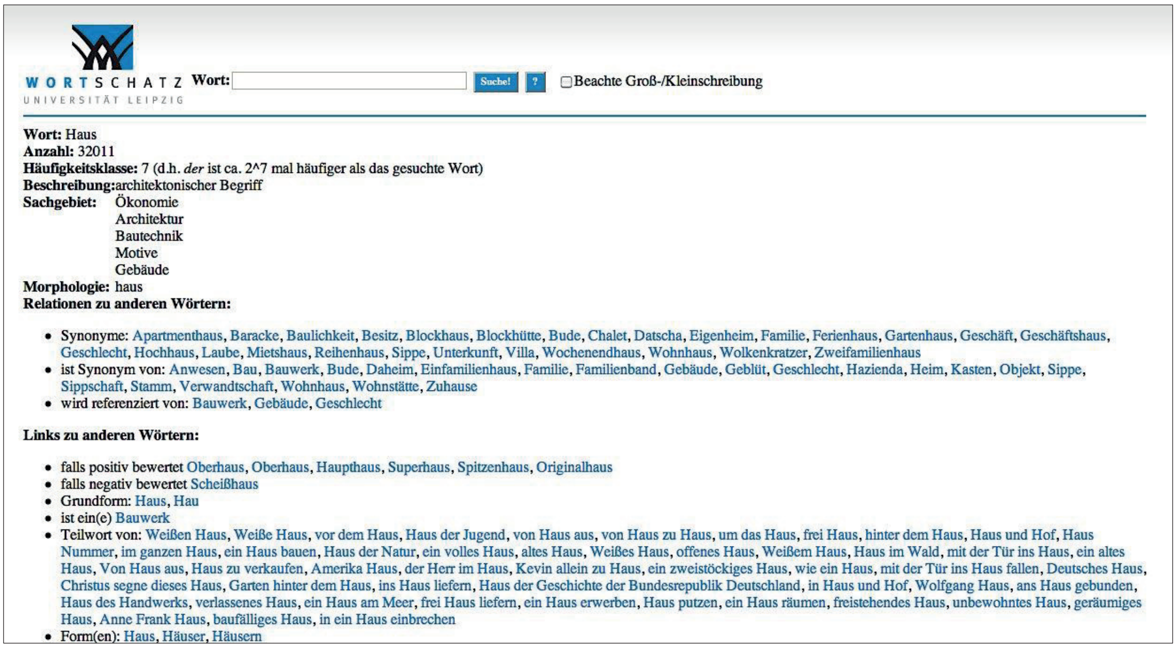
8.3.2 Referential form

The term referential form is defined as the form of the referential phrase that may either be pronominal or nominal phrases. Nominal phrases differ with regard to the article (definite vs. indefinite), the number and kind of modifiers, and nominal specificity. The definition of articles as well as modifiers is based on the German grammar and thus does not need much introduction. But the classification of nominal specificity in the specific context of the dollhouse assembly needs more description. In the next section the theoretical considerations for the classification are outlined before the final annotation categories are introduced. Section 8.3.3 focuses on the analysis of the referential process to study the dynamical change of object references over time. In section 8.3.4 the annotation procedure for explicit function assignment in mapping phrases is presented. In these sections the phenomena described in the literature are extended with regard to the specific nature of object assembly.

8.3.2.1 Nominal specificity

The topic of nominal specificity will not be addressed on the level of object class. Although the distinction between superordinate, basic level, and subordinate terms helps to get a first impression on the specificity of the concept that is expressed by them (as outlined in section 8.1.1.2) the interest was on a more fine-grained distinction. In order to highlight the concepts that are developed over the assembly process it is important to investigate nouns with regard to their specificity relating to the conceptual target domain. Thus the approach outlined in this thesis applies the framework of *representational metonymy* (Rieser 1997; Meyer-Fujara & Rieser 2005). As outlined in the first section of this chapter, the framework of *representational metonymy* distinguishes between two levels, i.e. the description level and representational metonymy. Terms that belong to the first level contain information on physical appearance; thus the information is unspecific with regard to the specific context. *Representational metonymy* contains conceptualized descriptions of an object, i.e. the object is described according to its function in the conceptualized target object. Those references are constructed based on background knowledge and are thus context specific.

In both experiments, the assembly and the instruction task, the goal concept is a house, specifically a two-story dollhouse. The concept dollhouse is reduced to the basic level term house because in the task no context is provided which would cue participants for the function 'dollhouse'. This can be expected to be different if dolls or small furniture would have been provided as well. Furthermore, the Wortschatzportal on German language was consulted. The 'Deutsche Wortschatzportal' provides detailed entries (e.g. frequency, description, domain, morphology, synonyms, forms, subordinate terms etc.) for German words (see Figure 8.7 for an impression). This information is automatically gathered from publicly available sources. The project was initiated and is supervised by the University of Leipzig.


WORTSCHATZ UNIVERSITÄT LEIPZIG Wort: Suche! ? Beachte Groß-/Kleinschreibung

Wort: Haus
Anzahl: 32011
Häufigkeitsklasse: 7 (d.h. *der* ist ca. 2^7 mal häufiger als das gesuchte Wort)
Beschreibung: architektonischer Begriff
Sachgebiet: Ökonomie
Architektur
Bautechnik
Motive
Gebäude
Morphologie: haus
Relationen zu anderen Wörtern:

- Synonyme: Apartmenthaus, Baracke, Baulichkeit, Besitz, Blockhaus, Blockhütte, Bude, Chalet, Datscha, Eigenheim, Familie, Ferienhaus, Gartenhaus, Geschäft, Geschäftshaus, Geschlecht, Hochhaus, Laube, Mietshaus, Reihenhaus, Sippe, Unterkunft, Villa, Wochenendhaus, Wohnhaus, Wolkenkratzer, Zweifamilienhaus
- ist Synonym von: Anwesen, Bau, Bauwerk, Bude, Daheim, Einfamilienhaus, Familie, Familienband, Gebäude, Geblüt, Geschlecht, Hazienda, Heim, Kasten, Objekt, Sippe, Sippschaft, Stamm, Verwandtschaft, Wohnhaus, Wohnstätte, Zuhause
- wird referenziert von: Bauwerk, Gebäude, Geschlecht

Links zu anderen Wörtern:

- falls positiv bewertet Oberhaus, Oberhaus, Haupthaus, Superhaus, Spitzenhaus, Originalhaus
- falls negativ bewertet Scheißhaus
- Grundform: Haus, Hau
- ist ein(e) Bauwerk
- Teilwort von: Weißen Haus, Weiße Haus, vor dem Haus, Haus der Jugend, von Haus aus, von Haus zu Haus, um das Haus, frei Haus, hinter dem Haus, Haus und Hof, Haus Nummer, im ganzen Haus, ein Haus bauen, Haus der Natur, ein volles Haus, altes Haus, Weißes Haus, offenes Haus, Weißem Haus, Haus im Wald, mit der Tür ins Haus, ein altes Haus, Von Haus aus, Haus zu verkaufen, Amerika Haus, der Herr im Haus, Kevin allein zu Haus, ein zweistöckiges Haus, wie ein Haus, mit der Tür ins Haus fallen, Deutsches Haus, Christus segne dieses Haus, Garten hinter dem Haus, ins Haus liefern, Haus der Geschichte der Bundesrepublik Deutschland, in Haus und Hof, Wolfgang Haus, ans Haus gebunden, Haus des Handwerks, verlassenes Haus, ein Haus am Meer, frei Haus liefern, ein Haus erwerben, Haus putzen, ein Haus räumen, freistehendes Haus, unbewohntes Haus, geräumiges Haus, Anne Frank Haus, baufälliges Haus, in ein Haus einbrechen
- Form(en): Haus, Häuser, Häusern

Figure 8.7: Screenshot of the entry *Haus* in the 'Wortschatzportal'[243].

The 'Wortschatzportal' defines a dollhouse to be a house. Based on this finding and the missing contextual clues for a dollhouse, the categorization of the identified nouns was based on the functional concept house. Two categories have been defined: *domain specific* terminology[244] and *domain unspecific* terminology. Furthermore, a third category has been introduced called *constraints* containing

[243] Available at: http://wortschatz.uni-leipzig.de/abfrage/; 24.10.2012.

[244] This category is equivalent to the concept of representational metonymy but this terminology is not adapted to highlight the distinction between the two categories – domain specific vs. domain unspecific – at first glance. It is also intended to foster understanding of the two distinct categories for readers who are unfamiliar to the framework of representational metonymy.

all nouns that referred to constraints that were induced by the task or the objects (for more detail on the notion of constraints see chapter 9).

Among the various nouns that were used to describe the objects, some could be classified easily. Among the clear candidates for the category *domain specific* were 'Etage' (story), 'Wand' (wall), 'Dach' (roof), 'Decke' (ceiling), or 'Fenster' (window). Among the clear candidates for the category *domain unspecific* were 'Ding' (thing) and all its synonyms, 'Bretter' (boards), or 'Gerüst' (gantry). However, there were also cases in which it was difficult to decide if it was specific or unspecific, such as 'Ebene'[245]. If it is assumed to be used within the conceptual frame of house, it is a synonym for 'story'. However, it can also denote the orientation (horizontal) or surface (smooth) of an object which would not be *domain specific* but part of a general conceptual frame. In this case it needs to be classified as unspecific.

After categorizing parts of the transcribed nouns intuitively the nouns classified as *domain specific* have been checked against the 'Deutsche Wortschatzportal' entries and the architectural ontology called IfcBuildingElements which was more specific in some respects including, for example, 'column' (Pfeiler) and 'railing' (Geländer) (see Figure 8.8). Furthermore, nouns that were compositions and contained house specific terms were classified as *domain specific*.

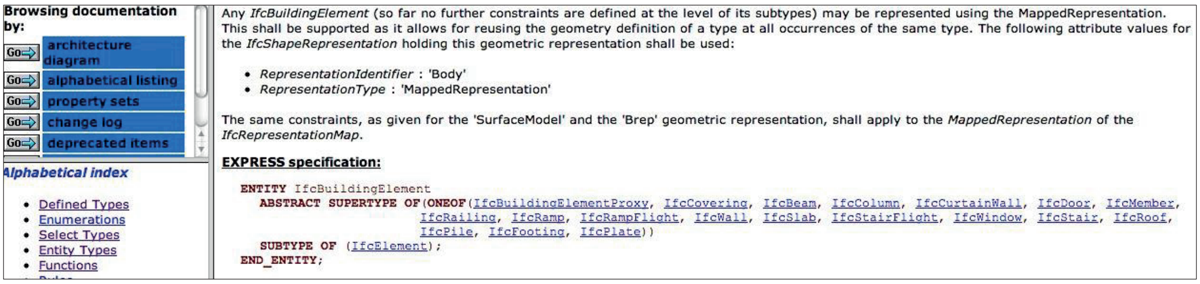


Figure 8.8: Screenshot of the entry *house* in the architectural ontology IfcBuildingElement[246].

Given both of these sources there were still 28 nouns for which it was difficult to find the appropriate category. Those 28 nouns were presented to 57 university students with the task to rate each word with regard to its specificity for the concept house (the original task is to be found in Figure 8.20 in the appendix). The scale was from 1 (very specific) to 6 (not specific at all). This scale allows for three equally well-defined intervals, i.e. 1,0-2,0 means specific with regard to the concept, 2,1-3,9 indicates an undecided category, and 4,0-6,0 indicates that an

---

[245] It is already difficult to translate this word because there is not one exact translation for it but different words for the different meanings, which it has in German.

[246] Available at: http://www.buildingsmart-tech.org/ifc/IFC2x3/TC1/html/ifcproductextension/lexical/ifcbuildingelement.htm#definition; 08.12.2009.

object is unspecific with regard to a 'house'. No fillers have been used because it was important to find out about their categorization of the respective objects but it was not important to distract them. Only those nouns that have been rated 2,0 or better were included in the category house-specific; these were 'Garten' (garden), 'Möbel' (furniture), 'Dachziegel' (tiles), and 'Inneneinrichtung' (interior of the house). This test helped to highlight nouns that are not typically house-specific and contributed to an objective categorization. Importantly, the presented words have been used with different frequency, e.g. 'Garten' is used by two participants, whereas 'Seite(n)' (side) has been mentioned by 28 participants. An overview of the categories and their members is provided in Table 8.34 in the appendix.

Members of the third category labeled 'constraints' were also difficult to define. For a broad orientation the categories defined by Norman (2002) have been used, namely physical, semantic, cultural, and logical constraints. During the analysis of the assembly process in terms of problem solving processes it became evident that the label 'constraints' was too specific because participants did not have to adhere to these constraints in order to build a house. Certainly some constraints were more binding such as physical constraints but some could also be disregarded without any negative consequences. Hence, the category was rephrased to *constraints and perceptual object features*. In some parts of the analysis this category will be collapsed with the category domain-unspecific nouns because the members of this class were not house specific but rather object and task specific.

8.3.2.2 Coding scheme: referential form

In German definiteness is encoded in definite articles, demonstrative pronouns, and possessive pronouns. Indefiniteness is encoded by the indefinite article and further indefinites, such as 'all', 'some', 'much', 'one' (Engel 2002:118). Additional markers of indefiniteness, which are found in the data, are 'irgendein' (any) and 'irgendwelche' (anything), and 'jedes' (each) (Erben 1972:218). Furthermore, Prince (1992:299) classifies quantifiers, such as numerals, to mark indefiniteness in English. This was adopted for German in the presented analysis. Thus, the two categories definite and indefinite noun phrase could be distinguished. In German plural nouns are only used with an article in cases in which a definite article would be used in the singular. In all other cases it is optional, i.e. indefinite references in the plural are used without an article (Engel 2002:122). Cases in which participants used a definite article with a plural noun (e.g. 'diese Teile' (those parts)) were annotated as definite nominal phrases. Consequently, all references that were coded as plural with no article were classified as indefinite references.

Combining all the presented detail into one coding scheme resulted in a number of categories. Table 8.6 provides an overview of the annotated features.

feature	variables	examples
nominal specificity	domain specific, domain unspecific, constraints and perceptual object features	Wand (wall) Teil (thing) Löcher (holes)
articles	1) definite, 2) indefinite, 3) numerical 4) 0 article	1) die Wand (the wall) 3) eine Wand (one walls)[247] 4) _ Wand (_ wall)[248]
grammatical number	1) singular 2) plural	1) die Wand (the wall) 2) (die) Wände ((the) walls)[249]
quantifiers	indefinite determiners[250]	alle Sachen (all things), keine Einkerbung (no groove)
numerals	numbers cardinal numbers	eins, zwei (one, two) erste Etage (first floor)
pronouns	pronouns demonstratives	a) der, die, das b) dieser, diese, dieses

Table 8.6: Overview of the annotated features on referential form.

8.3.3 Referential process: maintenance or change of conceptual state

As outlined above, references to the same object change over time. These changes might result from re-negotiations in conversation (Clark & Wilkes-Gibbs 1986) or re-conceptualization of objects (Wachsmuth & Jung 1996). In the literature four forms of referential change have been described: lexical entrainment, reduction, re-mention, and re-conceptualization. There may be no change but rather the repeated use of the negotiated referring expression for an item (Brennan & Clark 1996; Horton & Gerrig 2002, 2005). *Lexical entrainment* is a direct measure of

[247] Problems arose during the annotation process because 'eine' encodes the indefinite article (a) but at the same time the numeral (one). These two meanings cannot be differentiated based on the verbal data. Since there are numerous object parts and five of them even twice I concluded that the likelihood is very high for 'eine' to encode the numeral. Hence all instances of 'eine' have been coded as numerals.

[248] In these cases participants omitted the article which was needed.

[249] In German, the use of articles is optional in plural nominal phrases (Engel 2002:122).

[250] Classification according to Erben (1972).

referential continuity. However, there are also forms of change. First, *shortening* across references may either be on the semantic level by dropping modifiers (e.g. Clark & Wilkes-Gibbs 1986) or on the lexical level by use of pronominal forms in subsequent sentences or utterances. This strategy has been described as an indication for a referent to be in focus (Gundel et al. 1993) and as an indication of changed conceptualization (Horton & Gerrig 2002). Second, people have been observed to depart from a previous reference by adding new information or reducing information (Horton & Gerrig 2002). Third, people may offer a completely different conceptualization of the same object across trials. This is defined as reconceptualization (Wachsmuth & Jung 1995; Horton & Gerrig 2002). This later process can be investigated by studying nominal specificity, hence this process is called *reframing* in the analysis. Change of referential form by adding or reducing information is disentangled in the proposed analysis. Instances in which the nominal reference is changed but within the same conceptual domain are called *rephrasing*. Instances of reduced nominal references are annotated as belonging to the category *reducing*.

In order to study change in referential form, von Stutterheim et al. (1993:102) introduced three temporal points within discourse[251]. They argued that referential movement could be studied by looking at referential form in the introduction, at times of maintenance, and in re-mention. Maintenance was defined as "repeated reference to an object in the immediately following utterance" (von Stutterheim et al. 1993:108). The proposed temporal sequences were adopted by Kiefer et al. (1993). Although this approach provides an insight on the change of referential form this insight is only partial if the analysis is restricted to two references to the same object. Therefore, all references to an object up to eight times of mentioning were annotated in this study. This approach provides information on which objects were frequently referred to and which ones were mentioned only once or twice.

The terminology 'referential movement', as proposed by von Stutterheim et al. (1993), suggests activity on the side of the referential term that is not the case. Thus the term *referential process* is proposed to highlight the dynamic nature but to stress that it is passive on the side of references as it is the speaker/writer who is actively creating referential terms. The coded references were annotated with regard to change from the previous to the current form; eight categories were identified as outlined in Table 8.7.

[251] In von Stutterheim et al.'s (1993) study those are description and instruction texts.

category	description	change in conceptual state	example
no reference	the object has not been referred to at all		
single reference	the object has been referred to only once		
reducing	the previous form has been reduced either to a pronoun or by dropping modifiers	change to familiarity	das Teil – das (the piece – it)
reframing	the previously domain-unspecific term is now conceptualized within the goal frame, i.e. house	big change	ein Teil – eine Wand (one piece – one wall)
repeating	the previous form is repeated unchanged	no change	das Teil – das Teil (the piece – the piece)
rephrasing	a new form is introduced that is within the same domain	slight change	die Platten – dieses Ding (the boards – this thing)
reusing	taking up a previously used referential form	slight change	das Dach – das – das Dach (the roof – it – the roof)

Table 8.7: Coding categories for referential process.

In Table 8.7 an additional dimension is introduced, namely *change in conceptual state*. Since this thesis investigates the change of mental concepts over time and their representation in language it was important to make predictions about the relation between a change in lexical form and the change of mental concepts. It is proposed that the five categories of referential process resemble a continuum of change in conceptual state, i.e. from *repeating* to *reframing*. If references are simply *repeated*, there is no conceptual change. If references are *rephrased* or *reused*, there is a slight conceptual change. If people choose a different noun but from the same domain, the object may either be conceptualized a little differently or people may not remember what they have called the object before. If a term is reintroduced that has been used before a slight change can be observed in that participants recall the previously assigned noun and substitute the latest one by it. In this case the argument for a failed memory may hold as well if it is assumed that the person cannot recall the assigned name but the first one which may possibly be more conventional.

By *reducing* a noun phrase to pronominal form or by dropping additional information, which is encoded in modifiers, a big conceptual change is indicated. This change signals that an object is familiar and in the case of reduction to pronouns it also signals that the object is in focus. The biggest change, nevertheless, is assumed to be represented in the category *reframing* because in these cases the re-conceptualization of an unspecific object to a functional object within the goal domain is explicitly marked by the feature of nominal specificity.

Before presenting the results, it needs to be noted that the annotations reflect the change of referential form over time but they do not contain any information on actual time, i.e. the assembly time. Specifically with regard to *reframing*, this additional information is important to capture the actual time of re-conceptualization, i.e. at which time within the problem solving process. For capturing the dimension of *assembly time*, the participants' mention of instances of reframing was related to the protocol structure at large in an in-depth analysis. As summarized in chapter 5.3.1, there was a great difference in the amount of time that people needed to finish the assembly and give instructions. Comparing the number of annotated discourse units instead of performance time can standardize this difference. Thus the protocols, that resemble the assembly process, were divided into four quarters to structure the assembly process. The respective instance of *reframing* was located in one of these quarters and coded accordingly.

8.3.4 Functional mapping phrases

Based on the outlined expectations, the annotation scheme of functional mapping phrases contained the following features:

- nature of the source entity (exophoric, i.e. deictic, nominal, or missing (ellipses))
- mapping category (*direct mapping* or *representational mapping*)
- modal verbs in connection with main verbs (see *Figure 8.9*).

TN	information unit	source entity			target entity	mapping											
number	content	exopho	ellipses	noun		comparison (wie/		modal			relational verb					Ellipse	sp
						lemma		lem	Indi	Sub	lem	Prä	Ver	Fut	Sub	...ctiv	
8	9	gut das warn irgendwie zwei Etagen [?] das Dach	das			das Dach					sein		1				
8	45	also das werden dann die Rückwände für hinten sein	das			die Rückwände					sein			1			
8	47	so das ist die Rückwand	das			die Rückwand					sein	1					
8	62	das müsste schon die unterste Etage gewesen sein	das			die Etage		müssen		1	sein		1				

Figure 8.9: Screenshot of the annotation scheme for mapping between functional and goal domain.

This annotation scheme worked well for self-assembly data. However, the content-based analysis of mapping phrases in instructions (study 2) revealed that instructors used a third strategy. More specifically, they used a more implicit way of assigning function as illustrated in example (14).

> (14)„jeweils mit einem kurzen Stück einem ähm einem Wandelement"
> ("each with a **small piece** a uhm a **wall element**").

In this example, the source entity (a small piece) is assigned its function (a wall element) by the act of renaming after a hesitation. This third kind of mapping is more implicit than the previously introduced kinds. Drawing on the terminology that has been introduced in this chapter regarding the specific kinds of referential process, this type of mapping is called *reframing* because it represents the change from an unspecific term to a goal specific terminology.

In the following section, the results of the analysis of nominal form, referential process, and mapping phrases are presented. In section 8.4.1 the results regarding the different features of the referential form are outlined. Section 8.4.2 focuses on the results of the referential process throughout the assembly and the instruction. In section 8.4.3 the results regarding the analysis of mapping phrases are presented. In all of these sections the results on the first study, i.e. UOA are presented first. In this study the influence of the independent variable *prior information* is investigated. Then the results on instructions are provided. Based on these findings a comparison is made between results obtained in instructions to those obtained in the *verbal and visual goal* condition to study the influence of communicative intention[252]. After the presentation of the results, a final discussion concludes this chapter of analysis.

8.4 Results

The first sub-section provides a general overview on the number of references to objects throughout the assembly and the instruction. In the second sub-section, the results concerning referential form will be presented. As outlined in the analysis section, there were different aspects of referential form that were distinguished thus this sub-section is further divided. First, the analysis of nominal and pronominal phrases is presented (8.4.1.1). Second, the pronominal phrases are further distinguished into definite and indefinite ones (8.4.1.2). Third nouns are

[252] For a detailed discussion of the independent variable communicative intention see chapter 4.

differentiated into *domain specific* and *domain unspecific* nominal references (8.4.1.3). Then the results regarding referential change are summarized (8.4.2). Finally, the results on the mapping phrases are reported (8.4.3).

8.4.1 Referential form[253]

This presentation shall start with a technical note: the analysis can either be done on a by subject or a by object count. In a by subject count it is analyzed by which form one specific addressee referred to the objects. In the by object count, , it is evaluated which object was referred to in which way without distinguishing between individual participants. The results reported in this section, i.e. in the graphs and in the tables[254], are based on by subject counts because only this analysis allows to distinguish between conditions in the unaided assembly experiment (UOA). These findings were compared to the by object results in the statistical analysis and the findings were the same.

8.4.1.1 Number of references to objects throughout the assembly

Table 8.8 summarizes the raw frequency of references throughout the assembly process. The results revealed that most participants referred to an object only once or twice. Participants in the *verbal and visual goal* condition referred to objects more than twice most often. Participants in the *underspecified goal* and *verbal goal* condition referred to an object for a fourth time in only 14.9% and 13.3% of all cases respectively. Participants in the *verbal and visual goal* condition referred to an object a fourth time in 18.5% of the cases. As these numbers were fairly close and higher than those for fifth to eighth mention were, the analysis in the following sections is restricted to first to fourth mention.

More than one quarter (26.2%) of the participants who were provided with few prior information (*underspecified goal condition*) did not refer to an object once. This number was twice as large as the one for participants in the *verbal and visual goal* condition (12.8%). The difference was found to be statistically significant (Tukey $p = .016$; LSD $p = .003$).

---

[253] Preliminary analyses of a smaller data set were run in collaboration with Thora Tenbrink and Elena Andonova. Although the data sample was extended to the whole corpus and the analysis was further refined, I want to greatly acknowledge their contribution.

[254] Most of the tables are to be found in the appendix on this chapter.

condition	reference number	reference[255]	no reference	ellipsis
underspecified goal	1st	163 (73.76%)	58 (26.2%)	0 (0.0%)
underspecified goal	2nd	102 (46.15%)	119 (53.8%)	0 (0.0%)
underspecified goal	3rd	58 (26.24%)	162 (73.3%)	1 (0.5%)
underspecified goal	4th	33 (14.93%)	188 (85.1%)	0 (0.0%)
underspecified goal	5th	21 (9.50%)	200 (90.5%)	0 (0.0%)
underspecified goal	6th	13 (5.88%)	208 (94.5%)	0 (0.0%)
underspecified goal	7th	7 (3.17%)	214 (96.8%)	0 (0.0%)
underspecified goal	8th	3 (1.36%)	218 (98.6%)	0 (0.0%)
verbal goal	1st	152 (77.95%)	43 (22.1%)	0 (0.0%)
verbal goal	2nd	96 (49.23%)	99 (50.8%)	0 (0.0%)
verbal goal	3rd	57 (29.23%)	138 (70.8%)	0 (0.0%)
verbal goal	4th	26 (13.33%)	169 (86.7%)	0 (0.0%)
verbal goal	5th	13 (6.67%)	182 (93.3%)	0 (0.0%)
verbal goal	6th	9 (4.62%)	186 (95,4%)	0 (0.0%)
verbal goal	7th	6 (3.08%)	189 (96,9%)	0 (0.0%)
verbal goal	8th	4 (2.05%)	191 (97.9%)	0 (0.0%)
verbal and visual goal	1st	170 (87.18%)	25 (12.8%)	0 (0.0%)
verbal and visual goal	2nd	128 (65.64%)	67 (34.4%)	0 (0.0%)
verbal and visual goal	3rd	75 (38.46%)	120 (61.5%)	0 (0.0%)
verbal and visual goal	4th	36 (18.46%)	159 (81.5%)	0 (0.0%)
verbal and visual goal	5th	25 (12.82%)	172 (88.2%)	0 (0.0%)
verbal and visual goal	6th	11 (5.64%)	184 (94.4%)	0 (0.0%)
verbal and visual goal	7th	3 (1.54%)	192 (98.5%)	0 (0.0%)
verbal and visual goal	8th	1 (0.51%)	194 (99.5%)	0 (0.0%)

Table 8.8: Results on references to objects from one reference to eight references throughout the assembly; raw frequency with percentage in brackets.

8.4.1.2 Feature: nominal and pronominal phrase

- Study 1: UOA

If first to fourth reference were analyzed as a combined set of data, a significant effect of amount of prior information on pronominal and nominal forms could be observed, $L\chi^2$ (8, N = 2444) = 49.5, p = .00. Participants in the *underspecified goal*

[255] References contain nominal phrases, pronominal phrases, and numerals. A more detailed table distinguishing between these forms is provided in the appendix (table 115).

condition (138 cases) used pronominal forms significantly less frequently than participants in the *verbal and visual goal* condition (208 cases). Additionally, participants in the *verbal and visual goal* condition had significantly fewer instances of 'no reference' (see Table 8.9). Participants in the *verbal goal* condition used numerals significantly more often than participants in the other conditions (see Table 8.10).

condition	nominal phrase	pronominal phrase	no reference	numerals	ellipsis
underspecified goal	214	138	527	4	1
verbal goal	186	135	449	10	0
verbal and visual goal	199	208	317	2	0

Table 8.9: Distribution of nominal and pronominal forms within conditions in UOA, first to fourth reference; raw frequency.

condition	nominal phrase	pronominal phrase	no reference	numerals	ellipsis
underspecified goal	-0.2	-2.7	1.8	-0.7	1.1
verbal goal	-0.4	-1.5	0.9	2.2	-0.6
verbal and visual goal	0.6	4.4	-2.8	-1.4	-0.6

Table 8.10: Distribution of nominal and pronominal forms within conditions in UOA, first to fourth reference; standardized residuals.

The second part of the analysis investigated if the influence of amount of prior information on nominal or pronominal form could be observed in each of the four references individually. In 1st reference, the distribution of nominal and pronominal forms was similar in the three assembly conditions. Participants tended to use nominal phrases more frequently than pronouns (see Table 8.11 and Figure 8.10).

condition	noun phrase	pronoun	no reference	others
underspecified goal	49.78	22.62	26.24	1.36
verbal goal	51.28	22.56	25.64	0.51
verbal and visual goal	50.77	37.44	11.28	0.51

Table 8.11: Distribution of nominal and pronominal references within conditions in UOA in first reference; mean percentage.

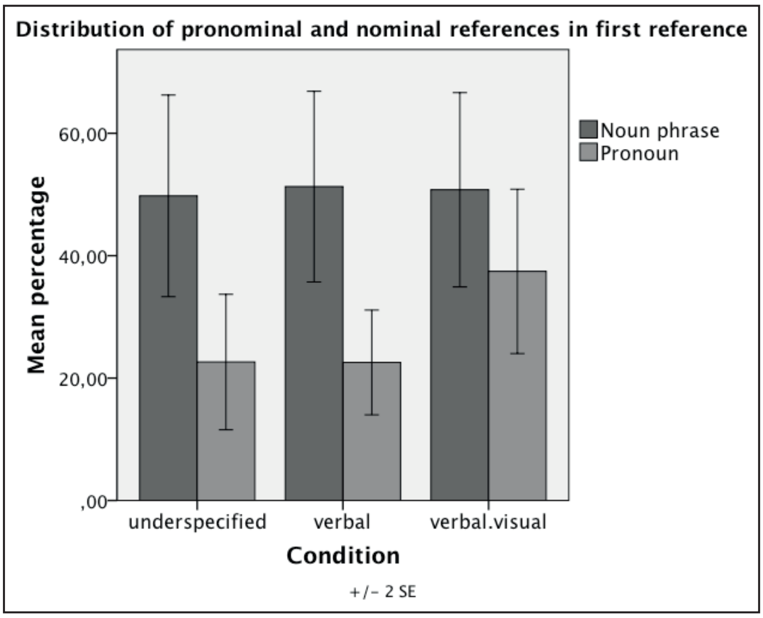


Figure 8.10: Distribution of pronominal and nominal references in first reference within conditions in UOA; mean percentage with standard error.

The difference between conditions concerning the use of pronouns got more marked in 2nd reference. Participants in the *verbal and visual goal* condition (37.4%) used pronominal forms more frequently than participants in the *underspecified goal* condition (19.5%) (see Table 8.12). However, this trend did not reach statistical significance; $F(2,44) = 2.58$, $p = .09$.

condition	noun phrase	pronoun	no reference
underspecified goal	26.39	19.53	54.07
verbal goal	21.53	27.18	51.28
verbal and visual goal	28.20	37.44	34.36

Table 8.12: Distribution of nominal and pronominal references within conditions in UOA in second reference; mean percentage.

There was no significant influence of amount of prior information on nominal and pronominal choice in 3rd and 4th reference (see Table 8.35 in the appendix for all results).

- <u>Study 2: Instructions</u>

Instructors had a clear tendency for using nominal phrases; 87.5% nominal references as compared to 5.3% pronominal references in 1st reference (see Table 8.13). This trend remained the same from 1st to 4th reference.

experiment	reference number	nominal phrase	pronominal phrase
instruction	1st	182 (87.5%)	11 (5.3%)
instruction	2nd	123 (59.1%)	38 (18.3%)
instruction	3rd	85 (40.9%)	33 (15.9%)
instruction	4th	67 (32.2%)	16 (7.7%)
instruction	5th	39 (18.8%)	8 (3.9%)
instruction	6th	21 (10.1%)	2 (1.0%)
instruction	7th	14 (6.7%)	0 (0.0%)
instruction	8th	7 (3.4%)	0

Table 8.13: Nominal and pronominal reference to objects in instructions in first to eights reference; raw frequency with mean percentage in brackets.

- Between *verbal and visual goal* condition and instructions

A one-way ANOVA revealed a significant influence of communicative intention on the dependent variables pronominal reference and nominal reference in 1st reference (see Figure 8.11). Instructors used significantly more nominal references (87.5%) than participants in the *verbal and visual goal* condition (49.2%), $F(1,29) = 14.16$, $p = .001$. Participants in the *verbal and visual goal* condition (37.4%) used significantly more pronominal forms than instructors (5.3%) in first reference, $F(1,29) = 21.40$, $p = .00$.

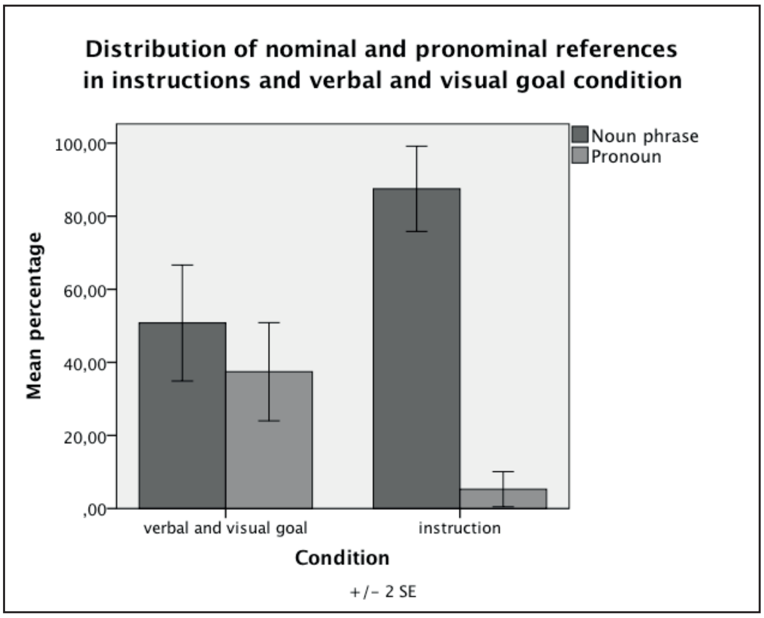


Figure 8.11: Distribution of nominal and pronominal references in *instructions* and *verbal and visual goal* condition in UOA in first reference; mean percentage with standard error.

The same significant differences were observed in 2nd reference. Instructors used significantly more nominal references (59.1%) than participants in the *verbal and visual goal* condition (28.2%) (see Table 8.12 on page 338), $F(1,29) = 8.75$, $p = .006$.

Participants in the *verbal and visual goal* condition (37.9%) used significantly more pronominal forms than instructors (18.3%) in second reference, $F(1,29) = 5.06$, $p = .03$.

In 3rd reference, the results revealed a significant effect of communicative attention on nominal references but no such effect on pronominal use. Instructors used significantly more nominal references (40.9%) than participants in the *verbal and visual goal* condition (19.5%), $F(1,29) = 4.83$, $p = .04$.

In 4th reference, the results revealed a significant effect of communicative attention on nominal references but no such effect on pronominal use. Instructors used significantly more nominal references (32.2%) than participants in the *verbal and visual goal* condition (5.6%), $F(1,29) = 10.40$, $p = .003$.

8.4.1.3 Feature: definiteness of nominal phrase

- Study 1: UOA

If 1st to 4th reference were analyzed as a combined set of data, a significant effect of amount of prior information on definiteness in nominal phrases could be observed, $L\chi^2$ (8, N = 615) = 29.94, $p = .00$. Participants in the *verbal and visual goal* condition used indefinite noun phrases significantly less frequently than participants in the other conditions, $z = -3.0$ (see Table 8.14 and Table 8.15).

condition	definite noun phrase	indefinite noun phrase	plural noun phrase	zero article noun phrase	numeral noun phrase
underspecified goal	93	59	37	25	4
verbal goal	105	51	21	18	1
verbal and visual goal	119	24	40	40	1

Table 8.14: Distribution of definite and indefinite noun phrases within conditions in UOA, first to fourth reference; raw frequency.

condition	definite noun phrase	indefinite noun phrase	plural noun phrase	zero article noun phrase	numeral noun phrase
underspecified goal	-1.8	1.7	0.4	0.8	1.3
verbal goal	0.4	1.3	-1.8	0.3	-0.6
verbal and visual goal	1.5	-3.0	1.4	-0.6	-0.

Table 8.15: Distribution of definite and indefinite noun phrases within conditions in UOA, first to fourth reference; standardized residuals.

Considering the distribution of categories annotating definiteness in 1st reference only, the same trend was observed. The results revealed that definite noun phrases were more frequent than indefinite ones in first reference (see Figure 8.12). There was no significant effect of amount of prior information on definiteness of nominal phrases in 1st, 2nd, 3rd, or 4th reference (see Table 8.16 for all results).

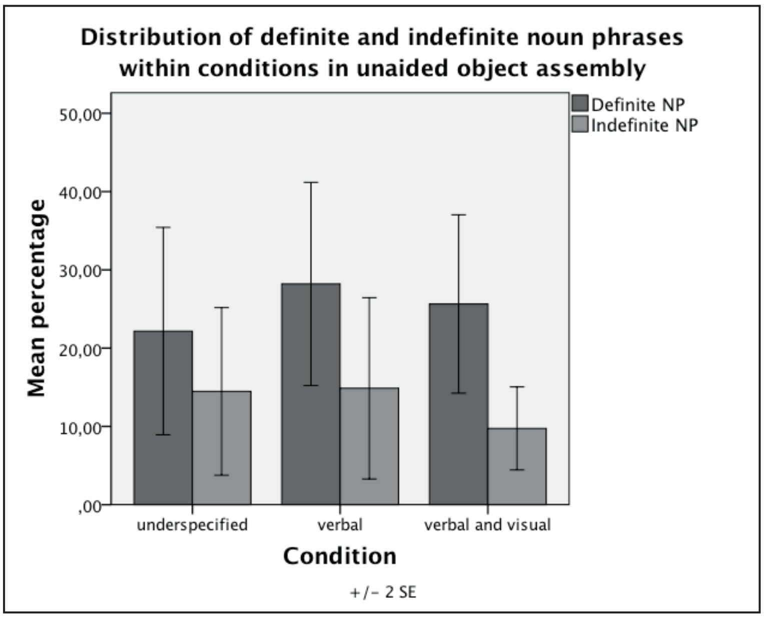


Figure 8.12: Distribution of definite and indefinite noun phrases within conditions in UOA in first reference; mean percentage and standard error.

condition	reference number	definite noun phrase	indefinite noun phrase	plural noun phrase	zero article noun phrase	numerals	ellipsis
underspecified goal	1st	51 (23.1%)	30 (13.6%)	14 (6.3%)	15 (6.9%)	3 (1.4%)	0 (0.0%)
underspecified goal	2nd	21 (9.5%)	17 (7.7%)	14 (6.3%)	7 (3.7%)	0 (0.0%)	0 (0.0%)
underspecified goal	3rd	10 (4.5%)	8 (3.6%)	7 (3.2%)	1 (0.5%)	0 (0.0%)	1 (0.5%)
underspecified goal	4th	8 (3.6%)	3 (1.4%)	6 (2.7%)	3 (1.4%)	1 (0.5%)	0 (0.0%)
verbal goal	1st	58 (29.7%)	29 (14.9%)	4 (2.1%)	12 (6.2%)	1 (0.5%)	0 (0.0%)
verbal goal	2nd	20 (10.3%)	12 (6.2%)	6 (3.1%)	5 (2.6%)	0 (0.0%)	0 (0.0%)
verbal goal	3rd	16 (8.2%)	10 (5.1%)	8 (4.1%)	1 (0.5%)	0 (0.0%)	0 (0.0%)

verbal goal	4th	11 (5.6%)	0	3 (1.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
verbal and visual goal	1st	50 (25.6%)	15 (20.0%)	21 (10.8%)	10 (5.1%)	1 (0.5%)	0 (0.0%)
verbal and visual goal	2nd	33 (16.9%)	4 (9.7%)	14 (7.2%)	3 (1.5%)	0 (0.0%)	0 (0.0%)
verbal and visual goal	3rd	26 (13.3%)	3 (4.1%)	5 (2.6%)	4 (2.1%)	0 (0.0%)	0 (0.0%)
verbal and visual goal	4th	10 (5.1%)	0 (0.0%)	1 (0.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Table 8.16: Results on the distribution of categories annotating definiteness of nominal phrase within conditions in UOA in first to fourth reference; raw frequency and mean percentage in brackets.

- Study 2: Instructions

Instructors had a clear tendency to use indefinite noun phrases from 1st to 4th reference (see Table 8.17).

condition	reference number	definite noun phrase	indefinite noun phrase	zero article noun phrase	numerals	ellipsis
instruction	1st	79 (38.0%)	103 (49.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
instruction	2nd	47 (22.6%)	73 (35.1%)	3 (1.4%)	0 (0.0%)	0 (0.0%)
instruction	3rd	32 (15.4%)	50 (24.0%)	3 (1.4%)	0 (0.0%)	0 (0.0%)
instruction	4th	31 (14.9%)	33 (15.9%)	3 (1.4%)	0 (0.0%)	0 (0.0%)

Table 8.17: Results definite an indefinite noun phrases in instructions from first to fourth reference, raw frequency with percentage in brackets.

- Between *verbal and visual goal* condition and instructions

The analysis revealed a significant effect of communicative intention on indefiniteness of noun phrases in 1st reference (see Figure 8.13). Instructors used significantly more indefinite references (49.6%) than participants in the *verbal and visual goal* condition (20.0%), $F(1,29) = 9.14$, $p = .005$.

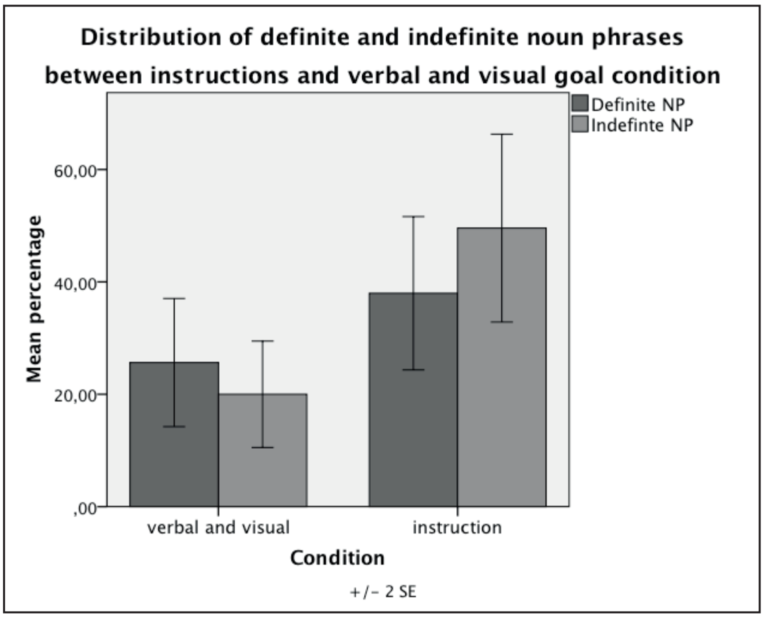


Figure 8.13: Distribution of definite and indefinite noun phrases in instructions and *verbal and visual goal* condition in UOA in first reference; mean percentage, standard error.

The analysis revealed the same effect of communicative intention on indefiniteness of noun phrases in 2nd reference (see Table 8.17). Instructors used significantly more indefinite references (35.1%) than participants in the *verbal and visual goal* condition (9.7%), $F(1,29) = 14.23$, $p = .001$.

The same effect of communicative intention on indefiniteness of noun phrases was revealed in 3rd reference. Instructors used significantly more indefinite references (24.0%) than participants in the *verbal and visual goal* condition (4.1%), $F(1,29) = 8.90$, $p = .006$.

The significant effect of communicative intention on indefiniteness of noun phrases was also revealed in 4th reference but indefinite noun phrases were very rare in the *verbal and visual goal* condition. Instructors used significantly more indefinite references (15.9%) than participants in the *verbal and visual goal* condition (0.5%), $F(1,29) = 8.78$, $p = .006$.

8.4.1.4 Feature: nominal specificity

- Study 1: UOA

Comparing the frequency of specific and unspecific nominal references from 1st through 4th reference, a significant effect of amount of prior information on specific nominal references was revealed, $\chi^2(2, N = 594) = 9.19$, $p = .01$. Participants in the *underspecified goal* condition used domain specific nouns less frequently from 1st to 4th reference (66 cases) than participant in the other two conditions, $z = -1.7$ (see Table 8.18).

condition	domain specific terminology (raw frequency)	domain specific terminology (standardized residual)	domain unspecific terminology (raw frequency)	domain unspecific terminology (standardized residual)
underspecified goal	66	-1.7	148	1.4
verbal goal	85	1.6	102	-1.3
verbal and visual goal	76	0.3	117	-0.2

Table 8.18: Raw frequency and standardized residuals of domain specific and domain unspecific nominal reference from first to fourth reference within conditions in UOA.

Considering 1st reference separately, the analysis of nominal specificity revealed that participants used *domain unspecific* nouns more frequently than domain specific nouns in all conditions (see Figure 8.14). There was no significant effect of amount of prior information on domain specific terminology when 1st, 2nd, 3rd, and 4th reference were analyzed individually.

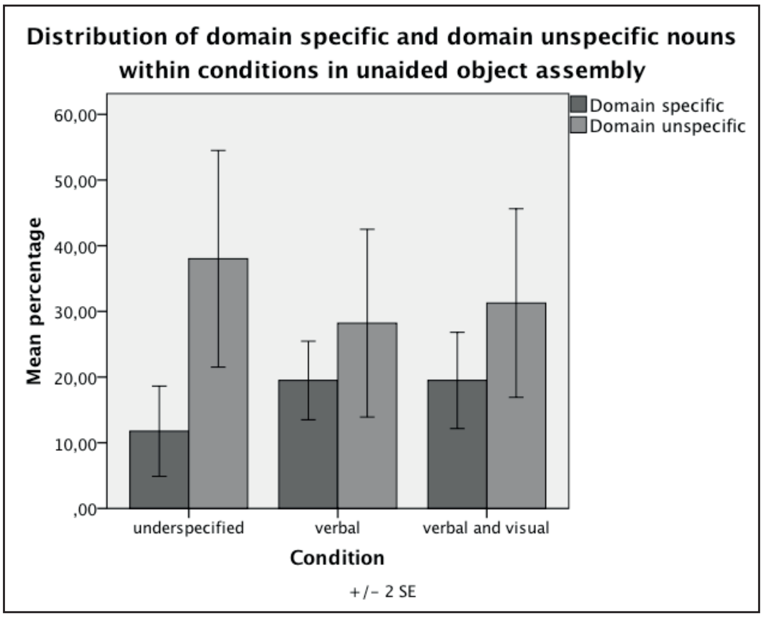


Figure 8.14: Distribution of domain specific and domain unspecific nouns within conditions in UOA in first reference; mean percentage and standard error.

Since the analysis focused specifically at the use of domain specific references, observed differences between conditions will be described although they are minor. Participants who were provided with unspecific information about the goal object used *domain unspecific* nouns (38.0%) slightly more often than participants who were provided with specific domain knowledge (29.2% in the *verbal goal* condition and 31.3% in the *verbal and visual goal* condition) in 1st reference (see Table 8.19). This tendency was mirrored in the use of domain specific terminology.

Domain specific nouns were used slightly more frequently in the *verbal goal* condition (19.0%) than in the *underspecified goal* condition (11.8%) in 1st reference.

Participants in the *verbal goal* condition were the only ones who used *domain specific* nouns more often than *domain unspecific* ones at some point. This was the case in second reference in which 13.3% of the references were *domain specific* as compared to 8.2% that were *domain unspecific*. In the *verbal and visual goal* condition, the frequency of *domain specific* and *domain unspecific* nominal references was similar (12.3% and 14.4% respectively). Although these differences between conditions were minor, it suggests that participants who were provided with least specific prior information used *domain unspecific* nominal references more often than participants who were provided with specific prior information.

condi-tion	referen-ce number	domain specific noun phrase	domain un-specific noun phrase	con-straints	pro-nominal phrase	no reference	others
A	1st	26 (11.8%)	84 (38.0%)	0 (0.0%)	50 (22.6%)	58 (26.2%)	3 (1.4%)
A	2nd	24 (10.9%)	34 (15.4%)	1 (0.5%)	43 (19.5%)	119 (53.8%)	0 (0.0%)
A	3rd	11 (5.0%)	15 (6.8%)	0 (0.0%)	32 (14.5%)	162 (73.3%)	1(0.5%)
A	4th	5 (2.3%)	14 (6.3%)	0 (0.0%)	13 (5.9%)	188 (85.1%)	1 (0.5%)
B	1st	37 (19.0%)	57 (29.2%)	2 (1.0%)	48 (24.6%)	43 (22.1%)	8 (4.1%)
B	2nd	26 (13.3%)	16 (8.2%)	1 (0.5%)	53 (27.2%)	99 (50.8%)	0 (0.0%)
B	3rd	14 (7.2%)	21 (10.8%)	0 (0.0%)	22 (11.3%)	138 (70.8%)	0 (0.0%)
B	4th	8 (4.1%)	6 (3.1%)	0 (0.0%)	12 (6.2%)	169 (86.7%)	0 (0.0%)
C	1st	34 (17.4%)	59 (31.3%)	3 (1.5%)	73 (37.4%)	25 (12.8%)	1 (0.5%)
C	2nd	24 (12.3%)	28 (14.4%)	2 (1.0%)	74 (37.9%)	67 (34.4%)	0 (0.0%)
C	3rd	12 (6.2%)	25 (12.8%)	1 (0.5%)	37 (19.0%)	120 (61.5%)	0 (0.0%)
C	4th	6 (3.1%)	5 (2.6%)	0 (0.0%)	25 (12.8%)	159 (81.5%)	0 (0.0%)

Table 8.19: Results referential form, variable: nominal specificity, within experiment 1 (all 3 conditions) (A: underspecified goal condition, B: verbal goal condition, C: verbal and visual goal condition); distribution over all coded forms.

- Study 2: Instructions

Instructors tended to use *domain unspecific* nouns from 1st to 3rd reference. The difference between *domain specific* and *domain unspecific* terminology was not as marked in 4th reference (see Table 8.20).

reference number	domain specific noun phrase	domain unspecific noun phrase	pronominal phrase	no reference
1st	35 (16.8%)	147 (70.67%)	11 (5.3%)	15 (7.2%)
2nd	38 (18.3%)	85 (40.87%)	38 (18.3%)	47 (22.6%)
3rd	29 (13.9%)	56 (26.92%)	33 (15.9%)	90 (43.3%)
4th	31 (14.9%)	36 (17.31%)	16 (7.7%)	125 (60.1%)

Table 8.20: Results on domain specific and domain unspecific nouns in instructions in first to fourth reference; raw frequency with percentage in brackets.

- <u>Between *verbal and visual goal* condition and instructions</u>

The analysis revealed a significant effect of communicative intention on *domain unspecific* nominal references in 1st reference (see Figure 8.15). Instructors used significantly more *domain unspecific* references (70.7%) than participants in the *verbal and visual goal* condition (31.3%), F (1,29) = 15.68, p = .00.

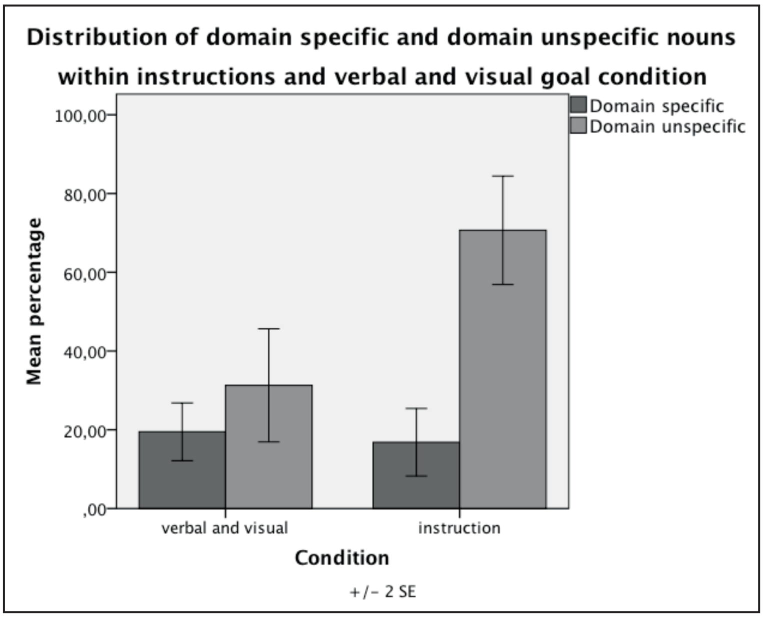


Figure 8.15: Distribution of domain specific and domain unspecific nouns in instructions and in the *verbal and visual goal* condition in UOA in first reference; mean percentage and standard error.

The analysis revealed the same effect of communicative intention on *domain unspecific* noun phrases in 2nd reference. Instructors used significantly more *domain unspecific* references (40.9%) than participants in the *verbal and visual goal* condition (14.4%), F (1,29) = 8.28, p = .007.

There was no statistically significant effect of communicative intention on *domain unspecific* noun phrases in 3rd reference. However, the effect was revealed again in 4th reference. Here instructors used significantly more *domain unspecific* references

(17.3%) than participants in the *verbal and visual goal* condition (2.6%), F (1,29) = 6.85, p = .014.

- Domain specific and domain unspecific nominal reference to specific objects

In order to assess the typicality of the different objects, nominal specificity of all objects in first reference was compared. The analysis focused on 1st reference only because it was assumed that first references highlighted the intuitive conceptualization of the respective object. The most culturally specific object was the object functioning as the roof. This object was most frequently referred to by the domain specific noun 'Dach' (roof). All other objects tended to be referred to by domain unspecific nominal or pronominal phrases (see Table 8.21).

condition	object(s)	domain specific nominal phrase	domain unspecific nominal phrase	pronominal phrase	others	no reference
under-specified goal	roof	7 (41.2%)	4 (23.5%)	2 (11.8%)	0 (0.0%)	4 (23,5%)
verbal goal	roof	12 (80.0%)	1 (6.7%)	1 (6.7%)	0 (0.0%)	1 (6.7%)
verbal and visual goal	roof	11 (73.3%)	0 (0.0%)	1 (6.7%)	0 (0.0%)	3 (20.0%)
under-specified goal	all others	19 (9.3%)	80 (39.2%)	48 (23.5%)	3 (1.5%)	54 (26.5%)
verbal goal	all others	25 (13.9%)	58 (32.2%)	47 (26.1%)	8 (4.4%)	42 (23.3%)
verbal and visual goal	all others	23 (12.8%)	62 (34.4%)	72 (40.0%)	1 (0.6%)	22 (12.2%)

Table 8.21: Results on nominal specificity in reference to the roof part and all other objects in UOA; raw frequency and percentage in brackets.

8.4.2 Referential process

8.4.2.1 Distribution of referential process categories

- Study 1: UOA

The analysis of referential process in 1st to 2nd reference highlighted a significant effect of amount of prior information on the dependent variable *repeating*, F (2,44) = 5.104, p = .01. Post-hoc tests revealed that participants in the *verbal goal* condition (10.3%) repeated references significantly less frequently than participants in the *verbal and visual goal* condition (29.2%) (see Table 8.22), Tukey p = .010 and LSD p = .004. Additionally, the less sensitive LSD post-hoc test revealed that participants in the *verbal and visual goal* condition used *repeating* more frequently than participants in the *underspecified goal* condition (15.4%) as well, LSD p = .025.

condition	no reference	no further reference	reducing	reframing	repeating	rephrasing	reusing	single reference
A	5 (2.3%)	0 (0.0%)	22 (10.0%)	17 (7.7%)	34 (15.4%)	27 (12.2%)	0 (0.0%)	116 (52.5%)
B	0 (0.0%)	0 (0.0%)	40 (20.5%)	21 (10.8%)	20 (10.3%)	15 (7.7%)	0 (0.0%)	99 (50.8%)
C	2 (1.0%)	0 (0.0%)	31 (15.9%)	16 (8.2%)	57 (29.2%)	23 (11.8%)	0 (0.0%)	66 (33.8)

Table 8.22: Distribution of categories of referential change within conditions in UOA in first reference; raw frequency with mean percentage in brackets (A: underspecified goal; B: verbal goal; C: verbal and visual goal).

No significant effect of amount of prior information was revealed in 2^{nd} to 3^{rd} reference. A trend was observed in 3^{rd} to 4^{th} reference. It was observed that prior information influenced frequency of *no further reference*. However, this trend did not reach statistical significance, F (2,44) = 2.697, p = .08. The LSD post-hoc test highlighted that participants in the *verbal and visual goal* condition (45.6%) had more instances of the category *no further references* than participants in the *underspecified goal* condition (29.9%), p = .03 (see Table 8.36 in the appendix).

The same trend was also observed in 4^{th} to 5^{th} reference and as before the amount of prior information had an effect on *repeating*. However, both of these trends did not reach statistical significance. A one-way ANOVA testing for the influence of amount of prior information on the dependent variable *no further reference* revealed a similar trend as before, F (2,44) = 2.754, p = .075. Again, the difference was between the two conditions that differed most with regard to amount of prior information and it was in the same direction. The LSD post-hoc test highlighted that participants in the *verbal and visual goal* condition (52.3%) had more instances of the category *no further references* than participants in the *underspecified goal* condition (34.4%), p = .024 (see Table 8.36 in the appendix).

A one-way ANOVA testing for the influence of amount of prior information on the dependent variable *repeating* revealed a minor trend, F (2,44) = 2.904, p = .07. The LSD post-hoc test revealed that participants in the *verbal goal* condition (2.1%)

repeated references less frequently than participants in the *verbal and visual goal* condition (8.2%), p = .021 (see Table 8.36 in the appendix).

- Study 2: Instructions

Similarly, instructors referred to 79% of the objects more than once and did so by repeated (26.0%) or reduced forms (20.7%) (see Table 8.23).

reference	no reference	no further reference	reducing	reframing	repeating	rephrasing	reusing	single reference
1st – 2nd	13 (6.3%)	0 (0.0%)	43 (20.7%)	23 (11.1%)	54 (26.0%)	46 (22.1%)	0 (0.0%)	42 (20.2%)
2nd – 3rd	13 (6.3%)	47 (22.6%)	27 (13.0%)	13 (6.3%)	34 (16.4%)	34 (16.4%)	11 (5.3%)	42 (20.2%)
3rd – 4th	13 (6.3%)	83 (39.9%)	4 (1.9%)	16 (7.7%)	31 (14.9%)	22 (10.6%)	10 (4.8%)	42 (20.2%)
4th – 5th	13 (6.3%)	119 (57.2%)	5 (2.4%)	4 (1.9%)	16 (7.7%)	17 (8.2%)	5 (2.4%)	42 (20.2%)

Table 8.23: Distribution of categories of referential change within instructions from first to fifth reference; raw frequency with mean percentage in brackets.

- Between verbal and visual goal condition and instructions

No effect of communicative intention on referential process was revealed in 1st to 2nd reference. A trend was observed, however, in 2nd to 3rd reference. In these references communicative intention had a slight effect on *reducing*, F (1,29) = 3.504, p = .071. Instructors tended to reduce third references more frequently (12.98%) than participants thinking aloud in the *verbal and visual goal* condition (5.1%).

A slight effect of communicative intention on the dependent variable *reusing* was observed in 3rd to 4th reference on, F (1, 29) = 3.505, p = 0.7.1 Instructors tended to reuse references from 3rd to 4th reference more frequently (4.8%) than participants thinking aloud in the *verbal and visual goal* condition (1.0%).

- Summary

Reviewing these findings, it can be summarized that prior information as well as communicative intention influenced the referential process slightly. Table 8.24 summarizes the most frequent referential process categories for both studies and the individual conditions.

reference	underspecified goal condition	verbal goal condition	verbal and visual goal condition	instruction
1st – 2nd	repeating (15.4%), rephrasing (12.2%)	reducing (20.5%), reframing (10.8%)	repeating (29.2%), reducing (15.9%)	repeating (26.0%), reducing (20.7%)
2nd – 3rd	repeating (10.0%)	repeating (11.3%)	repeating (17.9%), rephrasing (10.3%)	repeating (16.4%), rephrasing (16.4%)
3rd – 4th	repeating (5.9%)	repeating (5.2%)	repeating (8.2%)	repeating (14.9%), rephrasing (10.6%)

Table 8.24: Summary of the most frequent referential processes in the three conditions in UOA and in instructions; the most frequent referential process category and the second most frequent one are only listed if it accounts for more than 10% of the annotated cases.

8.4.2.2 Function assignment within the referential process in terms of reference process time

As pointed out in sub-section 2 of this chapter, the point in *reference process time* at which participants conceptualized the given objects within the conceptual domain 'house' was of interest to the theoretical discussion. Therefore, a closer analysis of the category *reframing* is reported in the following section.

- Function assignment from 2nd to 5th reference

Considering the occurrences of *reframing* within the first four references to an object, it was revealed that objects were most frequently reframed in second reference in UOA as well as in instructions (see Table 8.25). The frequency of *reframing* was almost equal between all three conditions and both experiments. Frequency of domain specific function assignment decreased over consecutive references.

reference	underspecified goal condition	verbal goal condition	verbal and visual goal condition	instruction
1st – 2nd	17 (7.7%)	21 (10.8%)	16 (8.2%)	23 (11.1%)
2nd – 3rd	9 (4.1%)	9 (4.6%)	9 (4.6%)	13 (6.3%)
3rd – 4th	4 (1.8%)	5 (2.6%)	2 (1.0%)	16 (7.7%)

Table 8.25: Occurrences of reframing in the three assembly conditions (UOA) and in instructions; raw frequency with mean percentage in brackets.

By adding the number of objects that were verbalized within the goal domain by reframing from second to fourth reference a measure of explicit function assignment [256] is established. In order to get an accurate percentage of conceptualized objects, the number of referred objects needed to be calculated as a first step. For participants in the *underspecified goal* condition the following equation was calculated: 221-58=163, i.e. number of all objects minus number of objects that were not referred to at all. This number of referred objects represents 100% of possible references and from these 100% the number of explicitly conceptualized objects was calculated.

The results highlighted that participants in the *underspecified goal* condition explicitly conceptualized 30 objects (18.4% of all 163 objects), participants in the *verbal goal* condition conceptualized 35 objects explicitly (23.0% of all 152 objects) whereas participants in the *verbal and visual goal* condition conceptualized 27 objects (15.9% of all 172 objects) explicitly. Instructors referred to 52 objects by functional terms (26.9% of all 193 objects).

Besides the number of objects that are functionally referred to, it is important to know how many participants, i.e. protocols, contributed to those numbers. In the *underspecified goal* condition, 13 out of 17 participants (76.5%) conceptualized at least one object functionally from second to fourth reference. Similarly, 11 out of 15 participants (73.3%) in the *verbal goal* condition conceptualized at least one object functionally from second to fourth reference. In the *verbal and visual goal* condition, fewer participants referred to an object functionally from second to fourth reference, i.e. nine out of 15 participants (60.0%). Even fewer instructors assigned function explicitly, namely nine out of 16 instructors (56.3%).

- Function assignment throughout the referential process

In order to evaluate the number of objects that were functionally referred to throughout the analyzed referential process, the objects that were referred to by specific nominal phrases in first reference were considered as well. The results revealed that participants in UOA and instructors conceptualized objects to a comparable amount in first reference (see Table 8.26 first row).

The frequency of object conceptualization from first to fifth reference showed that participants who were provided with *verbal goal* information conceptualized

[256] The term explicitly refers to the fact that those concepts are verbally expressed in a change from an unspecific term to a more functional term. Objects that are referred to by functional, i.e. domain specific terms, in first reference are not included in the first step of the calculation.

almost half of all objects in first reference (47.4%). Instructors conceptualized objects at a comparable frequency in first reference (45.1%). Participants who were provided with *verbal and visual* information referred to almost as many objects functionally as participants who were provided with no goal information (see Table 8.26 last row).

reference	underspecified goal condition	verbal goal condition	verbal and visual goal condition	instruction
1st	26	37	34	35
1st – 4th	30	35	27	52
sum explicitly conceptualized objects	56	72	61	87
objects without one reference	58	43	25	15
percent conceptualized objects	34.4%	47.4%	35.9%	45.1%

Table 8.26Overview of domain specific nouns in first reference, occurrences of reframing from first to fifth reference, objects which are not referred to at all, percent conceptualized objects from first to fifth reference in the three conditions (experiment 1) and instructions (experiment 2).

8.4.2.3 Function assignment within the referential process in terms of assembly time

So far, conceptualization was investigated in terms of *reference process time*. However, the assembly process itself also provides a time scale, referred to as *assembly time*. This time needs to be considered as well in order to get the full picture of object conceptualization over time.

The distribution revealed a difference between assembly conditions and think aloud and instructions (see Table 8.27). Participants who were provided with no specific information about the goal object (*underspecified goal* condition) tended to conceptualize objects in the second quarter of the assembly process. Participants who were provided with *verbal and visual* prior information conceptualized most objects in the last quarter of the assembly. Those participants who were given *verbal goal* information tended to conceptualize most objects in the third quarter.

Focusing on the first quarter, a marked difference in the amount of conceptualization was observed with regard to the amount of provided goal information. Those participants who were provided with verbal information assigned function to objects most often. Instructors tended to assign function explicitly most often in the fourth quarter of their instructions.

quarter	underspecified goal condition	verbal goal condition	verbal and visual goal condition	instruction
1st	4 (13.3%)	8 (22.9%)	2 (7.4%)	6 (11.5%)
2nd	13 (43.3%)	6 (17.1%)	7 (25.9%)	10 (19.2%)
3rd	4 (13.3%)	13 (37.1%)	5 (18.5%)	5 (9.6%)
4th	9 (30.0%)	8 (22.9%)	13 (48.1%)	31 (59.6%)

Table 8.27: Overview of occurrences of reframing in *assembly and instruction time* from 2nd to 4th reference in the different quarters of the annotated discourse units; distinguished between the three conditions in UOA and in instructions; raw frequency with percentage in brackets.

8.4.3 Functional mapping phrases

Overall 131 instances of mapping were identified and coded. Most instances were identified in protocols recorded in the *underspecified goal* condition (54 instances, 41.2%). Fewer instances were identified in protocols in the *verbal goal* condition (44 instances, 33.6%) and even fewer in protocols of the *verbal and visual goal* condition (33 instances, 25.2%).

8.4.3.1 Mapping phrases in process categories

The great majority of verbalizations of mapping processes was identified in the category *hypothesis* (83.1%). Fewer cases were revealed in *positive evaluations* (7.6%) and negative evaluations (3.1%). *Actions* did not contain any verbalizations of mapping processes (see Table 8.28).

problem solving process category	raw frequency	percentage
hypothesis	109	83.21
action	0	0
positive evaluation	10	7.63
negative evaluation	4	3.05
description of mental state	2	1.53
comment on object features	1	0.76
beginning	2	1.53
end	2	1.53
unrelated	1	0.76
sum	131	100

Table 8.28: Distribution of mapping phrases between process categories (mean raw frequency and percentage).

8.4.3.2 Kinds of mapping phrases and conditions

Two kinds of mapping phrases were distinguished in self-assembly protocols, namely *direct mapping* of the kind "**this** is a **wall**" and *representational mapping* "**this** looks like a **wall**" (see section 8.3.3 for more detail).

Participants in the *underspecified goal* condition used explicit mapping phrases most frequently (M = 3.1 cases, SD = 0.8). Participants in the *verbal and visual goal* condition assigned function explicitly least often (M = 1.8 cases, SD = 0.5) (see Table 8.38 in the appendix).

The frequency of the two kinds of mapping based on mean percentage highlighted that *direct mapping* was more frequent in all conditions than *representational mapping* (see Figure 8.16). It was most frequent in the *verbal goal* condition (71.9%, SD = 9.9%), less frequent in the *underspecified goal* condition (62.4%, SD = 10.5%) and least frequent in the *verbal and visual goal* condition (53.9%, SD = 11.7%) (see Table 8.39 in the appendix). The analysis of the different verbs used in phrases of *direct mapping* highlighted that 'sein' (to be) was used in the majority of cases (91.5%).

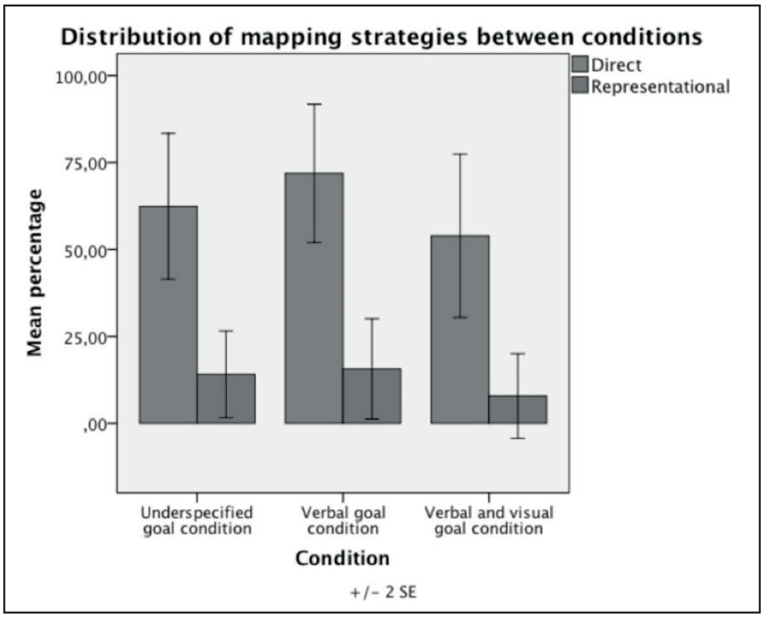


Figure 8.16: Distribution of mapping types between conditions (mean percentage with +/ - 2 standard errors).

Representational mapping was used to a comparable number in the *underspecified goal* (14.1%, SD = 6.2) and the *verbal goal* condition (15.6%, SD = 7.2%) and less frequently in the *verbal and visual goal* condition (7.8%, SD = 6.1%) (see Table 8.39 in the appendix). The verb analysis highlighted that 'aussehen wie' (look like) was the most frequently used relational term in *representational mapping* (59.1%).

Among the others were 'benutzen' (use) (13.6%) and 'verwenden' (use) (9.1%) which are synonymous in English.

8.4.3.3 Mapping phrases in instructions

23 instances of mapping were identified in 12 instruction texts. As described in the analysis section, a third kind of mapping was identified in instruction texts. This type was called *reframing*. The distribution of mapping types within instructions revealed a clear trend (see Figure 8.17). Instructors used *direct mapping* most frequently (36.5%, SD = 10.7) (see Table 8.29). The influence of experimental condition on the dependent variable mapping strategy was statistically significant, $L\chi^2$ (6, N = 154) = 36.42, p = .00.

Dependent variable	mean raw frequency	mean percentage
direct mapping	0.63 (0.20)	36.46% (10.67)
representational mapping	0.44 (0.18)	26.04% (10.43)
reframing	0.31 (0.12)	21.88% (9.09)
ellipsis	0.06 (0.06)	3.13% (3.13)

Table 8.29: Frequency of mapping types within instructions (mean raw frequency and percentage with standard deviation in brackets).

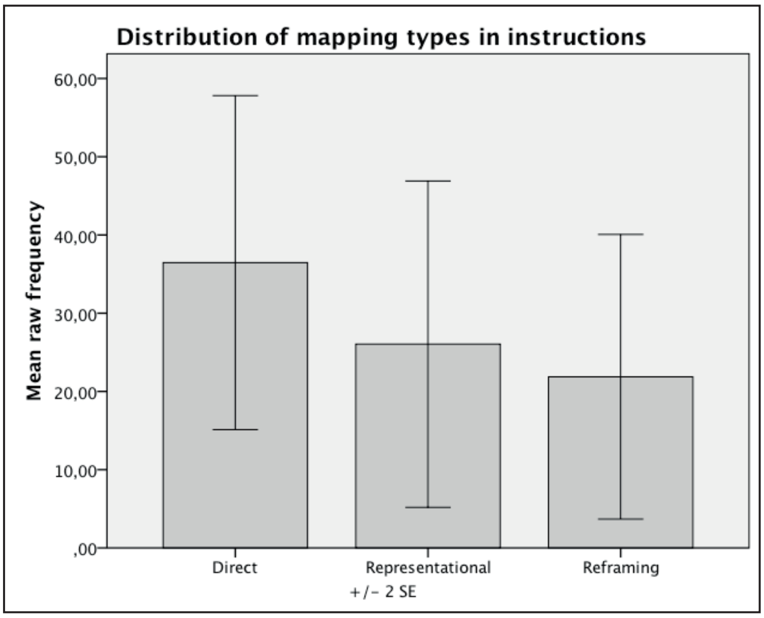


Figure 8.17: Distribution of mapping types in instructions (mean raw frequency with + / - 2 standard errors).

The strategy of *reframing* that was unique to instructions contributed most to this effect, z = 5.4, p < .001. Additionally, instructors also used instances of *representational mapping* (M = 0.4, SD = 0.2) significantly more frequently than participants in UOA (M = 0.4, SD = 0.8). Furthermore, the results revealed that *reframing* (21.9%, SD = 9.1) was used almost as frequently as instances of representational mapping. These results reflected a greater variability of mapping types in instructions than in self-assembly.

In phrases of *direct mapping* instructors used the verb 'sein' (to be) exclusively. In *representational mapping*, 'bilden' and 'darstellen' (both in the sense of 'represent') were used.

After providing a broad overview on the distribution of mapping types within the self-assembly experiment and in instructions, two additional features of mapping processes were investigated, namely use of modal verbs and type of source entity.

8.4.3.4 Distribution of modal verbs in mapping phrases between experiments

In all mapping processes 29 modal verbs were used by 18 participants. The use of modal verbs in mapping processes was equally distributed between conditions with regard to raw frequency of use (see Table 8.30). Instructors did not use modal verbs in mapping processes at all.

dependent variable	condition	raw frequency	mean raw frequency	mean percentage
no modal verb	underspecified goal	23	2.59 (0.72)	59.85% (11.27)
no modal verb	verbal goal	5	2.06 (0.57)	62.08% (10.81)
no modal verb	verbal and visual goal	9	1.24 (0.39)	41.10% (10.91)
modal verb	underspecified goal	11	0.59 (0.23)	16.62% (7.96)
modal verb	verbal goal	8	0.50 (0.18)	25.42% (9.29)
modal verb	verbal and visual goal	10	0.58 (0.23)	23.42% (8.97)

Table 8.30: Results on the dependent variable modal verb in mapping processes in unaided object assembly (raw frequency, mean raw frequency and percentage with standard deviation in brackets).

The analysis of modal verb type revealed that these were differently distributed between conditions. However, this difference did not reach statistical significance, $L\chi^2$ (4, N = 29) = 8.51, p = .075. This result could be explained by the low frequency of modal verbs. As these tendencies were highlighted, these are reported.

Three modal verbs were used in mapping processes, namely 'können' (can), 'müssen' (must), and 'sollen' (have to). Participants in the *underspecified goal* condition used all three modal verbs. However, the modal verbs 'können' (41.7%, SD = 16.0%) and 'sollen' (44.4%, SD = 18.6%) were more frequent than 'müssen' (13.9%, SD = 9.0%). Participants in the *verbal goal* condition showed a clear preference for 'können' (75.0%, SD = 7.1%), whereas participants in the *verbal and visual goal* condition used 'müssen' most frequently (55.6%, SD = 20.5%) (see Table 8.31).

dependent variable	condition	raw frequency	mean percentage & standard deviation
müssen	underspecified goal	2	13.89% (9.04)
müssen	verbal and visual goal	5	55.56% (20.49)
können	underspecified goal	5	41.67% (15.69)
können	verbal goal	6	75.00% (17.08)
können	verbal and visual goal	3	22.78% (18.09)
sollen	underspecified goal	4	44.44% (18.59)
sollen	verbal goal	2	25.00% (17.08)
sollen	verbal and visual goal	2	16.67% (16.67)

Table 8.31: Results on the dependent variable modal verb type in mapping processes within the different conditions in UOA (raw frequency with mean percentage in brackets and standard deviation of percentage).

8.4.3.5 Object reference and functional terminology in mapping phrases between experiments

The second part of the analysis concerned the nature of the source entity. As outlined in the analysis section, these could either be deictic, i.e. referring to an object in the environment, or nominal.

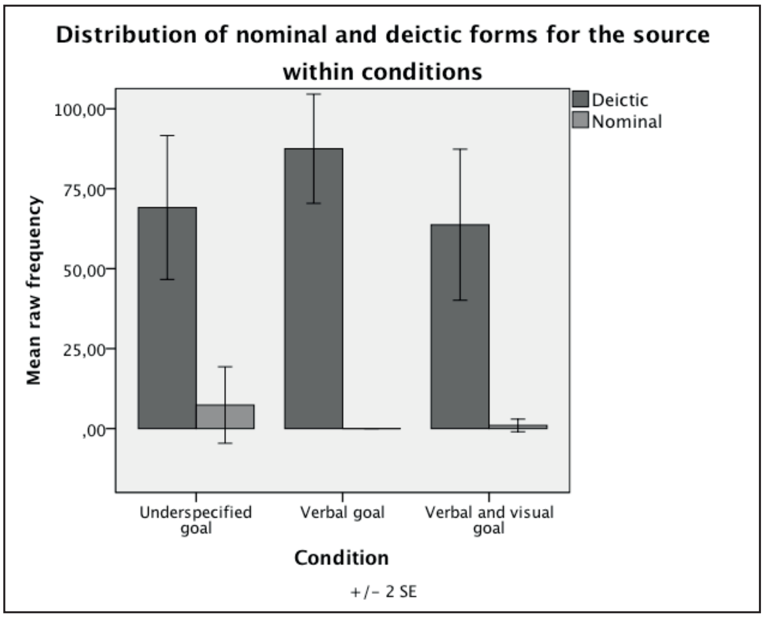


Figure 8.18: Nature of source entities in UOA (mean percentage with +/ 2 standard errors).

The clearest effect of condition was observed on the nature of the source entity. A chi-square test revealed a significant effect of communicative intention on the dependent variable *nominal source* (see Figure 8.18), $L\chi^2$ (6, N = 154) = 57.7, p = .00. Instructors used nominal terms to refer to the source entity (M = 0.9, SD = 0.9) more frequently than participants in the self-assembly scenario (M = 0.1, SD = 0.2), z = 7.5, p < .001 (all results are reported in Table 8.40 in the appendix; see Table 8.42 in the appendix for standardized residuals).

Deictic expressions were most frequently used in all three self-assembly conditions, i.e. *underspecified goal* (69.1%, SD = 11.3), *verbal goal* (87.5%, SD = 8.54), and *verbal and visual goal* (63.7%, SD = 11.8) (see Table 8.41 in the appendix). Instructors used deictic references significantly less often than assemblers, z = 2.6, p < .01. The results further highlighted that participants in the *verbal goal* condition used deictic references only.

The analysis of the functional term that was assigned to the object revealed that five functional terms or their synonyms were used most frequently, namely 'Boden' (floor), 'Dach' (roof), 'Wand' (wall), 'Etage' (story), and 'Haus' (house). This list represents two perspectives, i.e. whereas one perspective focuses on parts of the goal structure (floor, roof, and wall), the other perspective focuses on the whole structure (house and story).

A chi-square test revealed a significant effect of amount of prior information on perspective expressed by the assigned functional term, $L\chi^2$ (4, N = 131) = 13.122, p = .01. Participants in the *underspecified goal* condition tended to consider the whole structure of the goal object (M = 1.6, SD = 0.6) (see Table 8.32). All other participants in unaided assembly tended to refer to functional parts of the goal object (see Figure 8.19).

The difference remained statistically significant if instructors were considered as well, $L\chi^2$ (6, N = 154) = 13.4, p = .04. This finding highlights that instructors tended to refer to functional parts of the goal object rather than on the structure, too.

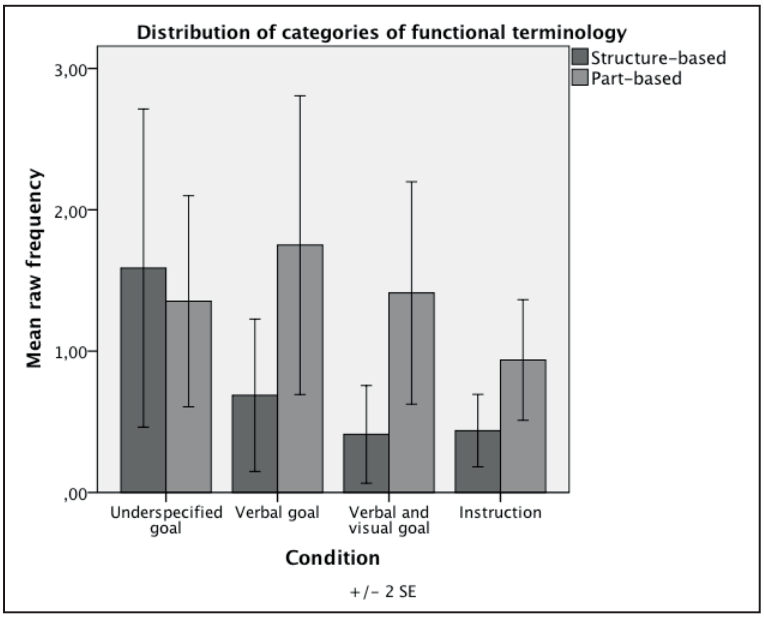


Figure 8.19: Distribution of functional terminology expressing structure-based and part-based perspectives (mean raw frequency with +/ - 2 standard errors).

perspective expressed in functional term	condition	mean raw frequency
whole goal structure	underspecified goal	1.59 (0.57)
whole goal structure	verbal goal	0.69 (0.27)
whole goal structure	verbal and visual goal	0.41 (0.17)
whole goal structure	instruction	0.44 (0.13)
parts of goal structure	underspecified goal	1.35 (0.37)
parts of goal structure	verbal goal	1.75 (0.53)
parts of goal structure	verbal and visual goal	1.41 (0.39)
parts of goal structure	instruction	0.93 (0.21)

Table 8.32: Frequency of categories of functional terminology (raw frequency with standard deviation in brackets).

8.5 Discussion

Research question 1 investigated which referential form participants chose when referring to objects that did not suggest a specific function right away. The difference between referential form was investigated in think aloud data and in instructions, i.e. in self-oriented speech vs. addressee-oriented speech. Against prior expectations, no influence of amount of prior information was observed in the analysis of individual references.

Although the results were presented individually in the previous section, those findings about referential form are combined in the discussion to provide a picture on the general form of nominal reference. This analysis revealed a significant effect of amount of prior information on definiteness of nominal phrase and pronominal use. Participants in the *verbal and visual goal* condition tended to use fewer indefinite noun phrases than participants in the other conditions. They rather used more pronominal forms than the other participants did. Participants in the *underspecified goal* condition used pronominal forms least frequently. Furthermore, participants in the *verbal and visual goal* condition had the lowest frequency of 'no references' in first to fourth reference. Table 8.33 summarizes the most frequent forms in the three assembly conditions and in instructions from 1st to 4th reference.

condition	reference number	most frequent referential form	example
underspecified goal	1	definite noun phrase with unspecific noun	diese Sachen (those things) [60][257]
verbal goal	1	definite noun phrase with unspecific noun	die Teile (those parts) [11]
verbal and visual goal	1	definite noun phrase with unspecific noun	die ganzen Zwischenstücke (all of those parts) [24]
Instruction	1	indefinite noun phrase with unspecific noun	die Einzelteile (those single parts) [20]
underspecified goal	2	indefinite noun phrase with unspecific noun	dieses Ding (this thing)
verbal goal	2	pronoun	das (it)
verbal and visual goal	2	pronoun	die (it)
Instruction	2	indefinite noun phrase with unspecific noun	dieses Brett (this board)
underspecified goal	3	pronoun	die (it)
verbal goal	3	definite or indefinite noun phrase with unspecific noun	das (it)
verbal and visual goal	3	definite noun phrase with unspecific noun, pronoun	die (it)
Instruction	3	indefinite noun phrase with	*das* (*that*)

[257] All examples in the respective condition are taken from one protocol. The participant's ID is provided in brackets. In the case of instructions it was not possible to select a prototypical case. The deviation from the prototypical example is indicated in italics.

		unspecific noun	
underspecified goal	4	definite noun phrase with unspecific noun	das Ding (that thing)
verbal goal	4	definite noun phrase with unspecific noun, pronoun	das Ding (that thing)
verbal and visual goal	4	pronoun	das (it)
Instruction	4	definite or indefinite noun phrase with unspecific noun	dieses Brett (this board)

Table 8.33: Summary of the most frequent referential forms (nominal vs. pronominal, definiteness, nominal specificity) in the conditions in UOA and in instructions in 1st to 4th reference.

Going back to the proposed German givenness hierarchy it can be observed that three cognitive statuses were encoded in referential form in the analyzed data. The least restricted cognitive status (type identifiable) is encoded in indefinite noun phrases. Definite noun phrases, in contrast, encode the cognitive status of uniquely identifiable, and pronouns signal that an object is in focus. These three cognitive statuses can be distinguished into two broader categories, namely familiar within the assembly context (pronouns and definite noun phrases) and generally identifiable but unfamiliar within the assembly context (indefinite noun phrases). Participants engaged in the assembly task referred to objects as familiar within the assembly context but not in focus in first reference. This suggests that the individual object parts were identified on the descriptive level[258] but they were not familiar and in focus yet because otherwise they would have been referred to by pronouns.

Instructors used indefinite noun phrases indicating that those objects can be identified by type but they were unfamiliar within the given task context. This is supported by von Stutterheim et al. (1993:111) who pointed out that indefinite noun phrases encode the status ‘to be identified’. Since instructors themselves were familiar with the goal object and the function of the individual parts, it can be assumed that they tailored their reference for an uninformed listener. This data revealed a notable influence of communicative intention on referential form. Participants speaking for themselves indicated a higher cognitive status of a referent than instructors. This finding suggests that the objects were more familiar to participants thinking aloud than the familiarity that was assumed by instructors for their listener.

[258] This terminology refers back to the opposition between descriptive level and the domain specific level of *representational metonymy* (Rieser 1996).

In 2nd reference, the picture in self-assembly got more diverse, i.e. the influence of amount of prior information showed. Participants who did not have any goal specific prior information (*underspecified goal* condition) used referential forms encoding the least restrictive cognitive status (type identifiable) indicating that the objects were less recognizable within the assembly context than assumed in first reference. Objects were described on a basic level as objects that could be identified by type but were unfamiliar within the specific context. This was different for participants who were provided with specific goal information. Those self-assemblers used pronominal forms signaling that the referred object was in focus. Participants who were provided with verbal information also conceptualized objects within the goal domain in second reference. 20% of the participants used definite noun phrases with a domain specific noun. This was the most explicit indication for conceptualization within the goal domain. The occurrence of pronominal forms and definite domain specific nominal phrases at a comparable frequency supports the claim that pronouns signal not only the cognitive status of 'in focus' but also conceptualization and familiarity within the specific assembly context. Instructors still assumed the initial unfamiliar cognitive status of objects for their listener.

Those participants in the *underspecified goal* condition, who referred to an object for a third time (26.7%), signaled that the objects were known and conceptualized within the assembly context. This was also observed for participants in the *verbal and visual goal* condition. Participants in the *verbal goal* condition signaled familiarity as well as general type knowledge about the objects to a comparable degree. This suggests that those participants held a possible goal concept and thus identified objects within the assembly task. Other participants, in contrast, had difficulties in categorizing objects within the assembly context hence they referred to its type. However, this insecurity was resolved by fourth reference; here all references by participants engaged in the self-assembly task signaled familiarity within the assembly context. At this point in referential process time, instructors also assumed familiarity for some objects whereas other objects were still referred to by a less specific form.

To conclude, the analysis of referential form revealed that amount of prior information as well as communicative intention influenced referential form. Participants who were provided with *verbal and visual goal* knowledge referred to objects with forms indicating familiarity within the assembly context over all four references. Change in cognitive status was observed in references by participants who were provided with *unspecific goal* knowledge or *verbal goal* information. This finding confirms parts of the initial hypothesis about referential form. It suggests that participants who had much prior information used forms signaling familiarity more frequently and constantly than participants in the other two conditions. However, unexpectedly, they did not use more specific nominal forms. One possible explanation can be found in Ungerer and Schmid's (2006) observation that

> "with regard to its cognitive function, the superordinate category HOUSE does not highlight a single or a few attributes of its 'members', as is typical of type-of superordinates (ROOM for LIVING ROOM, BEDROOM, KITCHEN etc.). Its strength is its assembling function, the notion that the category HOUSE is thought of as being composed of LIVING ROOM, BEDROOM etc." (Ungerer & Schmid 2006:88)

Following this argument it was not the attributes (e.g. wall, door, window) that came to mind first when looking at the presented picture. Given that the structure of the house needed to be assembled, the broader structures of rooms were not prominent in the beginning but the parts of a house needed to be conceptualized. Thus, the low number of specific nouns might mirror the encountered difficulties in conceptualizing the individual parts as attributes of the concept house.

Communicative intention had the greatest influence on referential form in that objects were referred to by the least restricted form for a long time (first to third reference) before some of the objects were assumed to be familiar. This observed referential behavior by instructors did not confirm the initial hypothesis. Rather than introducing the objects in first and second reference explicitly concerning their function and assuming them as known thereafter, instructors used the least restricted form in the first three references. These findings contradict the expectations about nominal specificity. Recalling the findings presented by Meyer-Fujara and Rieser (2005) that are contradicted by the presented findings, it supports the claim that the given object parts did not resemble conventional house parts. Thus, domain specific references would not facilitate identification of the referent. Assuming that instructors conceptualized the objects for themselves during assembly, this fallback on terms on the description level can be interpreted as an indicator for audience design.

The findings on definiteness also contradicted the formulated assumptions that were based on data reported in Mangold-Allwinn et al. (1995:36), who reported that instructors used definite noun phrases in first mention. The differences might be attributed to differences in design. In their experiment, Mangold-Allwinn et al. (1995) showed an assembly video to the instructors and asked them to instruct a fictional listener afterwards. Regarding prior information the data presented in this chapter contradicted Kiefer et al.'s (1993) findings. Kiefer et al. (1993) reported that instructors with action-oriented prior knowledge used more subordinate terms. This was interpreted to highlight higher levels of conceptualization of objects. In my analysis this level of conceptualization was expected to be reflected in a high frequency of definite noun phrases. As this expectation was not confirmed, the second factor, i.e. that of motivation, needs to be considered more closely.

In the present study, instructors assembled the house themselves and they were introduced to the person who they were to instruct. Thus, two factors can be identified that might account for the observed differences. First, participants in my study were provided with action-oriented prior information and second the motivation was different. In the present study, instructors knew the person they were instructing and the assembly was done while listening, hence the finished product could be evaluated by the instructor. Interestingly, participants who instructed the specific addressee and who knew the goal object verbally, seem to have assumed a listener who did not have any context specific prior information (uninformed listener) whereas instructors who did not know anything about their addressee, as in Kiefer et al. (1993), seemed to assume some prior information and possibly even background knowledge. This observation stresses that assumptions about the addressee's knowledge state need to be carefully considered and discussed in the interpretation of experimental findings collected in the referential communication paradigm and in comparisons between studies.

Confirming the hypothesis that culturally constrained objects were more frequently conceptualized, it was observed that one object, namely the roof part, was conceptualized significantly more explicitly in first reference than the other objects. Participants in self-assembly as well as instructors referred to this part by domain specific nouns significantly more frequently than to all other objects. However, it was pointed out to me that the cultural bias might have been strengthened by the upright position in which the roof part was placed on the table. This influence was not accounted for in the presented study but more caution will be paid to this possible confounding factor in future experimental studies. Despite this potential bias, the observation that instructors referred to this specific object functionally in the introductory phase indicates that instructors made use of general cultural background knowledge in their references. Although the general observation held for participants in the *underspecified goal* condition, the results also indicated that those participants referred to the object functionally least frequently. This finding suggests that prior information has an effect on function assignment even in cases in which cultural constraints strongly suggest a certain function.

Research question 2 investigated referential change in general. This process of change was then studied in more detail by focusing on the differences between the two independent variables of amount of prior information and communicative intention. Furthermore, it was of great interest to shed light on the influence of prior information on the time of object conceptualization. In general, the results revealed that repeating was the most frequent process in the *underspecified goal*, *verbal and visual goal* condition, and in instructions in first reference. Since repeating does not signal any change in conceptual state, it can be hypothesized that those participants did not change their concepts. However, if the second most frequent processes were considered as well, the data revealed that changes to familiarity were expressed in the *verbal and visual goal* condition and in instructions. Participants in the *verbal goal* condition tended to reduce initial references and some participants assigned functional terms to objects at this stage of the referential process as well. Thus, participants who were provided with verbal goal information showed the greatest change in conceptual state in first reference. Their referential choice indicated change towards familiarity and towards a domain specific conceptualization. In all subsequent references, repeating was most frequent in both studies indicating no conceptual change. Only slight changes were observed by change of referential form (rephrasing) in instructions.

The finding that *repeating* was more frequent than reducing carries an important implication because *reducing* signals a change in conceptual state whereas *repeating* signals maintenance of an expression and concept. Horton and Gerrig (2002) pointed out that consistency of expression is a principle of speech economy. This phenomenon can be interpreted from two perspectives, namely the speaker's internal production perspective and the addressee's perspective. Concerning the first perspective, Bock (1986:379) described that people used the same structure repeatedly in cases in which production of new structures would result in hesitations and pauses. Fukumura and van Gompel (2012) stated more explicitly that "repeating the recently mentioned word is usually easier than producing a word that has not been mentioned" (Fukumura & van Gompel 2012:1305). Horton and Gerrig (2002) observed a tendency for repetitions in instructions but they did not discuss the implications of this observation. Referring to data that was presented in Horton and Gerrig (2005a), Brennan and Clark (1996) proposed that instructors used the same structures to facilitate identification for the listener.

The presented data highlighted that if conceptual changes happened, it was early within the referential process, i.e. in first to second reference. Overall, great changes, as expressed in the category of *reframing*, were rarely identified. Nevertheless, the results revealed that participants who were provided with *verbal goal* information conceptualized almost half of all 13 objects (47.7%) from first to fourth reference and throughout the assembly process (*assembly time*). This number is a little higher than function assignment that was observed in instructions (45.1%). Instructors showed a tendency to reframe objects late in the instructions, i.e. in the fourth quarter of the instruction time. This observation supports the findings that instructors started marking objects as familiar only at a late stage in the referential process. Possibly this observation is an effect of participants' general knowledge about the generic structure of an instruction, i.e. assume familiarity of an object only after it was properly introduced. Given the similarity of overall conceptualization, the notion of the uninformed listener that was proposed in the discussion on referential form needs to be reassessed. The observed similarity in overall conceptualization suggests that instructors assumed their listeners to have general domain knowledge regarding components of houses as well as general knowledge about building toy houses.

Research question 3 investigated at which point participants conceptualized the objects within the goal domain, either for themselves or for an addressee. First, it needs to be noted that instructors and listener in the presented study were provided with two sources of information about the goal object: the experimental task instructions and the verbal instructions. In the experimental task instruction the experimenter introduced the goal object verbally but without highlighting it (for more detail see chapter 4). Furthermore, the instructor may provide the addressee with additional prior information. In order to assess how often the goal object was explicitly restated at the beginning of the instruction, introductory sentences of the instructions were analyzed at the content-level. Five out of 16 instructors restated the instructions saying that they would assemble a house[259] with the listener. This finding suggests that instructors noticed which background information was provided to their listeners and that they tailored their references accordingly. Recalling the discussion on referential form in instructions and combining it with the presented discussion, converging evidence was found for the proposal that instructors kept track of the shared discourse history and tailored their references specifically for the introduced listener.

[259] One instructor referred to a doll's house when starting the instruction by saying „wir bauen heut zusammen ein Puppenhaus auf" (we will assemble a dollhouse together today) [B17].

Participants who were provided with no specific goal information used references indicating conceptualization to a similar degree as participants who were provided with *verbal and visual goal* information (34.4% and 35.9% respectively). It can be assumed that participants who did not know anything about the goal object had difficulties in assigning function to the individual parts. Since they did not know anything about the nature of the goal object, they needed to create a mental image of possible objects first. This was expressed in the frequent use of domain unspecific description terms. This assumption is supported by the finding that there was little interpersonal variation with regard to instances of *reframing* in this condition. Interestingly, participants without information about the goal object assigned function by reframing most frequently in the second quarter of the assembly process. Participants who were provided with much prior information tended to reframe objects in the fourth quarter. This unexpected observation can be explained in two ways. First, those participants did not need to conceptualize the individual parts explicitly because they matched them with the provided external model mentally. This explanation is supported by the findings that participants in this condition used pronouns more frequently indicating that the objects were clearly represented in memory, as noted by Fukumura and van Gompel (2012:1306). The second explanation would assume that participants encountered difficulties in matching the provided objects to the remembered picture (external image) and thus drew on unspecific terms on the description level. This explanation can be supported by the observation that participants in this condition did not use domain specific nouns more frequently than other participants. Furthermore, they tended to explicitly assign function late in the assembly process (in the fourth quarter), which may indicate initial identification and classification problems. Based on the reviewed observations, it can be assumed that both explanations contribute to the findings.

To conclude, the hypothesis on the time of conceptualization was not confirmed because participants in all three conditions tended to reframe initial references more frequently than subsequent references. Furthermore, it was observed that participants who were provided with an external model tended to conceptualize objects later in the assembly process than participants who were provided with verbal goal information only or none at all. With regard to instructors, the expectations were confirmed regarding reference process time. However, contrary to the expectations, findings on instruction time revealed that instances of *reframing* were most often observed late in the instruction text. This finding mirrors the findings on referential form which highlighted that instructors used forms that signal familiarity only late in the referential process. Overall, the data seems to confirm that

> "internal representations are constructed, maintained and restructured along with changes that occur when assemblies are constructed from a variety of multi-functional construction objects." (Wachsmuth & Jung 1996:348)

Beyond this general observation, the presented analyses shed light on the time course of conceptual changes and the influence of prior information on referential change behavior. Furthermore, the presented data also highlighted differences between self-oriented and addressee-oriented speech. The discussed differences between think aloud data and addressee tailored speech support the claim, for example made by Ericsson and Simon (1993), that think aloud data reflect unaltered thought processes in the sense that no conscious changes are made for an assumed addressee.

Further specific features of language use in think aloud protocols have been identified with regard to the process of mapping the functional goal domain to the physically present object as investigated in research question 4. The majority of mapping instances was identified in protocols by participants who were provided with no prior information (*underspecified goal* condition). Participants who were told about the goal object and saw the picture (*verbal and visual goal* condition used explicit mapping least often.

Overall *direct mapping* was most frequent in all conditions suggesting that assemblers were certain in their assignment. Taking the overall distribution of mapping phrases into account, the results highlighted that participants with fewer background knowledge expressed more certainty in their assignment. This finding suggests that knowing the goal object and holding specific mental representations constraints the options for matching physical parts and function. It can be assumed that these participants had concrete expectations as to the number and shape of objects that would be presented to them. This interpretation is supported by the analysis of modal verb type that revealed that participants who were shown the picture used 'müssen' (must) most frequently. The choice of *must* signals a high certainty in function assignment. Furthermore, it implies that the certainty was based on some external source, possibly the memory of the picture. However, if the provided objects did not match the participant's expectations, function assignment might be difficult and therefore not done. Participants who were provided with unspecific goal information or a verbal clue seemed to be more flexible in their associations that were expressed in a higher number of mapping phrases. It seems that those participants expressed lower certainty by using modal verbs in their mapping phrases rather than by choosing *representational mapping* that signals an even lower level of certainty. These findings suggest that the amount of prior information has an influence on certainty of mapping as expressed in the frequency of mapping phrases and modal verb choice.

In addition to the amount of prior information, the data revealed a significant effect of communicative intention on mapping phrase structure as well. The comparison of think aloud protocols to spoken instructions revealed systematic differences in the reference term for the source entity, i.e. object domain, the strategies for mapping the domain specific function to the respective physical object, and the functional term that was assigned. The frequency analysis revealed that instructors used mapping phrases less frequently than assemblers did. There are two possible interpretations for this finding. First, it implies that instructors focus on the systematic procedure of the assembly rather than on the goal structure as a whole. Daniel and Tversky (2012) showed that instructors omit explicit information on object parts and sequential order when they were told to be brief but they did not omit information about the step-by-step action sequence. Thus, Daniel and Tversky (2012:318) concluded that “the critical information to include in instructions, then, is the sequence of actions on objects that users need to perform to correctly assemble the object”. These findings support the first interpretation. However, instructors might believe that functional terms would not facilitate object identification because the functional relation between the physical object and its position in the goal structure was not intuitive (von Stutterheim et al. 1993). Therefore, instructors might have concluded that functional terminology would rather result in additional cognitive load for their addressee. This interpretation seems reasonable given that the wooden objects did not suggest any prototypical usage unrelated to the specific context of this dollhouse assembly. These two interpretations are not exclusive but may both explain the observed behavior.

A significant effect of communicative intention was also observed on referential form to the physical object. In self-assembly the physical object is often referred to by deictic expressions. As pointed out before, this referential choice signals the assembler’s engagement with the physical world and his speaker centered perspective. Instructors used nominal phrases to refer to the physical object most frequently. This finding may suggest that instructors did not rely on the visual support in function assignment but rather used the strategy of verbally identifying the physical object before assigning its function.

In instructions, function is either assigned by one of the two strategies identified in think aloud protocols (*direct mapping* and *representational mapping*) or by using a third strategy, namely *reframing*. In *reframing* the referential term was elaborated by a functional term without a verb phrase in between. In most cases, hesitation markers separated the two references. This additional strategy of assigning function was identified post-hoc during the analysis. After the strategy was identified an additional example of this structure was identified in the literature, e.g. "da ist dieses Baufix, diese Baufixschraube" ("there is this baufix"[260], "this baufix screw") (Rieser 1996:10). In this example, a more explicit reference replaced the initial one without any interaction between instructor and addressee. The structure is similar to those cases of *reframing* that were identified in the presented assembly data only without a marker of hesitation. This kind of mapping is the least explicit one and it does not allow for any conclusions about certainty of assignment. Whereas assemblers tended to express certainty in function assignment, instructors also expressed tentative assignment by *representational mapping* and they used *reframing*. The use of *representational mapping* and *reframing* was not markedly different. This finding was surprising because instructors could have expressed a high level of certainty in their assignment because they assembled the house, i.e. goal object, themselves prior to the instruction task; an activity that should result in expert knowledge about the assembly and the objects' functions.

The use of *representational mapping* supports the interpretation that the objects were not prototypical for their specific function. Instructors encode this information by stating that an object represents the specific function in this specific context of the assembly. Possibly this is a strategy to implicitly ask the addressee for acceptance. The strategy of *reframing* from unspecific terms to goal specific terms can also be interpreted as a strategy designed for the addressee. This conclusion is based on the observation that only instructors used *reframing* and comparable phenomena are described in the literature on interaction. In communicative settings, instructors are observed to rephrase references when their addressee signals difficulties by hesitations or questions for clarification (e.g. Clark & Wilkes-Gibbs 1986; Horton & Gerrig 2002). However, although there was no interaction in the experiment, instructors still spontaneously felt the need to conceptualize the objects with regard to its function within the assembled goal structure. This observation can be explained by assuming that instructors have expert knowledge regarding the different functions but they adopt the addressee's

[260] Baufix is a brand name.

perspective. Since the addressee does not know about the function, the object is identified first before offering expert information about the object's function. This interpretation suggests that *reframing* is a communicative strategy that is specific to the goal of shared attention and object identification.

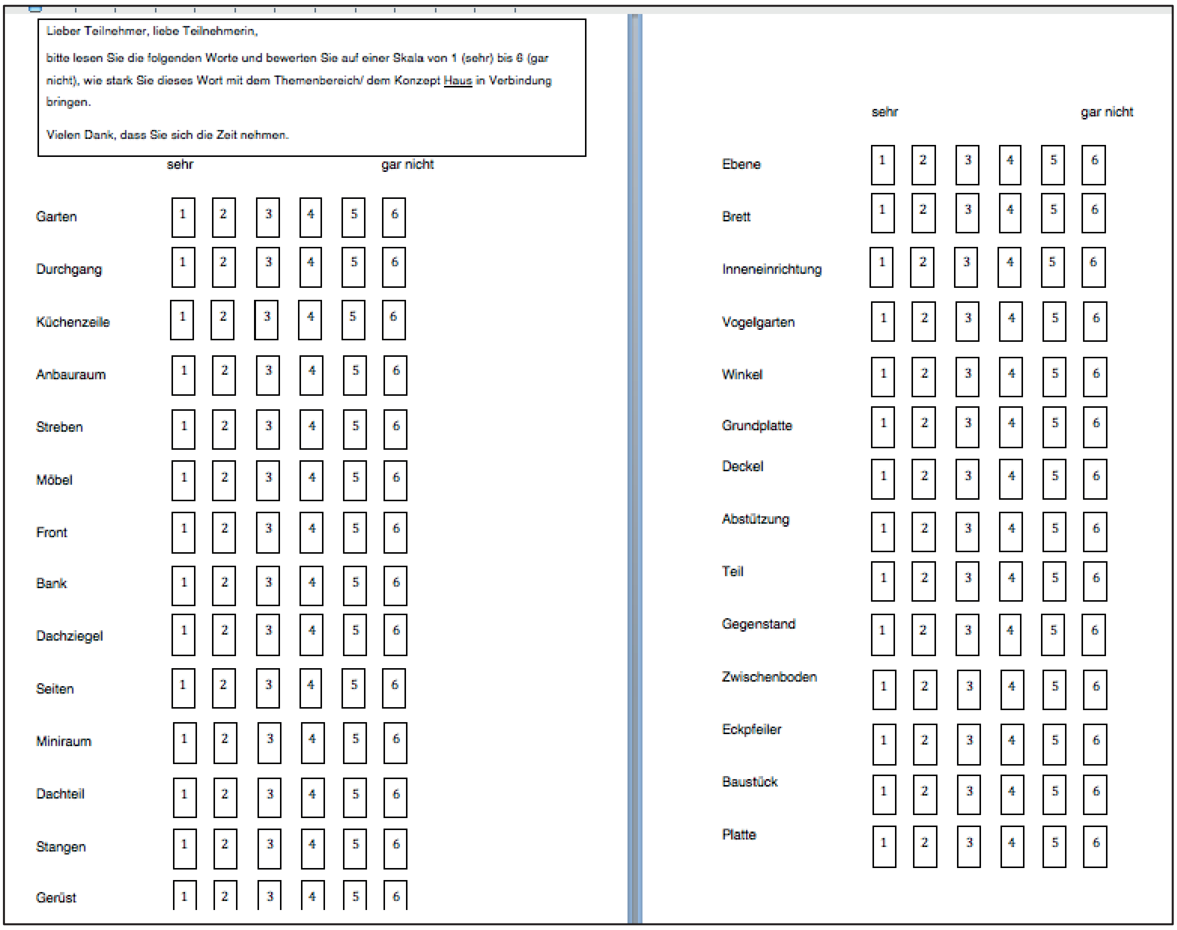
Additionally, a clear effect of amount of prior information as well as communicative intention on type of functional terminology assigned to the physical object was observed. Participants who were provided with *underspecified goal* information tended to assign terms that refer to structural parts of the goal object (e.g. 'room', 'story') or the goal structure itself ('house'). Instructors as well as participants who were provided with more explicit goal information, in contrast, tended to assign function of parts, such as 'wall', 'roof', and 'floor'. This finding allows for two readings. First, it suggests that the complete structure of the goal object was more prominent in the participants' minds who needed to construct the goal structure from scratch than in participants' minds who already held a mental representation. Apparently, those participants needed to conceptualize each individual part in relation to the overall structure. Second, participants who did not have precise information about the goal referred to already assembled object structures because they did not know the nature of the overall structure before conceptualizing the assembled parts. This interpretation suggests that the perception of the assembled parts leads to an insight regarding the overall structure or to the confirmation of the assembler's hypotheses on its nature. Participants who were provided with some information about the goal structure matched the functional parts that are commonly associated with a house to the objects at hand without the need to focus on the overall structure.

8.6 Conclusion

In sum, the detailed analysis of referential form to objects throughout the assembly highlighted the influence of prior information and communicative intention on referential choice. The findings highlight that participants who were provided with much prior information conceptualized objects as familiar earlier than participants who were not provided with an external model. This observation suggests that participants integrated the given objects into the remembered goal structure thereby conceptualizing them within the assembly context. Participants who were only told about the nature of the goal structure marked references as familiar within the given context later in the process. The finding that the one sole object that was prototypical, namely the roof part, was referred to as roof very frequently in all conditions suggests that cultural knowledge is activated while solving the task. These findings indicate that most of the provided objects were not prototypical for any specific function in the assembly of a house, thus participants needed to conceptualize them throughout the assembly. This is supported by the observation that instructors chose referential terms that did not assume any specific familiarity with the object in the given assembly context. They referred to objects in descriptive terms over consecutive references before they assumed some of them as familiar.

Taking a look beyond this chapter, the observations on referential form in think aloud protocols strengthen the findings reported in chapter 6 on problem solving processes. In this previous chapter, the content-based analysis highlighted that the specific experiment of object assembly investigated in this thesis presented participants with a difficult task that resulted in problem solving behavior. The analysis reported in this chapter contributed new insights on the difficulties of conceptualization of object parts within the specific domain of a 'house'. At the methodological level, the analysis highlighted that the analysis of referential form and referential change throughout the task should not be restricted to the first one to two instances of mention because in cases in which conceptualization is difficult, the moment of context specific interpretation would not be captured. Moreover, the methodological approach illustrated how phenomena of referential change as observed in interactive settings can be adopted and extended to describe phenomena of change in think aloud protocols, i.e. self-directed speech. Therefore, this analysis provides methodological tools for future analyses, such as categories of referential change and types of explicitly verbalized function assignment along with their distinct linguistic representation.

8.7 Appendix



Lieber Teilnehmer, liebe Teilnehmerin,

bitte lesen Sie die folgenden Worte und bewerten Sie auf einer Skala von 1 (sehr) bis 6 (gar nicht), wie stark Sie dieses Wort mit dem Themenbereich/ dem Konzept <u>Haus</u> in Verbindung bringen.

Vielen Dank, dass Sie sich die Zeit nehmen.

	sehr					gar nicht
Garten	1	2	3	4	5	6
Durchgang	1	2	3	4	5	6
Küchenzeile	1	2	3	4	5	6
Anbauraum	1	2	3	4	5	6
Streben	1	2	3	4	5	6
Möbel	1	2	3	4	5	6
Front	1	2	3	4	5	6
Bank	1	2	3	4	5	6
Dachziegel	1	2	3	4	5	6
Seiten	1	2	3	4	5	6
Miniraum	1	2	3	4	5	6
Dachteil	1	2	3	4	5	6
Stangen	1	2	3	4	5	6
Gerüst	1	2	3	4	5	6

	sehr					gar nicht
Ebene	1	2	3	4	5	6
Brett	1	2	3	4	5	6
Inneneinrichtung	1	2	3	4	5	6
Vogelgarten	1	2	3	4	5	6
Winkel	1	2	3	4	5	6
Grundplatte	1	2	3	4	5	6
Deckel	1	2	3	4	5	6
Abstützung	1	2	3	4	5	6
Teil	1	2	3	4	5	6
Gegenstand	1	2	3	4	5	6
Zwischenboden	1	2	3	4	5	6
Eckpfeiler	1	2	3	4	5	6
Baustück	1	2	3	4	5	6
Platte	1	2	3	4	5	6

Figure 8.20: Material categorization task.

domain unspecific (generic)		domain specific			constraints and percpetual object features	
super-ordinate	synonyms	*super-ordinate*	*synonyms*	*subordinate*		*synonyms*
Stück(e)			Stock(werk), Etage(n), Geschoss	Erdgeschoss	Einkerbungen	
Brett(er)	Platte	Wand	Häuserwände	Vorderwand	Eckpunkte	
Deckel				Querwand	Schräge	
Ding(er)				Trennwände für Toiletten-kabinen	Schlitze	Rillen

Doppelschienen				Innenwand	Löcher	
Ecke(n)				Seitenwände	Bolzen	
Teil(e)	Part			Zwischenwände	Stöpsel	
Sachen				Rückwand	Kerben	
Eckpfeiler				Trennwand	Schrauben	
Gegenstand	Gegenstände			Raumteiler	Stecker	
Kanten				Eckwand	Stift	
Gerade				Außenwand	Dübel	
Halbkugeln	Gnubbel, Noppen, Nubbsies, Pinökel	Basis	Grundfläche(n),	Decke, Fundament	Stöpsel	
Holz	Holzding			Fußboden	Playmobil	
Keil				Zwischendecke	Maßstab	
Seiten				Boden	Stabilität	
Stangen				Fenster	Verbindung	
Streben				Tür		
Abstützungen				Küchenzeile	Verstärkung	
Krempel	Zeug			Inneneinrichtung	Steckverbindung	
Materialien				Treppe	Vertiefung	
Sticker	Zeichen, Schild, Aufkleber, Etikett	Gebäude		Haus	Keil	
Punkte	Kontaktpunkte			Puppenhaus; Spielzeughaus		
Rest				Treppenhaus		
Konstruktur	Gerüst			Mehrfamilienhaus		

Bild				Traumhaus; Villa		
Regal				Balken		
Zweiteilung				Pfeiler		
Winkel		Dach		Dachschräge		
Ständer	Säulen			Dachziegel		
Träger	Stützen			(Dach)giebel		
Leiste		Raum	Zimmer			
Zweier		Garten				
Bühne		Möbel				
Füße	Füßchen					
Ganze						
Teilstrecke						
Rahmen						
Durchgang						
Bank						
Vogelgarten						
Anbauraum						
Miniraum						
Front						
Ebene						
Grundplatte						
Zwischen-boden						

Table 8.34: Overview nouns and their categorization.

8.7.1 Feature: nominal and pronominal phrase

condition	reference number	nominal phrase	pronominal phrase	no reference	numerals	ellipsis
A	1st	110 (49.8%)	50 (22.6%)	58 (26.2%)	3 (1,4%)	0 (0.0%)
A	2nd	59 (26.7%)	43 (19.5%)	119 (53.8%)	0 (0.0%)	0 (0.0%)
A	3rd	26 (11.8%)	32 (14.5%)	162 (73.3%)	0 (0.0%)	1 (0.5%)
A	4th	19 (8.6%)	13 (5.9%)	188 (85.1%)	1 (0.5%)	0 (0.0%)
A	5th	10 (4.5%)	11 (5.0%)	200 (90.5%)	0 (0.0%)	0 (0.0%)
A	6th	5 (2.2%)	8 (3.6%)	208 (94.5%)	0 (0.0%)	0 (0.0%)
A	7th	3 (1.3%)	4 (1.8%)	214 (96.8%)	0 (0.0%)	0 (0.0%)
A	8th	0 (0.0%)	3 (1.4%)	218 (98.6%)	0 (0.0%)	0 (0.0%)
B	1st	103 (52.8%)	48 (24.6%)	43 (22.1%)	1 (0.5%)	0 (0.0%)
B	2nd	43 (22.1%)	53 (27.2%)	99 (50.8%)	0 (0.0%)	0 (0.0%)
B	3rd	35 (17.9%)	22 (11.3%)	138 (70.8%)	0 (0.0%)	0 (0.0%)
B	4th	14 (7.2%)	12 (6.2%)	169 (86.7%)	0 (0.0%)	0 (0.0%)
B	5th	6 (3.1%)	6 (3.1%)	182 (93.3%)	1 (0.5%)	0 (0.0%)
B	6th	6 (3.1%)	3 (1.5%)	186 (95,4%)	0 (0.0%)	0 (0.0%)
B	7th	6 (3.1%)	0 (0.0%)	189 (96,9%)	0 (0.0%)	0 (0.0%)
B	8th	1 (0.6%)	3 (1.5%)	191 (97.9%)	0 (0.0%)	0 (0.0%)
C	1st	96 (49.2%)	73 (37.4%)	25 (12.8%)	1 (0.5%)	0 (0.0%)
C	2nd	54 (27.7%)	74 (37.9%)	67 (34.4%)	0 (0.0%)	0 (0.0%)
C	3rd	38 (19.5%)	37 (19.0%)	120 (61.5%)	0 (0.0%)	0 (0.0%)
C	4th	11 (5.6%)	25 (12.8%)	159 (81.5%)	0 (0.0%)	0 (0.0%)
C	5th	6 (3.1%)	17 (8.7%)	172 (88.2%)	0 (0.0%)	0 (0.0%)
C	6th	5 (2.6%)	6 (3.1%)	184 (94.4%)	0 (0.0%)	0 (0.0%)
C	7th	1 (0.5%)	2 (3.0%)	192 (98.5%)	0 (0.0%)	0 (0.0%)
C	8th	1 (0.5%)	0 (0.0%)	194 (99.5%)	0 (0.0%)	0 (0.0%)

Table 8.35: Results referential form, variable: nominal vs. pronominal form in UOA from first to fourth reference; raw frequency with percentage in brackets (A: underspecified goal condition, B: verbal goal condition, C: verbal and visual goal condition)

8.7.2 Referential process

condition	reference	no reference	no further reference	reducing	reframing	repeating	rephrasing	reusing	singel reference
A	2nd – 3rd	5 (2.3%)	42 (19.0%)	17 (7.7%)	9 (4.1%)	22 (10.0%)	7 (3.2%)	3 (1.4%)	116 (52.5%)
A	3rd – 4th	5 (2.3%)	66	4 (1.8%)	4 (1.8%)	13	11 (5.0%)	2 (0.9%)	116

			(29.9%)			(5.9%)			(52.5%)
A	$4^{th} - 5^{th}$	5 (2.3%)	76 (34.4%)	5 (2.3%)	6 (2.7%)	9 (4.1%)	1 (0.5%)	3 (1.4%)	116 (52.5%)
B	$2^{nd} - 3^{rd}$	0 (0.0%)	38 (19.5%)	10 (5.1%)	9 (4.6%)	22 (11.3%)	13 (6.7%)	4 (2.1%)	99 (50.8%)
B	$3^{rd} - 4^{th}$	0 (0.0%)	70 (36.1%)	4 (2.1%)	5 (2.6%)	10 (5.2%)	2 (1.0%)	4 (2.1%)	99 (50.8%)
B	$4^{th} - 5^{th}$	0 (0.0%)	83 (42.6%)	2 (1.0%)	4 (2.1%)	4 (2.1%)	2 (1.0%)	1 (0.5%)	99 (50.8%)
C	$2^{nd} - 3^{rd}$	2 (1.0%)	49 (25.1%)	10 (5.1%)	9 (4.6%)	35 (17.9%)	20 (10.3%)	4 (1.0%)	66 (33.8)
C	$3^{rd} - 4^{th}$	2 (1.0%)	88 (45.1%)	8 (4.1%)	2 (1.0%)	20 (10.3%)	7 (3.6%)	2 (1.0%)	66 (33.8)
C	$4^{th} - 5^{th}$	2 (1.0%)	102 (52.3%)	2 (1.0%)	1 (0.5%)	16 (8.2%)	5 (2.6%)	1 (0.5%)	66 (33.8)

Table 8.36: Distribution of categories of referential change within conditions in UOA in second to fifth reference; raw frequency with mean percentage in brackets. (A: underspecified goal; B: verbal goal; C: verbal and visual goal).

experiment	participant	instances of reframing	% reframing respective object
A	10	1	1,0
A	13	1	1,0
A	15	1	1,0
A	19	8	8,2
A	23	1	1,0
A	25	1	1,0
A	32	13	13,3
A	35	2	2,0
A	42	1	1,0
A	46	2	2,0
A	50	1	1,0
A	52	2	2,0
A	60	2	2,0
B	11	2	2,0
B	12	2	2,0
B	16	1	1,0
B	17	4	4,1
B	29	3	3,1
B	31	4	4,1

B	33	1	1,0
B	47	4	4,1
B	53	2	2,0
B	54	7	7,1
B	62	5	5,1
C	14	1	1,0
C	18	1	1,0
C	20	1	1,0
C	24	6	6,1
C	28	9	9,2
C	40	1	1,0
C	48	2	2,0
C	49	3	3,1
C	59	3	3,1
Instruction	2	4	7,7
Instruction	4	3	5,8
Instruction	6	6	11,5
Instruction	9	4	7,7
Instruction	12	1	1,9
Instruction	15	12	23,1
Instruction	17	15	28,8
Instruction	20	2	3,8
Instruction	21	5	9,6

Table 8.37: Overview of which participants used reframing and how often it was used in first to fourth reference in UOA (A: underspecified goal condition, B: verbal goal condition, C: verbal and visual goal condition, UOA: unaided object assembly) and in instructions; raw frequency and percentage of reframed objects.

8.7.3 Mapping process

dependent variable	condition	mean	standard deviation
mapping total	underspecified goal	3.12	0.81
mapping total	verbal goal	2.50	0.54
mapping total	verbal and visual goal	1.82	0.46
direct mapping	underspecified goal	2.59	0.71
direct mapping	verbal goal	1.94	0.37
direct mapping	verbal and visual goal	1.65	0.47

representational mapping	underspecified goal	0.53	0.17
representational mapping	verbal goal	0.56	0.27
representational mapping	verbal and visual goal	0.12	0.08
ellipsis	underspecified goal	0.06	0.06
ellipsis	verbal goal	0.06	0.06
ellipsis	verbal and visual goal	0.06	0.06

Table 8.38: Results of mapping types sorted by conditions in self-assembly (mean and standard deviation of raw frequency).

dependent variable	condition	mean	standard deviation
direct mapping	underspecified goal	62.40	10.49
direct mapping	verbal goal	71.88	9.94
direct mapping	verbal and visual goal	53.92	11.74
representational mapping	underspecified goal	14.07	6.24
representational mapping	verbal goal	15.63	7.21
representational mapping	verbal and visual goal	7.84	6.08

Table 8.39: Results of mapping types sorted by conditions in self-assembly (mean and standard deviation of percentage).

dependent variable	condition	mean	standard deviation
deictic source	underspecified goal	3.00	0.82
deictic source	verbal goal	2.50	0.54
deictic source	verbal and visual goal	1.76	0.43
deictic source	experiment self-assembly	2.42	2.54
deictic source	instruction	0.50	0.63
nominal source	underspecified goal	0.12	0.08
nominal source	verbal and visual goal	0.59	0.06
nominal source	experiment self-assembly	0.06	0.24
nominal source	instruction	0.94	0.85

Table 8.40: Results of nature of source entity in both experiments (mean and standard deviation of percentage).

dependent variable	condition	mean	standard deviation
deictic source	underspecified goal	69.12%	11.25
deictic source	verbal goal	87.50%	8.54
deictic source	verbal and visual goal	63.73%	11.81
deictic source	instruction	34.38%	11.02
nominal source	underspecified goal	7.35%	5.97
nominal source	verbal and visual goal	0.98%	0.98
nominal source	instruction	53.12%	11.71

Table 8.41: Results of nature of source entity in both experiments (mean and standard deviation of percentage).

Condition/Category source entity	deictic source	nominal source	ellipsis
underspecified goal	0.8	-1.7	-0.8
verbal goal	0.6	-2.3	1.0
verbal and visual goal	0.4	-1.5	0.6
instruction	-2.6	7.5	-0.9

Table 8.42: Standardized residuals for condition × category source entity.

9 Features of analysis: Constraints

9.1 Introduction

Traditional accounts of problem solving describe problem solving as a search through a problem space that is defined by the initial state, the goal state, and numerous intermediate states that are created by the application of operators (Newell & Simon 1972). Newly created states are evaluated with regard to their resemblance to the goal state. This evaluation process either results in the decision to proceed if the goal state is not reached or in ending the process if the goal state is reached.[261] Since the search is based on entities in the search space, the representation of this search space is crucial. As Kirsh (1996:448) concludes

> "with a good representation, a problem may be easy to solve, requiring little search, but with a bad representation, the problem may be almost

[261] See more on Newell and Simon's (1972) theory on human problem solving in chapter 2.

> impossible to solve in that form and may require inordinate amounts of search, calculation, and recall of states."

This is especially the case if information about the individual states is missing resulting in an incomplete search space.

Traditionally, psychologists tend to study special cases of problem solving such as playing chess (e.g. Newell & Simon 1972) or dynamic decision tasks such as running an imaginary tailorshop (e.g. Roth 1985). In these well-defined tasks[262] complex cognition can be investigated in planning processes and actions. Knauff and Wolf (2010) define complex cognition as involving thinking, reasoning, problem solving, and decision making under complex conditions. Complex conditions are characterized by numerous dependent and independent variables, such as norms, time, participants, and objects. Everyday activities also involve complex cognition but they represent ill-defined problems in most cases (Holland et al. 1996:11). Holland et al. (1996:12) specify that solving ill-defined problems "depends on the parallel activity of multiple pieces of knowledge that both compete with and complement each other in revising the problem representation".

There are a number of differences between problem solving tasks that are traditionally studied and every day problems. First, in contrast to traditionally studied problem solving tasks, daily problem solving does not involve much planning (Norman 2002) because mental simulation and planning pose high cognitive load on the problem solver. Thus, in order to perform the numerous daily tasks efficiently in a reasonable amount of time, memory load needs to be reduced to a minimum. The theory of minimizing cognitive load is supported by experimental research reported in Kirsh (1995; 2010) and Norman (2002). These studies highlight that people combine internal knowledge, external information, and constraints to come up with fast solutions for everyday tasks (Norman 2002:55) as well as newly encountered problems. Second, since everyday problems are ill-defined in nature, the problem solver is confronted with incomplete knowledge and uncertainty about constraints. Third, everyday problems involve a high level of activity.

This chapter provides a brief summary of problem solving in terms of embodied cognition. It will focus on one source of information that is deployed in problem solving by interacting with the environment, namely the use of object properties,

[262] In well-defined problems the initial as well as the goal state are known to the problem solver. Additionally, potential operators are known and need to be arranged in a meaningful way to solve the problem successfully. Ill-defined problems are characterized by missing information on initial state, goal state, operators, and/or constraints (e.g. Holland et al. 1996:11).

more specifically constraints. No definition of constraints has been proposed in the literature so far. However, based on the literature review, the following working definition is defined for this thesis. *Constraints* are devices that help to reduce the number of possibilities regarding the actions that can be performed in a specific context. As the field of embodied cognition is very young, not much empirical research has been reported so far. A general overview concerning the nature, functions, and use of constraints as reported in the literature will serve as a starting point for this explorative analysis. Based on previous findings and open questions, research questions and expectations will be formulated. After a description of the annotation procedure the results will be presented. The results of the explorative approach will be discussed in the light of the posed research questions in the final sub-section.

9.2 State of the art

Researchers surveyed people's problem solving behavior in their naturalistic environments and they found that people apply a variety of strategies for extracting useful information. One of those strategies is searching for affordances and constraints inherent in objects (Norman 2002). A second strategy is the use of cognitive artefacts, i.e. "physical objects made by humans for the purpose of aiding, enhancing, or improving cognition" (Hutchins 1999:126)[263]. Those artefacts may either be structural, i.e. remembering kinds of glasses to remember a list of drinks (Hutchins 1999), or physical, such as a knot in a handkerchief.

Another strategy that is connected to the concept of cognitive artifacts and that has been frequently observed is the arrangement and rearrangement of the external world in order to foster remembrance or gain new insights (Kirsh 1995; Kirsh 1996). The controlled manipulation of resources in the environment reduces demands on working memory, visual spatial memory, and visual search (Kirsh 1996:442). Hutchins (2005) refers to one of Kirsh's examples in which he describes that people arrange objects in a line during disassembly and when reassembling the object they assemble the objects as laid out in the line but in reverse order. This example illustrates a strategy in which an external structure is created, i.e. a

[263] At the end of the encyclopedia entry Hutchin's (1999:127) broadens this narrow definition by stating that "the cognitive artifact concept points not so much to a category of objects, as to a category of processes that produce cognitive effects by bringing functional skills into coordination with various kinds of structure." But in order to grasp the general concept of cognitive artifacts, the narrow definition as quoted above is equally well suited.

line, in order to remember the sequential order of assembly of specific objects, i.e. a sequence of actions and objects needed. Hutchins (2005) argues that this creation and use of cognitive artefacts represents cases of conceptual blending resulting in material anchors.[264] Whereas conceptual blending involves much cultural knowledge, such as lines representing temporal sequences, the basic strategy of rearrangement can already be used effectively by an 18 month-old child. Kirsh (1996) reports the example of a little girl who solves her puzzle by remembering the spatial locations of the individual objects pictured on the pieces (e.g. the cat in the upper right). In order to test her problem solving behavior, Kirsh (1996) rotated the puzzle by 180 degrees and observed her behavior again. After two unsuccessful tries and one seemingly accidental correct placement the little girl rotated the entire board by 180 degrees and started reusing her initial placement strategy. According to Kirsh (1996), this example illustrates that taking actions to reduce mental effort can be an effective problem solving method.

Observations such as the little girl rotating her board support the claim proposed by researchers in the field of situated and embodied cognition. They state that human cognitive abilities such as thinking, remembering, and understanding a language are not based on the manipulation of abstract symbols in the mind but rest on an interactive process between humans and the environment (e.g. Zhang & Norman 1994; Hutchins 1995; Pecher & Zwaan 2005; Kirsh 2009; Vallée-Tourangeau & Wrightman 2010). Kirsh (2009) points out that although the use of externalized representations is acknowledged and theoretically discussed by observing "that the environment provides organization for cognitive activity, that the world enables and supports such activities" (Kirsh 2009:303) there is a need to explain the mechanisms underlying the interaction between internal and external processes. In this claim Kirsh summarizes the core objective of situated cognition. The fact that it has been formulated by Colleen Seifert (1999) in a review on situated cognition ten years before illustrates that there is still much work to do in order to understand human cognition as being shaped by an interactive process between the individual and the environment. One basic assumption postulated in embodied cognition and supported in experimental research is that "perception and action are considered central to higher cognition" (Pecher & Zwaan 2005:2); not solely being input and output devices. Steffensen (forthcoming) takes this assumption as the starting point for his concept of *interactivity* which he defines as

[264] Material anchors are created "if conceptual elements are mapped onto a material pattern in such a way that the perceived relationships among the material elements are taken as proxies (consciously or subconsciously) for relationships among conceptual elements" (Hutchins 2005:156).

"sense-saturated coordination that contributes to human action" (Steffensen 2013:195). In this line of thought perception is proposed to take place while acting and action is performed while perceiving, thus constituting "embodied and emotional action-perception cycles" (Steffensen 2013:199).

Researchers in the embodied cognition community argue that the interaction between the environment and the individual cannot be adequately observed in data collected in laboratory studies because the environment is organized by the experimenter. In this controlled setting participants are not likely to rearrange the environment or ask for additional material such as pen and paper if it is not provided. But the attempt to collect data in natural settings poses a number of challenges. First, the data collection is time consuming, e.g. as described in Hutchins (1995). Second, challenges are presented by the analysis of the data. These challenges arise from the nature of everyday problems that are structurally different from experimental tasks. They are not well-structured since "human life unfolds in an indefinite problem space" (Steffensen 2013:196). Therefore identifying the interesting and important aspects in the data needs expertise and time. Steffensen (2013) and Kirsh (2008) present different methods to tackle the problem. Whereas Kirsh (2008) focuses on mathematical measures to assess cost and benefit, Steffensen (2013) presents a method that describes action and interaction on a very detailed and fine-grained level. Steffensen (2013) illustrates how the method of *Cognitive Event Analysis* can be used to analyze data collected "in the wild"[265] to gain "a more realistic view on human problem-solving" (Steffensen 2013:218).

Cognitive Event Analysis is designed for analyzing video and audio recordings featuring one or more people working out a problem in their natural environment, e.g. in an office. In a very fine grained analysis it combines the analysis of body movements (such as gaze, body posture and orientation, motor movements), language (content, intonation, and timing), and interaction (eye contact, verbal). Steffensen (2013) argues that this work-intensive data analysis is necessary because only by looking at the different modalities of interaction between people and their environment it seems possible to reveal the mechanisms of real-world problem-solving.

[265] Hutchins (1995:xiii) stresses that these studies need to focus on "human cognition in its natural habitat – that is, to naturally occurring culturally constituted human activity".

The great amount of modalities that are analyzed results from the assumption that a Distributed Cognitive System[266] (hereafter DCS) is composed of at least five components, namely persons, technologies[267], artifacts[268], procedures[269], and narratives[270] (Steffensen 2013:201). Steffensen (2013) states that the DCS makes use of artefacts "because of their material affordances (…), or because human beings interpret them as giving off information" (Steffensen 2013:202). As described above, it is the latter feature that drives the use of constraints. Therefore, constraints can be considered to be one member of the category 'artefacts'.

In a recent study, Steffensen (2013) applied this method on a video-taped interaction between two office mates. His analysis highlighted that problem-solving in an ill-defined problem scenario did not proceed in an analytical or logic-based linear structure as assumed in traditional problem solving theories, such as Newell and Simon's (1972). The observed problem solving process was rather a creative, near-chaotic solution-probing process. In this process both participants generated hypotheses on possible solutions and those were probed "until they, *post festum*, observed that one of the probes fits the problem space" (Steffensen 2013:31).

Encouraged by these findings a number of researchers (e.g. Hutchins 1995; Clark 2008; Kirsh 2009) conclude that 'realistic problem solving' can only be studied in more naturalistic studies, studies 'in the wild' at best. Hutchins (1995: xiv) argues that only in the everyday world it is possible to observe how human cognition adapts to its natural surroundings. The experiment that is investigated in this thesis does not meet the latter criterion because it has been run in an experimental setting. However, in the presented study, participants are recorded assembling a two-story dollhouse in three scenarios that are equally likely to occur in real life. The assembly task exhibits the following characteristics of everyday problems ((a) and (b)) and interactive problem solving ((c) and (d)):

[266] Steffensen (2013:199) defines a Distributed Cognitive System as "a self-organising entity that arises as human beings co-engage through interactivity, and connect up brains, bodies and aspects of the environment".

[267] This thought is in line with Kirsh (2010) who suggests that using technology on external representations helps to foster new insights and reduces cognitive load.

[268] The notion of cognitive artifacts has been proposed by Hutchins (1999) first.

[269] The notion that procedures are an important way of interaction between the environment and the mind has also been stated by Kirsh (1995).

[270] Narratives may help to constrain possible behavior by providing "folk-theoretical schemes for understanding causality, ethics, emotions etc." (Steffensen 2013:203).

(a) the task is action oriented,

(b) it encourages few planning processes because objects can be rearranged,

(c) the objects can be used as a model, and

(d) thinking with things is encouraged because no constraints regarding the arrangement and possibility to disassemble parts are posed by the task instructions.

The opportunity to manipulate physical objects externally changes mental processes because some object features are recognized by interacting with them only. Knowing about the characteristics of transformation of objects is important for imagining specific object arrangements or activities that represent intermediate states on the way towards the final goal state (Klix 1976:27).

9.2.1 Functions and use of constraints

Three basic strategies that are used by people to solve problems in interaction with the environment have been presented in the previous sub-section, namely externalization of thoughts, rearrangement of the environment, and the active search for affordances and constraints inherent in objects (Norman 2002). Kirsh (1995:43) explains that recognition and meaningfulness of affordances[271] depends on individuals, their experiences, and their knowledge. Contrary to Kirsh's (1995) reasoning, Norman (2002:82) observed that participants with different prior knowledge, i.e. experts and novices, used the same methods and strategies when solving everyday tasks. Participants were observed to make use of different constraints in experimental settings. These constraints were provided by the individual object parts, the situation, and culturally learned assumptions about objects. Norman (2002) argued that affordances help to reduce cognitive load in everyday activities.

Norman's (2002)[272] assumption about the facilitating effects of constraints by extending the mind towards the outside world is in the line of thought of researchers in distributed and embodied cognition (e.g. Agre & Chapman 1987; Kirsh 1995; Clark 2008). Kirsh (2009), more specifically, discussed the function of hints in the context of problem solving. He defined hints as "verbal and nonverbal

[271] Kirsh (1995:43) defines affordances to be "a dispositional property of a situation defined by a set of objects organized in a set arrangement, relativized to the action repertoire of a given agent. Agents perceive an affordance, when they register that one of their possible actions is feasible in a situation".

[272] Note that Donald A. Norman presented his arguments in 1988 already, the first year of the publication of "The design of everyday things". Hence he was among the first researchers to formulate this theory.

cue[s] that act like a heuristic bias on search" (Kirsh 2009:291). In their function as heuristic bias, they are important in the process of candidate generation and evaluation. This proposal draws on a thought that Kirsh formulated in 1995 already by stating that "prima facie, choice is the product of search – visual search for the actions that are currently available, and mental search of the desirability of those available actions" (Kirsh 1995:43).

Norman (2002), Kirsh (1995), and Steffensen (2013) investigated how people made use of external information in order to perform and act efficiently. Norman (2002) approached the topic by observing the process and results of the design of everyday things in which constraints are deliberately designed as clues for the user. Kirsh (1995; 1996; 2010) investigated how people actively used space to structure thought, cue the temporal order of actions, limit the possibilities of actions, and focus the mind on certain objects or actions. In a recent study, Steffensen (2013) observed that the physical manipulation of material artifacts supported the problem solving process by establishing an external focus of attention for two people who were engaged in a task (Steffensen 2013:19). This analysis, as outlined in the previous section, highlighted that this new focus of attention extended and restructured their problem space.

To sum up, the literature review illustrates that external cues, such as constraints, help to reduce the amount of possibilities and thus generate a manageable amount of solutions by constraining necessary decisions. People have been observed to actively search for cognitive artifacts, such as constraints, because they have experienced that these artifacts provide important information, e.g. about spatial arrangements.

9.2.2 Nature of constraints

Focusing on the nature of constraints, Norman (2002) distinguished between physical, semantic, cultural, and logical constraints. These constraints were assumed to determine the possibility of performing certain actions. *Physical constraints* describe properties, which make certain actions impossible thereby reducing the number of possible actions to be performed. For example, a certain amount of holes allows for the exact same number of screws to be inserted and fastened thereby limiting their number and possible arrangement of objects. Norman (2002:84) points out that physical constraints are specifically valuable because no special training is required to use them efficiently.

Semantic constraints are bound to the situation, which may allow only specific actions to be performed. Norman's example (2002:85) refers to his experiments in which he asked participants to assemble a lego motorcycle. In this specific scenario there is only one proper location and orientation for the policeman, i.e. he needs to sit on the seat facing front. This example illustrates that semantic constraints are based on world knowledge and knowledge about the situation thereby providing powerful and important clues about the task (Norman 2002:85).

Cultural constraints are based on learned and accepted cultural conventions, such as handling screws. The rotation of screws for fastening and loosening them is conventionally bound, i.e. screw clockwise to fasten and counterclockwise to loosen. Furthermore, people conventionally tend to hide the bolts in the less visible parts of an object, such as bottom, side, or interior, whereas the screws are shown in the front or on top of a piece (Norman 2002:62). The concept of cultural constraints, as proposed by Norman in 1988, is supported by Hutchins' (2011) notion of cultural practices. Cultural practices are defined as "emerging products of dynamic distributed networks of constraints" (Hutchins 2011:441). By these constraints they shape the way in which people perceive, i.e. see, hear, smell, and taste, the world "by highlighting what to attend to and what to see when attending" (Hutchins 2011:441). Constraints may either be mental and internal, physical as arising from the mechanics and physiology of the body or provided by the engagement with material artifacts, or mental and physical as arising in the interaction with other people (Hutchins 2011:441). Throughout his paper Hutchins (2011) stresses the great influence of cultural practices on people's situated actions. *Logical constraint*s result in straightforward decisions: "only one piece left, only one possible way to go" (Norman, 2002:86).

Norman (2002) defined these classes of constraints based on observations in handling everyday devices, such as computers, VCR, or film projectors, and by observing participants assembling a lego motorcycle from 13 pieces. Throughout his book Norman (2002) illustrated that choices in operating devices and in assembly tasks were guided by these constraints that rule out specific actions and make others more likely.

Considering the experimental set-up investigated in this thesis a fifth kind of constraint is proposed, namely *task induced constraints*. This category includes constraints that are imposed on the task by the instructions given to the participants. Possible candidates for this category can be identified by an analysis of the task instructions. If the instructions state that all objects need to be used, this presents the constraint that no object can be left. If the instructions state that the task needs to be solved in ten minutes, this statement represents a time constraint.

Based on the reported observations that people engaged in a problem solving task make extensive use of the environment, this part of the thesis investigates the use of object specific constraints in self-assembly. A fine-grained content based analysis of references to constraints in the elicited verbal reports will serve as a starting point for studying the nature, function, and use of constraints in unaided object assembly. Furthermore, these three features of constraints will also be examined in the instruction texts to investigate which constraints were evaluated to be important clues for others. After introducing the investigated research questions in the next section a description of the analysis procedure and the different categories that were used in the analysis are provided. Based on the reported findings, a discussion will highlight the answers that can be gained by this exploratory analysis.

9.3 Research questions and expectations

First, emerging research questions that can be investigated by analyzing the given data set will be formulated. Second, expectations that are raised by findings and theories reported in the literature review will be stated.

The five research questions address two different fields of research; four research questions[273] investigate questions concerning the nature and function of constraints in the specific experimental context (RQ 1 and 3-5) whereas one question is raised by the theory of embodied cognition (RQ 2).

- Research question 1: Which constraints are mentioned? Can Norman's (2002) classification be adapted to classify them?

This initial question concerns the general nature of constraints that can be identified in the specific context of the dollhouse assembly. Looking at the individual object features representatives for each of Norman's (2002) categories can be identified. Norman (2002:82;224) reports one experiment that can be compared to the one presented in this thesis in two respects. First, participants were also presented with a small amount of objects, i.e. 13 lego parts, that did not suggest any prototypical function. Second, they were not given any information

[273] At some points within this sub-section the term research question is abbreviated by RQ for reading comfort.

about the final product, i.e. a motorcyle. Therefore, this experiment is comparable to the *underspecified goal* condition in the experimental design in this thesis.

The following constraints are assumed to be noticed and mentioned by participants assembling the dollhouse. Based on Ericsson and Simon's (1984 (1993)) *theory of verbalization of thinking*[274], a distinction needs to be made between noticing an object and mentioning it in the verbal reports. Constraints might be noticed but this perception might not be conscious and thus it is not mentioned in the think aloud protocol. Furthermore, perception might be conscious and thus verbalized in concurrent verbalization but it might be forgotten at the end of the task and thus it is not reported in the retrospective report. To gain as much information about the perception of constraints as possible participants were asked guided questions about noticing and using specific constraints at the end of the experimental session (for more detail on the procedure see chapter 4). In the following paragraphs all theoretically possible constraints are introduced.

The bevel that characterizes each object belonging to the second story (see Figure 9.1 and see Figure 9.2) needs to be interpreted as a *physical constraint* because it is needed to securely position the roof part on top of the second story (see Figure 9.3). It is assumed that the great majority of participants notices the bevels because they are very salient. Furthermore, it is assumed that their function as *physical constraints* is noticed.

[274] For more information on their theory see chapter 3.

Figure 9.1: Object that goes on the right hand side at the front of the board in the second story.

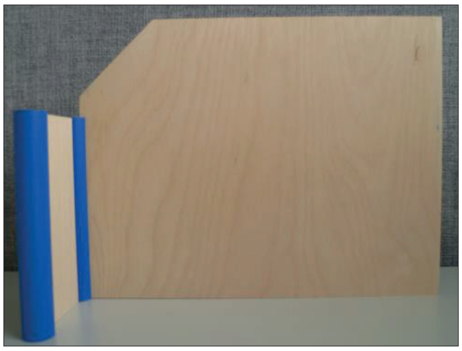


Figure 9.2: Object that goes on the left hand side at the back of the board in the second story.

Figure 9.3: The second story of the dollhouse.

There are boreholes in the boards (see Figure 9.4) and at the bottom of the objects. Normally screws would be inserted into these boreholes to connect and fasten the objects. These boreholes can be interpreted as *semantic constraints* because they suggest the spatial arrangement of objects on the board (e.g. see Figure 9.9 on page 396). They are not defined as physical constraints because the dollhouse can be successfully assembled without using them as a guide. It is expected that boreholes will be noticed by most participants but as they do not serve their normal function, i.e. holding screws, it is expected that they will be mentioned only rarely.

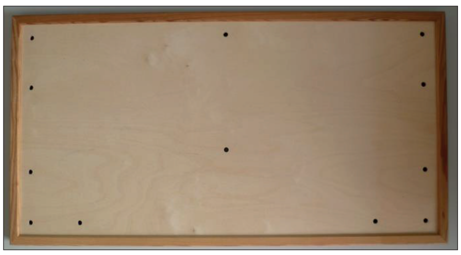


Figure 9.4: One of the two boards that need to be used as floors.

Two *cultural constraints* can be identified in the dollhouse assembly, namely the roof part and the label attached to it (see Figure 9.5). The roof clearly suggests a house whereas the label 'Selecta Spielzeug' suggests that it is a children's toy. If participants are very attentive, the label may provide a clue for a dollhouse. It is assumed that the roof part will serve as a salient clue for building a house. The label is expected to be noticed and mentioned only seldom because it does not have any facilitating effect of the assembly.

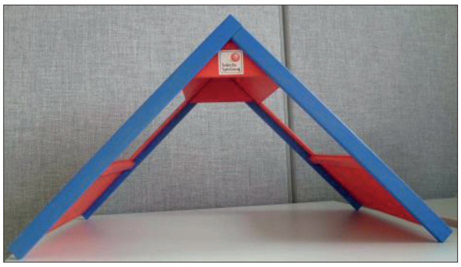


Figure 9.5: Roof part of the dollhouse.

Additionally, the objects provide two kinds of *logical constraints*. The first one is the groove (see Figure 9.6) that is characteristic of the two objects functioning as middle parts on both stories (see Figure 9.7). If it is used, the two boards in the back can be inserted on both sides of this object part (see Figure 9.8). If this is done, the back row of the story is stable even without screws. The use of a groove can be assumed to be cultural knowledge because it is a technical device that is commonly used in constructions. Therefore, it is assumed to be frequently mentioned if it is noticed.

The second logical constraint is the supporting system (see Figure 9.10) that can be found on top of each object that belongs to the first story (see Figure 9.9). If the objects are placed with the supporting system facing up, as shown in Figure 9.9, it helps to stabilize the second board on top of them. Both of these constraints help to stabilize the dollhouse structure but they are not necessary to assemble a sensible structure or a house. However, as these are very salient features, it is expected that participants will frequently refer to the supporting system.

Figure 9.6: Object that is placed in middle position between the two larger objects in the second story.

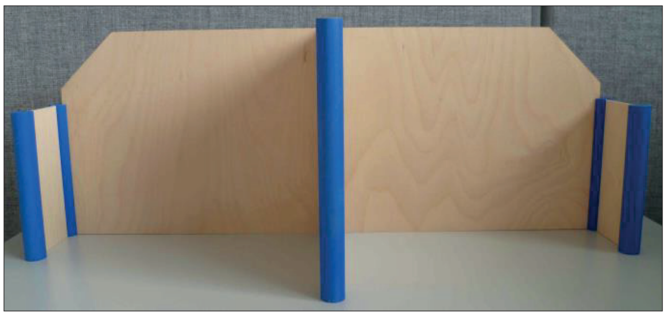


Figure 9.7: Arrangement of the back row of the second story.

Figure 9.8: Close-up of the left hand side of the arrangement of the back row of the second story illustrating the function of the groove.

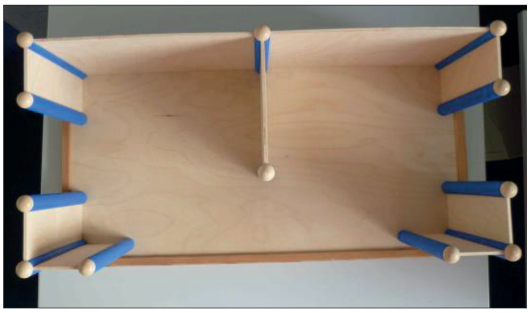


Figure 9.9: First story of the dollhouse looked at from a bird's-eye view.

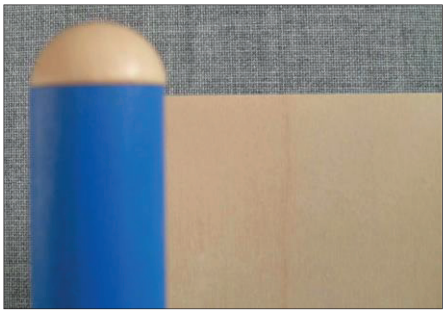


Figure 9.10: Close-up of the supporting system that is found on each object belonging to the first story.

Two additional constraints can be inferred from the instruction in the specific scenario, i.e. *task induced constraints*. First, the instruction explicitly states that no screws can be used, thus participants should not search for connecting devices. Furthermore, this implies that stable connections need to be made by different means. Second, in one condition (*verbal and visual goal* condition) a picture is shown to participants to visualize what the goal object looks like[275]. Matching the objects to the remembered picture limits the possibilities concerning their function and spatial arrangement. Therefore, it is expected that participants will mention the picture or their memory of the pictured dollhouse throughout the assembly using it as a guide.

- Research question 2: In which part within the assembly process do participants mention constraints?

Two temporal dimensions need to be distinguished to formulate expectations regarding this question, i.e. the overall assembly process and the problem solving process as described by its process categories, such as *hypothesis* and *action* (see chapter 6). If the overall assembly process is considered, constraints are likely to be mentioned at all times. This expectation is based on the proposed embodied and emotional *action-perception cycle* (Steffensen 2013) which states that action is

[275] For more details on the design see chapter 4.

accompanied by perception and vice-versa. Therefore, object features may be noticed at any time during the assembly. Nevertheless it might be possible that the amount of prior information has an influence on noticing specific constraints, such as the bevel. Possibly, participants in the *verbal and visual goal* condition notice them earlier because they match new objects with an already existing external model. It can be assumed that the function of the bevels within the goal structure becomes clear to those participants at an earlier stage in the assembly process. Thus they mention them early on. Participants who do not have any prior information, in contrast, may notice object features early in the assembly phase because they may take more time for inspection in the beginning in order to get a good overview of the objects that are available. Moreover, they may be more attentive to object features because they are looking for information to guide the assembly.

Considering the specific structure of problem solving processes, constraints are likely to be mentioned in 'description of object features' (for more detail on the identified problem solving processes see chapter 6. Assuming that constraints are identified as clues regarding function and location of objects, associations about them are likely to be verbalized in 'hypotheses'.

- Research question 3: How does the discovery of constraints influence the problem solving process?

Generally, constraints are assumed to limit the number of possibilities (Kirsh 2009) and provide information about the assembly structure. The discovery and understanding of constraints as clues regarding function and position of objects is expected to foster their strategic use by some participants. For example, it might be possible that participants use the boreholes and match the objects according to the boreholes on their bottom part. However, it is equally plausible to assume that participants do not use constraints if they cannot be used with regard to their conventional function, such as 'boreholes hold screws'. Additionally, participants may use physical object features for sorting objects.

- Research question 4: Which constraints serve which function in the specific assembly context?

Previous theoretical considerations and to date limited experimental research on the function of constraints do not allow for directed hypotheses. Nonetheless, it can be assumed that some constraints are specifically likely to foster a certain function. Assuming that constraints observed in the assembly scenario are used to limit the number of possibilities, the perception of grooves is likely to foster function assignment. Drawing on the notion of cultural practices (Hutchins 2011) the function of grooves can be assumed to be well known to anyone who has assembled furniture before. The assignment of a function to stabilize the arrangement of objects may be fostered by noticing and conceptualizing bevels. Additionally, salient object features, such as the supporting system and the bevels, are likely to be used to distinguish between objects.

- Research question 5: Which constraints are evaluated as specifically helpful and are used in instructing other people on the same assembly?

Assuming that instructors use constraints for the same strategic reasons as assemblers, for example for sorting and guidance concerning spatial arrangements, the following constraints are expected to be mentioned. First, instructors are assumed to use the boreholes as markers for indicating the spatial layout of the assembly product. Second, instructors are expected to use the supporting system and the bevel as characteristic features for distinguishing between the given objects.

9.4 Analysis

This section introduces the procedure developed for analyzing the collected data in order to identify constraints, define them, and describe their use as well as their influence on participants' behavior. Four different data sets were analyzed to answer the research questions posed above, namely think aloud protocols, retrospective reports, instructions, and guided questions in which participants were questioned about noticing and using specific constraints (for more detail see chapter 4). Table 9.1 provides a general overview on the research questions, necessary analysis, and the data sets that are investigated.

research question	necessary data analysis	data set
Which constraints are mentioned? Can Norman's classification be adapted to classify them?	identify, describe, define, and classify nature of constraints (task specific and general classification)	think aloud data, retrospective reports
In which part within the assembly process do participants mention constraints?	analysis of process episodes containing discourse units in which constraints are mentioned and analysis of temporal sequence of constraints across the whole assembly phase	think aloud

How does the discovery of constraints influence the problem solving process?	identify, describe, define, and classify function of constraints	think aloud, retrospective report, guided questions
Which constraints serve which function in the specific assembly context?	frequency analysis of categories describing the nature of constraints within categories describing the function of constraints	think aloud data, retrospective reports, instruction texts
Which constraints are evaluated as specifically helpful and are used in instructing other people on the same assembly?	identify the constraints that are mentioned and their function in the text	instruction texts

Table 9.1: Summary of general procedure and the data sets that were investigated to answer the research questions posed above.

Before the analysis is described in detail, it shall be discussed which information can be gained from the four data sets to answer the research questions. As outlined at various points throughout this thesis, think aloud protocols contain the participant's verbalized thoughts during the assembly. Regarding constraints these protocols highlight which object features are noticed. If they are considered to be important and meaningful, they will be re-mentioned throughout the assembly process. Therefore, two mental states can be distinguished. Constraints are either verbalized as they are noticed and attended to or they are mentioned because they are used, i.e. they serve a specific function within the assembly process. The strategic use of constraints can be studied in participants' retrospective reports because those may contain more meta-information as participants may start reflecting on their behavior in retrospect.

In order to assess if participants noticed constraints that the author considers important, i.e. the groove, the boreholes, and the label 'Selecta Spielzeug', few guided questions were posed after the retrospective report. These answers may highlight that constraints have been noticed but were not mentioned in the think aloud protocol because the participant did not evaluate them to be important or facilitating in the specific assembly context.

The recorded instructions allow to assess the importance and function that instructors assign to individual constraints in order to facilitate the assembly for their partner. As instructors assembled the dollhouse in an explorative assembly phase themselves, these functions may reflect the functions identified in think aloud protocols. If this were the case, the results would allow for more general conclusions about the nature and function of the identified constraints.

9.4.1 Qualitative Analysis: Content based

9.4.1.1 Think aloud protocols and instructions

The first step of the analysis was done in the general annotation process as described in chapter 8. In this initial annotation process the category *constraints* was annotated according to the assumed constraints introduced and defined above. Furthermore, as this annotation was based on content analysis other instances in which participants referred to object features functioning as constraints were identified and annotated as well. In a second step, all annotated instances of constraints identified in think aloud protocols and instructions were combined in one table specifying the participant's ID, the condition, the corresponding discourse unit, and the discourse unit content itself (see Figure 9.11 for an example).

	A	B	C	D
1	TN	C	DU ID	DU
2	8	C	66	ach deswegen hatte der auch so eine Einkerbung
3	8	C	75	und dann hier dieses auch mit so ner Einkerbung [?]
4	8	C	34	und dann sind hier noch mal solche Schrägen
5	8	C	54	die hier ham solche Einkerbungen da
6	8	C	68	okay weil die hier solche Einkerbungen hatte

Figure 9.11: Screenshot of first step preparing for the analysis; 1st column participant ID, 2nd column condition (here *verbal and visual goal* condition), 3rd column ID discourse unit, and 4th column discourse unit, i.e. actual verbalization.

Then the discourse unit's content was coded concerning the nature of the specific constraint. In more general terms, classes of constraints were defined according to Norman's (2002) classification and each constraint was coded in this respect as well. The first classification of constraints, as identified in the data, is referred to as *task specific nature* because these categories are specific to the dollhouse assembly task. The second classification that is based on Norman's (2002) categories is referred to as *general nature* because these categories are not dependent on the investigated task. In a third step, the process category for the respective discourse unit was extracted from the annotation scheme for process categories (for more detail see chapter 6) and the number of overall process units in the protocol was recorded. In a fourth step, the discourse units were analyzed with regard to functions that constraints serve in the participant's assembly. One participant stated „die hier ham solche Einkerbung" ("those ones here have such grooves") indicating that he/she noticed the grooves (see Figure 9.12). The individual descriptions were later classified in order to define more general categories of function. The example above was classified as *perception* because the participant simply mentions the object feature.

The analysis of occurrences on constraints within the problem solving process is based on the identified problem solving episode for the respective discourse unit. In order to investigate the occurrence of constraints within the overall assembly process, the number of discourse units is divided into quarters and the discourse unit is defined within the respective quarter.

	A	B	C	D	E	F	G	H	I	J	K
1	TN	C	DU ID	DU	Constraint be	contraint allge	constraints (c	type constrain	Process categ	Funktion	# DU
2	8	C	66	ach deswegen hatte der auch so eine Einkerbung	Einkerbung	Einkerbung	groove	logical	Hypothesis	insight	82
3	8	C	75	und dann hier dieses auch mit so ner Einkerbung [?]	Einkerbung	Einkerbung	groove	logical	action	perception	
4	8	C	34	und dann sind hier noch mal solche Schrägen	Schräge	Schräge	bevel	physical	Hypothesis	perception	
5	8	C	54	die hier ham solche Einkerbungen da	Einkerbung	Einkerbung	groove	logical	comment on obj	perception	
6	8	C	68	okay weil die hier solche Einkerbungen hatte	Einkerbung	Einkerbung	groove	logical	comment on obj	insight	

Figure 9.12: Screenshot of final annotation scheme for think aloud protocols; 1st to 4th column cf. previous figure, 5th column constraint as verbalized in discourse unit, 6th column general category of constraint, 7th column English category, 8th column type of constraint, 9th column process category of discourse unit, 10th column category function, and 11th column sum discourse units in think aloud protocol.

The same annotation scheme was used for instructions. The categories that were defined for think aloud protocols could be adapted for the majority of annotations. If instructions contained new functions, additional categories were defined. The annotation scheme is identical except for the missing column specifying the process category because the problem solving dimension of instructions was not investigated (see Figure 9.13).

	A	B	C	D	E	F	G	H	I	J
1	TN	Cond	DU ID	DU	Constraint be	contraint allge	constraint (en	type constrain	Funktion	# DU
2	2	instructi	37	<- wie du auch siehst haben die hier so eine Rinne eine Fuge	Rinne	Einkerbung	groove	logical	description	125
3	2	instructi	45	<- und jetzt kannst du auch die Sperrholzplatten in diese Fugen / reintun	Fugen	Einkerbung	groove	logical	placement	
4	2	instructi	73	<- dass wir jetzt nur noch so Teile mit einer Schräge haben	Schräge	Schräge	bevel	physical	distinguish	
5	2	instructi	8	dass du auf der rechten Seite zwei Löcher vor dir hast	Löcher	Löcher	hole	semantic	orientation	

Figure 9.13: Screenshot of the final annotation scheme for instructions; 1st to 8th column cf. previous figure, 9th column category function, and 10th column sum discourse units in instruction.

9.4.1.2 Retrospective report

All utterances that contain references to constraints in the retrospective reports were extracted and annotated. The extracted references were annotated with regard to the task specific nature of the constraint as well as to its function as outlined above (see Figure 9.14 for an example). The previously identified categories were used and new categories were defined if necessary.

	A	B	C	D	E	F	G
1	TN I	Cond	Utterance	Constraint	Constraint	Constraint	Function
2	11	B	dann habe ich diese abgeschraegten Teile gesehen / und da dachte ich mir natuerlich das die irgendwie zu dem Schraegdach gehoeren muessen also ins obere Stockwerk	abgeschrägte Teile	Schräge	bevel	insight
3	11	B	diese blauen Rundhoelzer solche Ritze haben da muss man irgendwie Bretter reinschieben koennen	Ritzen	Einkerbung	groove	insight
4	11	B	und hab gesehen gibs ja nochmal sowas und das muss es fuer das obere Stockwerk auch geben	nochmal sowas	Gleichheit	analogy	insight

Figure 9.14: Screenshot annotation of retrospective reports; 1st column participant ID, 2nd column condition (here *verbal goal* condition), 3rd column utterance in which constraint is mentioned, 4th column constraint as mentioned in the utterance, 5th column constraint task specific category German, 6th column constraint task specific category English, 7th column category function constraint.

9.4.1.3 Guided questions

The answers to the guided questions were annotated as follows (see Figure 9.15 for an example). Each question contained two sub-questions. The first sub-question asked if the constraint was noticed. The second sub-question asked if the constraint was used as a guide in the assembly. The answers to the first sub-question were coded as noticed or unnoticed. The answers to the second sub-question were coded as 'no', 'yes', 'partially', 'irritated', or 'uncertain'. Additionally, interesting comments about the respective constraint were collected in case that they might be relevant for a future analysis.

	A	B	C	D	E	F	G	H	I	
1	TI	co	question 1:	question	question	questior	questior	comment groove	comment borehole	comm
3	13	A	noticed	no	noticed	no	unnoticed		boreholes were not used since there was no connector	
4	15	A	noticed	no	noticed	no	unnoticed		ja die hab ich richtig das äh stimmt da hab ich ja auch noch drüber nachgedacht ob ähm die Bohrlöcher halt eine Bedeutung haben und ob man da irgendwas reinstecken kann aber ich hab nichts gefunden was in der Größe passen würde	
5	19	A	noticed	irritated	noticed	yes	noticed bu	ähm joa zumindestens hatte mich ihr [?] n bisschen irriitiert weil das irgendwie ähm ich weiß nicht also also es hat mich bei der Anordnung glaub ich gestört also wenn [?] zumind ja vielleicht hat ich mehr Freiraum irgendwie insofern dann zu sagen ähm ich steck überhaupt irgendwas zusammen anstatt drei vier Teile zu haben aber also es hat mir glaub ich wenn ich wenn ichs nicht ähm wenn ich das nicht zusammenstecken könnte hätt ich glaub ich schneller irgendwie ne Anordnung gefunden denk ich weil so gabs mehr Möglichkeiten das anzuordnen also da wars irgendwie so ich hätte jetzt irgendwie ne andere Raumteilung vornehmen können oder so und das war mit wenns ähm mh also mit also ohne stecken wärs unmöglich gewesen dann hätte man feste ja feste Elemente gehabt und dann ähm wär garnicht die Frage gewesen wo man jetzt ne Grenze setzt [?] glaub ich weniger ne Chance	die Bohrlöcher die hab ich auch wahrgenommen ja also erst hab ich angefangen die immer falschrum aufzusetzen	
6	21	A	noticed	uncertain	noticed	yes	noticed & used		sogar eher als das Stecksystem	

Figure 9.15: Screenshot annotation scheme guided questions; 1st column participant ID, 2nd column condition, 3rd column answer to first part of question 1, 4th column answer to second part of question 1, 4th column answer to first part of question 2, 5th column answer to second part of question 2, 5th column answer to question 3, 6th column interesting remarks about the groove, 7th column interesting remarks about the boreholes, 8th column interesting remarks about the label.

Example (1) exemplifies the category 'partially', example (2) exemplifies the category 'irritated', and example (3) exemplifies the category 'uncertain'. All three examples are answers to the first question, concerning the use of the groove.

(1) „teilweise ja aber teilweise nein weil also es diente auf jeden Fall als Orientierung dass da die Rillen drin sind und dass man eben diese also bei dieser Kante halt ähm die Möglichkeit hat das da tatsächlich draufzustecken" ("partly yes but partly no because well it helped as a guide for sure that those grooves are there and that one has the possibility to fasten these edges in there") [50]

(2) „ähm joa zumindestens hatte mich hier n bisschen irriitiert weil das irgendwie ähm ich weiß nicht also also es hat mich bei der Anordnung glaub ich gestört" ("uhm well at least I was a little irritated because somehow uhm I don't know well well it interfered with the arrangement I think") [19]

(3) „ein wenig nicht wirklich weil dieses an der Seite also so wie ichs zusammengebaut hab bleibt ja ein Stecksystem frei an beiden Seiten und das hat mich dann doch n bisschen verunsichert" ("a little but not really because at the sides well at least the way I assembled it leaves a connecting system unused at both sides and this unsettled me a bit") [23]

9.4.2 Quantitative Analysis

The different categories that were identified in the qualitative analysis for describing the nature and function of constraints in the different data sets will be analyzed by descriptive as well as inferential statistics. First, raw and mean frequencies will be reported to observe the general distribution of the individual categories in each data set. If categories are adequate to describe findings across data sets, these will be combined to evaluate the distribution in a larger data sample. To highlight if specifically salient differences are statistically significant, inferential statistics such as the chi-square test are reported.

9.5 Results

As outlined in the previous section there are numerous different steps of identification, description, and categorization involved in answering the research questions. In order to present the data in a coherent and comprehensive way, this section is divided into a number of sub-sections. At the broadest level the results are distinguished into quantitative results and qualitative results. In section 9.4.1 the identified categories for describing the nature and function of constraints are presented and defined. This presentation distinguishes between the different data sources. The categories describing the nature of constraints are reported for think aloud protocols, retrospective reports, and instructions. The described functions of constraints are investigated in those three data sets as well.

In the section on quantitative results (9.4.2) the distribution of these different categories within the different experimental conditions is reported. In addition to mere frequency accounts the results of inferential tests such as the chi-square test are reported. Three sub-parts are distinguished here, namely *nature and function of constraints* from *occurrence of constraints*, and *answers to guided questions*. In the section on *occurrence of constraints* their distribution in the individual problem solving processes and across the whole assembly will be reported.

However, before the results are presented in the different sections a general overview on the number of protocols (think aloud and retrospective), instructions, and guided questions that are analyzed will be provided.

9.5.1 Details on analyzed data sets

50 think aloud protocols were analyzed. In ten of those protocols no constraints were mentioned; more precisely in five protocols in the *underspecified goal* condition, two protocols in the *verbal goal* condition, and 3 protocols in the *verbal and visual goal* condition. In the remaining 40 protocols 160 constraints were referred to. The participant who used most constraints referred to 18 constraints while thinking aloud. On average, participants assembling the house for themselves referred to constraints 3.2 times (SD = 3.7). Participants in the *verbal goal* condition (M = 3.6 times, SD = 4.4) did so frequently whereas participants in the *underspecified goal* condition referred to constraints less frequently (M = 2.6 times, SD = 3.1). Participants in the *verbal and visual goal* condition (M = 3.4 times, SD = 3.7) were closer to those participants who were provided with some information about the nature of the goal object (see Figure 9.16). In all three conditions one outlier was observed. The one in the *verbal goal* condition deviated markedly from the mean, i.e. with 18 references to constraints.

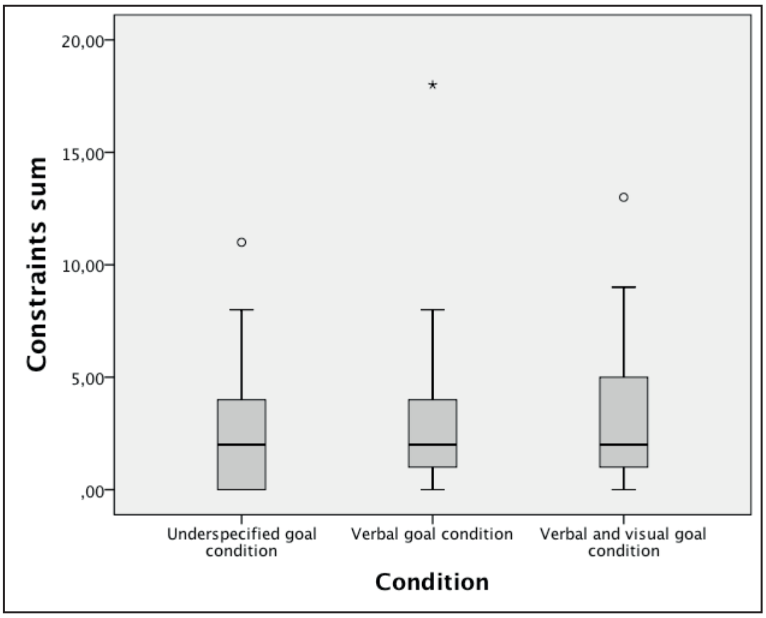


Figure 9.16: Plot sum of constraints mentioned in think aloud protocols; distinguished by condition.

Corresponding to the 50 think aloud protocols, the 50 retrospective reports of those participants were analyzed. Eleven participants did not mention any constraints in their retrospective report; four protocols in the *underspecified goal* condition, six protocols in the *verbal goal* condition, and one protocol in the *verbal and visual goal* condition. One participant did not refer to constraints in his think aloud protocol and did not mention any constraints in his retrospective report. All other participants mentioned constraints in at least one of the verbal reports.

The remaining 39 participants mentioned 111 constraints in total. Constraints were referred to up to eight times with an average of 2.22 constraints in retrospective reports (SD = 2.1). Participants in the *underspecified goal* condition referred to constraints least often (M = 1.6, SD = 1.5) and participants in the *verbal and visual goal* condition most frequently (M = 3.2, SD = 2.4). This difference was statistically significant, $t(27) = -2.41$, $p = .023$. Participants in the *verbal goal* condition (M = 1.8, SD = 1.9) referred to constraints less often than participants with very specific information. this difference did not reach statistical significance, $t(30.4) = -1.89$, $p = .069$. Although a great range can be observed in all three conditions, no outlier was identified in either of them (see Figure 9.17).

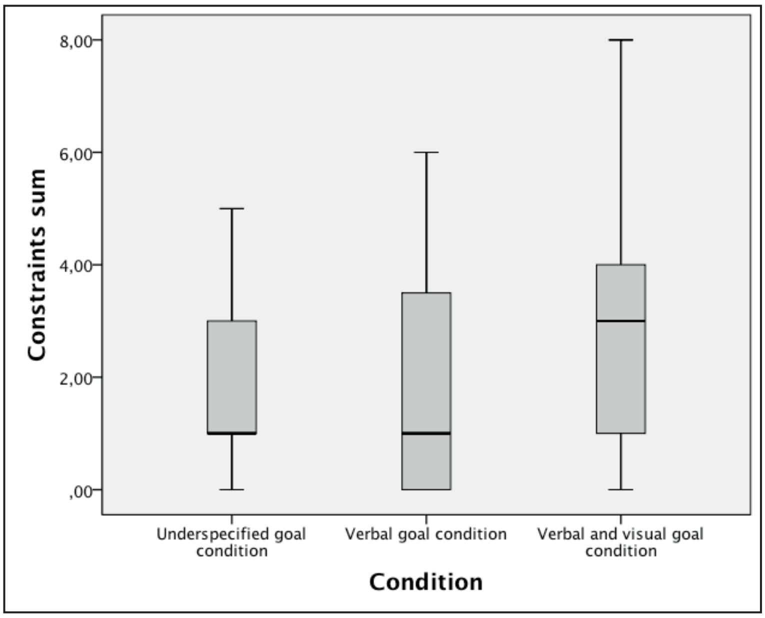


Figure 9.17: Plot sum of constraints mentioned in retrospective reports; distinguished by condition.

16 instructions were analyzed. Those instructions contained 136 references to constraints. All instructors referred to constraints at least once and the instructor who used constraints most frequently referred to constraints 20 times in his instruction. Instructors referred to constraints 8.5 times on average (SD = 5.7). Although there is a great range no outliers was identified (see Figure 9.18).

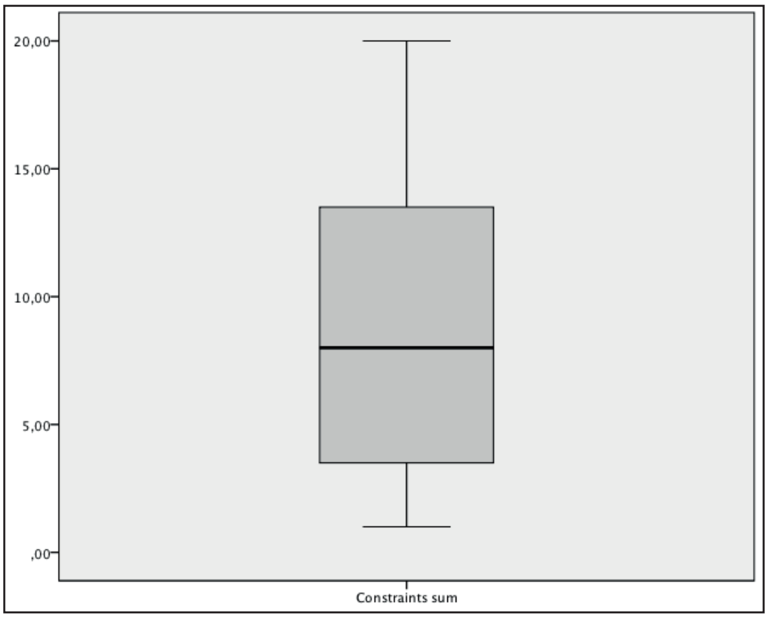


Figure 9.18: Plot sum constraints mentioned in instructions.

9.5.2 Qualitative analysis

This section provides an overview of the categories that were identified to describe the nature of constraints and their reported function. As outlined when stating the research questions and in the section of analysis, the nature of constraints can be defined with respect to two dimensions, namely with respect to the task specific dimension and the general dimension. In the presentation of the results the task specific nature will be presented first and then the general nature in terms of Norman's (2002) classification will be considered. The reported nature and function of constraints were analyzed in the think aloud protocols first, thus those results are reported in the first sub-section. Then the adapted and novel categories for the instructions are presented. Finally, the constraints that were identified in retrospective reports will be outlined.

9.5.2.1 Nature of constraints: task specific and general

9.5.2.1.1 Think aloud protocols

The analysis of the think aloud protocols revealed nine categories of constraints (see Table 9.2 for a summary). Participants used different words to refer to constraints, especially for those for which no conventional name exists, such as the supporting system. The identified categories contained the assumed constraints and two additional ones. The category *connections* contains references to objects that were either physically present (example (4)) or that participants searched for (examples (5) and (6)).

(4) „scheint mir so ne Art Verbindungsteil zu sein" (it seems to me that this is some kind of a connecting piece) [10] [276]

(5) „wo finde ich Verbindungsteile? " ("where do I find connecting pieces") [10]

(6) „ist ähm doch irgendwo n Verbingsstück? " ("is there uhm some connecting piece somewhere?") [54]

Participants referred to connectors either when noticing that no such connecting devices were provided (example (7)) or when searching for them (example (8)).

(7) „es gibt wohl keine Stecker" ("apparently there are no connectors") [16]

(8) „ich bräuchte irgendwelche Bolzen oder so" ("I would need some bolts or so") [24]

Whereas references to connectors referred to connecting devices such as screws, connections were rather general in nature. Connections did not relate to a physical object but referred to the concept of a connection, i.e. something that connects to objects.

[276] The numbers in square brackets denote the participant's ID from whom the example was extracted.

constraint (English category and German correspondent)	references used in protocols[277]	definition	previously expected or unexpected
label (Aufkleber)	Aufkleber, Schild, Etikett, Zeichen	Label 'Selecta Spielzeug' (see Figure 9.19)	Expected
supporting system (Auflagerundungen)	Noppen, Halbkugeln, Kontaktpunkte, Knubbel, Nubbsies, Kugel, Noppen, Pinökel	Small half ball like wooden pieces used to rest the next board on them (see Figure 9.10).	Expected
groove (Einkerbung)	Einkerbungen, Schlitze, Rillen, Kerben, Vertiefung, Verbindung, Steckrille, Fugen	In one rectangular piece with pillars on the sides there is a long narrow cut on each side of the pillar. These grooves are there to hold the two adjacent walls (see Figure 9.6 and Figure 9.8).	Expected
picture (Foto)	Bild, Farbfoto	Picture shown to the participant in the *verbal and visual goal* condition.	Expected
borehole (Löcher)	Löcher, Punkte	Either the boreholes in the large boards or the boreholes on the bottom of the wooden pieces[278].	Expected
bevel (Schräge)	Schräge	All parts belonging to the attic have bevels on the sides.	Expected
connector (Stecker)	Bolzen, Stöpsel, Schrauben, Stecker, Stift, Dübel	Something that connects the different parts to the base.	Unexpected

[277] Since these are specific German terms that do not have direct English correspondences no translations will be provided.

[278] The second option is hardly referred to hence both options are subsumed into one category.

connection (Verbindung)	Verbindungsteil, Verbindungsecken, Steckverbindung, Stecksystem, Verbindungsstücke	Connections that can be made. In this context it refers the function that grooves and connectors have within the assembled building.	Unexpected
no screwing (kein Schrauben)	ohne Schrauben	The statement that no screws are provided.	Expected
others	Playmobil, Maßstab, Stabilität, Verstärkung, Gedanken	Category that contains all references that cannot be classified in any of the other categories.	

Table 9.2: Constraints identified in think aloud protocols: examples and definitions.

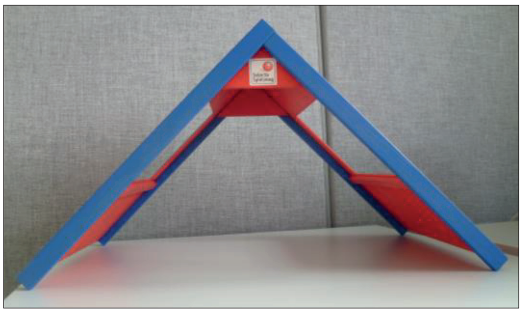


Figure 9.19: Roof part with the label 'Selecta Spielzeug' on it.

9.5.2.1.2 Instructions

Five constraints were identified in instructions. All of them could be classified according to the same categories as those identified in think aloud protocols. Sometimes the nouns that were used to refer to the individual constraints were a little different. There were no references to the label or missing connecting devices (see Table 9.3 for a summary).

constraint (English category and German correspondent)	references used in protocols	definition
supporting system (Auflagerundungen)	Knubbel, Pinökel	Small half ball like wooden pieces used to rest the next board on them.
groove (Einkerbung)	Einkerbungen, Fräsnaht, Fuge, Kerbe, Naht, Nut, Reingefräst, Rillen, Rinne, Schiene, Schlitz, Spalte	In one rectangular piece with pillars on the sides there is a long narrow cut on each side of the pillar. These grooves are there to hold the two adjacent walls.

borehole (Löcher)	Löcher, Punkte, Bohrlöcher, Lochbohrung	Either the boreholes in the large boards or the boreholes on the bottom of the wooden pieces.
bevel (Schräge)	Schräge	All parts belonging to the attic have bevels on the sides.
connection (Verbindung)	Steckverbindung	Connections that can be made. In this context it refers the function that grooves and connectors have within the assembled building.
others	Ecken	Category that contains all references that cannot be classified in any of the other categories.

Table 9.3: Constraints identified in instructions: examples and definitions.

9.5.2.1.3 Retrospective report

Nine constraints were identified in retrospective reports. Seven categories could be adapted from the previous categorization in think aloud protocols and instructions. Additionally, participants noticed the similarity between objects as well as between object arrangements. They reported that they used this observation for sorting and assembling the same arrangement again. Therefore, this category was called *analogy*. Furthermore, participants reported that they noticed the roof part and that it suggested building a house or that it was recognized from the picture. An overview on the different categories with examples is provided in Table 9.4.

constraint (English category and German correspondent)	references used in protocols	definition
supporting system (Auflage-rundungen)	Halbkugeln, Kuhlen, Holzdinger, Kugeln, Pinöpel, Knubbel, Gnubbsies	Small half ball like wooden pieces used to rest the next board on them.
groove (Einkerbung)	Einkerbungen, Ritzen, Rillen, Nuten, Schlitze	In one rectangular piece with pillars on the sides there is a long narrow cut on each side of the pillar. These grooves are there to hold the two adjacent walls.
borehole (Löcher)	Löcher, Markierung	Either the boreholes in the large boards or the boreholes on the bottom of the wooden pieces.
bevel (Schräge)	Schräge, abgeschrägte Teile, abgeschrägte Dinger, schräge Seiten, Teile ohne Ecken, schräge Platte, Schnittkante	All parts belonging to the attic have bevels on the sides.

connection (Verbindung)	Verbindungsstück, Stecksystem	Connections that can be made. In this context it refers the function that grooves and connectors have within the assembled building.
roof (Dach)	Dach, rotes Teil	The roof part is recognized as such and associated with a house that needs to be assembled.
analogy (Gleichheit)	ähnliche Teile, gleiches Schema, gleiche Teile, genauso wie unten, jedes Teil doppelt	Similarity between objects and object arrangements is recognized and strategically used.
others	Puppenhaus, Querstreben, Größe, Ecken, Platten	Category that contains all references that cannot be classified in any of the other categories.

Table 9.4: Constraints identified in retrospective reports: examples and definitions; new categories are highlighted.

9.5.2.2 Function of constraints

Three data sets were investigated in order to study the function of constraints, i.e. think aloud protocols, instructions, and retrospective reports. The findings for each of them will be presented in separate sections.

9.5.2.2.1 Think aloud protocols

The verbalized use of constraints by participants who thought aloud while assembling the dollhouse was classified according to five categories. In some cases participants referred to a constraint, such as the picture, when drawing general conclusions, such as „auf dem Bild sah das gar nicht so schwer aus“ (“it did not look that difficult on the picture”) [14] (*statement*). Some participants verbalized seeing an object feature but did not conceptualize it as a constraint (*perception*). For some participants the recognition of a specific object feature led to insights about its function within the assembly structure (*insights*). Some participants made strategic use of constraints, e.g. by recognizing the boreholes and using them as guidance as to where objects should be placed (*strategy*). At the other end of the continuum some participants were left puzzled about the function of perceived features (*puzzlement*). Table 9.5 provides a definition for each function with examples; here the examples are translated into English but the German original is reported in Table 9.32 in the appendix.

category	definition	examples
insight	understanding the function of a feature; assigning function to an object;	“oh that is why it has such a groove here” [2] “it seems to me that this is some

	no connectors are provided.	kind of connecting piece" [10] "apparently there are no connectors" [10]
strategy	recognizing and making use of object features such as holes or grooves; using world knowledge to search for things that may be useful.	"that they are on top of the holes" [14]; "those have such grooves here so that I can slide in here" [23] "I would need some bolts or so" [24]; "where do I find connecting pieces?" [20]
perception	object features are perceived; objects and its assigned function are perceived.	"and then this one here with such a groove" [8] "oh and here is a nice connecting system unused" [17]
puzzlement	asking oneself about the function of a perceived feature or placement of an object.	"those knobs do they need to stand on the bottom?" [10]; "and what are these holes good for?" [44]
statement	stating a fact; evaluative statement about an object referring to the memorized picture; repetition of task instructions.	" it didn't seem that difficult on the picture" [14] "but it looks similar to the picture" [27] "without screwing/ without screws" [31]

Table 9.5: Classification of functions identified in think aloud protocols; examples are translated.

9.5.2.2.2 Retrospective reports

Eight different functions could be identified in the assemblers' retrospective reports, ranging from perception of constraints on the one end of a continuum to strategic use of them on the other end. Some participants described that they observed specific object features (*description*) or simply mentioned noticing them (*perception*). Some used these features to distinguish between objects (*distinguish*). For some participant the function of constraints became evident during the assembly (*function assignment*) and in some cases constraints fostered insights about the assembly (*insight*). In contrast to those participants, some were left puzzled about the function of constraints (*puzzlement*). In some cases constraints were used strategically either by consciously disregarding them, using them for sorting objects, or using them for orientation as where to put objects (*strategy*). Table 9.6 provides an overview of the identified categories with some examples for illustration; again the German original wordings are reported in the appendix (see Table 9.36).

category	definition	examples
description	object features are described; general thoughts about observed features are reported.	"I noticed these half ball like pieces ('Halbkugeln') here" [16] "then I thought about how it could work without screwing ah without screws because there are these holes in the boards" [31]
perception	features are reported to have been observed.	"because I saw that there are ah holes for screws" [21] "but prior to that I saw the edges" [27]
distinguish	object features are used to distinguish between objects.	"the I noticed that uhm some boards are beveled and those I thought uhm and the others straight" [25] "afterwards I placed these uh pieces without bevels" [31]
function assignment	function is assigned to objects by means of constraints; function of constraints is explained.	"this red roof for example reminded me of a roof" [10]; "well I noticed pretty fast that this has bevels here, that this needs to belong to the roof" [18] "it is difficult to arrange them differently if one wants to put these connecting pieces in between" [19]; "the I understood pretty fast that the beveled sides need to go on top in order for the roof to be fastened and that the whole thing has two-stories" [23]
insight	conclusions about objects are based on perceived constraints.	"then I understood that the pieces that I assembled in the lower part probably need to be placed in the upper part because those are beveled because the shape would fit then" [47]; "but the walls were not beveled and the roof would not have fit on top therefore it was logical that the walls with the balls on the columns needed to go down" [59]
puzzlement	constraints cannot be conceptualized and leave the assembler puzzled.	"I was irritated by those balls here because I thought they are so nicely curved they need to be placed somewhere on the top" [20]; "and along the way I asked myself why some are curved here and others are not" [22]; "ah I was irritated that some have such balls on top and others don't" [50]
strategy	strategic use of constraints by consciously disregarding them; using them for orientation e.g. in putting & arranging	"then I put the picture out of my mind a little" [20] "then I was guided by those balls ah those balls on top and thought ok that is they cannot be placed on top thus I know what is up and what is down" [24]; "stuck them together in a way that the holes fit" [26]; "at some point the

	objects; using them to sort objects.	insight came something yes right I sorted it according to those holes and looked which fits where" [32] "right then we tried to separate between these beveled walls and ah balls" [16]; "and next I at these bevels I noticed them next and then I thought all right this needs to be connected to the roof and sorted uhm the pieces that need to belong to the second story and to the bottom one" [24]

Table 9.6: Categories of functions identified in retrospective reports: definitions and examples (in English translation).

- Think aloud and retrospective reports

The overview in Table 9.6 highlights that the majority of functions that were described for think aloud protocols were also to be found in reports about the assembly, i.e. *insight*, *puzzlement*, *strategy*, *perception*. Furthermore, three additional categories were identified in retrospective reports. One newly defined category was very similar to that of *perception*, namely *description*. The distinction between perception and description was based on the observation that some discourse units did not contain verbs of perception (e.g. *see*, *notice*, or *recognize*), thus they were defined as *description*. The other two novel categories were more interesting in terms of the analysis of the influence of constraints on the assembly process. In their retrospective reports participants mentioned that they used constraints to *distinguish* between objects and they remembered that constraints helped them to *assign function* to objects or arrangements of objects. Two classes of functions can be distinguished with regard to the level of conceptualization that they express. Whereas perceiving and describing constraints signals the *level of sensing*, all other functions involve an active use of constraints which involves the *level of interpretation* (see Table 9.7 for a summary).

data set	class of perception at the level of sensing	class of usage at the level of interpretation	others
think aloud	perception	strategy, insight, statement	puzzlement
retrospective reports	perception, describe	strategy, insight, distinguish, function assignment	puzzlement

Table 9.7: Summary categories defining the function of constraints mentioned in think aloud protocols and retrospective reports distinguishing between *class of perception*, *class of interpretation*, and *others*.

9.5.2.2.3 Instructions

The functions of constraints that were observed in instructions differ from those described in think aloud protocols but they are similar to those identified in retrospetive reports. Whereas participants assembling the dollhouse verbalized noticing an object feature, instructors described object features (*description*). Pointing out object features served different functions in instructions, namely distinguishing objects, assigning function, and orienting objects. In order to distinguish objects, their salient features were stressed (*distinguish*). In some cases instructors described an object before assigning its function within the goal structure (*function assignment*). Some features, such as the *boreholes* in the boards, were used to coordinate the position of the object (*orientation*). Similarly, object features, especially boreholes, were used to describe the spatial arrangement and placement of other objects (*placement*).

Some instructors told the partner about their own difficulties during the assembly (*comment about self*). Similar to those participants assembling the dollhouse for themselves some instructors pointed out the strategic use of specific constraints, such as *boreholes* or the *groove* (*strategy*). Table 9.8 provides definitions and examples for each category; the German original wordings are reported in the appendix (see Table 9.39).

category	definition	examples
description	a) object features are pointed out b) the current assembly status is described	a) "as you can see those have a groove here" [2] b) "there are no bevels positioned in here yet" [3]
distinguish	Features are highlighted to distinguish the object in focus from others. This is done by means of comparison or contrast.	"all these elements that we previously used expect with a bevel on them" [3]
function assignment	Highlighting or mentioning the function that a feature serves in the whole assembly.	"those two wooden ah those two back elements ('Rückbauelemen-te') need to be put in this groove" [3]
orientation	Object features are used to explain the orientation of an object.	"so that you have two holes on the right hand side in front of you" [2]
placement	Object features are used to describe where and how to position an object in relation to others.	"exactly on top of these three holes" [2]; "and slide in the boards on the left and right hand side of the blue columns" [6]
comment about self	Comment about instructor's own assembly process.	"and I was wrong about the bevels all the times" [5]; "I also paid attention that those holes are down on this board" [5]

strategy	Point out object features and explain their strategic use for the assembly process.	"that we pay attention to those boreholes" [20]

Table 9.8: Classification of functions identified in instructions.

9.5.3 Quantitative analysis

The categories that were defined in the previous sub-section were analyzed with regard to their distribution within each kind of verbal report (think aloud and retrospective) and within instructions. For think aloud protocols and retrospective reports results are reported distinguishing between conditions. Furthermore, for those categories that can be identified in different kinds of verbal reports (see Table 9.9 for an overview) the difference between those reports was investigated. Moreover, in order to study which constraints serve which function, the frequency of categories defining the nature of constraints was investigated within the categories defining their function.

data set in which categories are identified	category (nature of constraints)	category (function of constraints)
think aloud protocols, retrospective reports, & instructions	supporting system, groove, hole, bevel, connection, others	strategy
think aloud & retrospective report		perception, puzzlement
think aloud protocols	label, picture, connector, no screwing	insight, statement
retrospective report	roof, analogy	
instruction		orientation, placement, comment on self
retrospective report & instruction		description, distinguish, function assignment

Table 9.9: Overview on categories for constraints regarding their nature and function as identified in the different verbal reports.

9.5.3.1 Nature and function of constraints

9.5.3.1.1 Think aloud protocols

The first part of this section focuses on the distribution of the categories describing the nature of constraints and the second part presents results on the distribution of categories describing the function of constraints.

Across all conditions *boreholes* (52 times) were the most frequently mentioned category, followed by *bevel* (30 times), and *groove* (21 times) (see Table 9.10).

Differences between conditions showed in these three most frequently mentioned constraints (see Table 9.28 in the appendix). A statistically significant effect of condition on the dependent variables *groove* and *bevel* was observed, $L\chi^2$ (16, N = 149) = 30.9, p = .01[279]. Participants in the *underspecified goal* condition mentioned *grooves* more frequently than all other participants (M = 0.6, SD = 1.2), p < .05[280]. Participants in the *underspecified goal* and *verbal goal* condition referred to *bevels* more often than participants in the *verbal and visual goal* condition (see Figure 9.20), p < .01.

Category	raw frequency	mean	standard deviation
borehole	52	1.04	1.92
bevel	30	0.60	1.50
groove	21	0.42	0.95
supporting system	13	0.26	0.63
connector	12	0.24	0.56
picture	11	0.22	0.58
connection	8	0.16	0.58
label	8	0.16	0.62
no screwing	4	0.08	0.34
others	1	0.02	0.14
all	160	3.20	3.70

Table 9.10: Results frequency of categories describing the nature of constraints in think aloud protocols; sorted from most to least frequent.

[279] The category *picture* was excluded from the statistical test because it can only be mentioned by participants in the *verbal and visual goal* condition hence it was certain to observe a significant main effect there. However, this significant effect would not be representative for considering differences between categories of constraints that can potentially be observed in all conditions.

[280] For standardized residuals see Table 9.29 in the appendix.

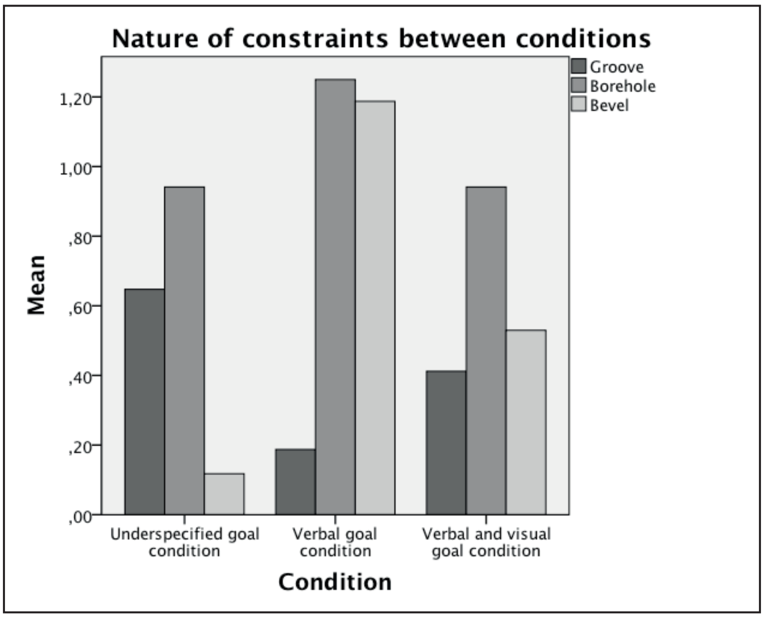


Figure 9.20: Bar plot displaying the distribution of mean frequency of the three most frequent categories describing the nature of constraints.

So far constraints were defined in task specific terms but they can also be defined in more general terms by adapting the classification proposed by Norman (2002). In his classification he distinguished between *logical*, *physical*, *semantic*, and *cultural* constraints. Additionally, the category *task specific* was introduced. The results revealed that participants thinking aloud mentioned *semantic constraints* most frequently (72 cases, M = 1.4, SD = 2.1) (see Figure 9.21). Less frequently they referred to *logical constraints* (34 cases, M = 0.7, SD = 1.2) (see Table 9.30 in the appendix). A similar distribution was observed in all three conditions (for more detail see Table 9.31 in the appendix).

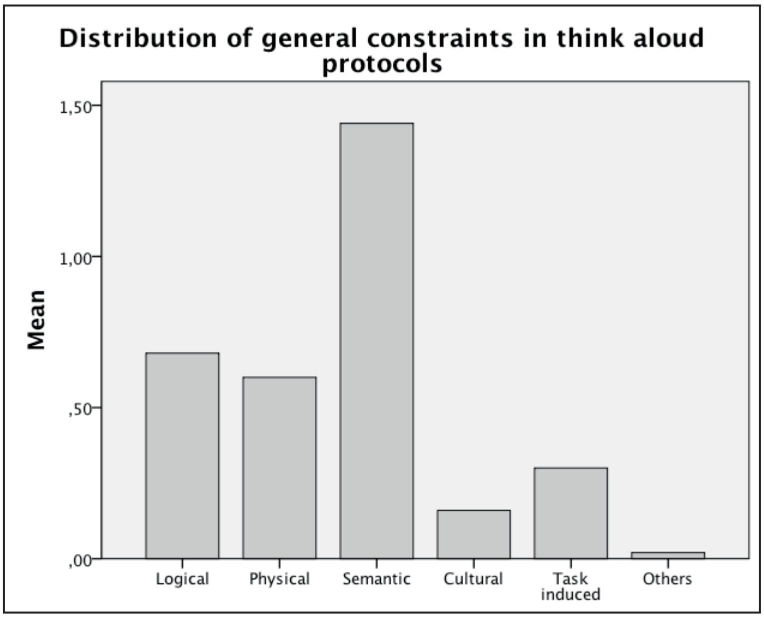


Figure 9.21: Bar plot displaying the distribution of general constraints in think aloud protocols; mean raw frequency.

- Function in think aloud protocols

Five categories were identified that described the function of constraints verbalized in think aloud protocols, namely *insight*, *strategy*, *perception*, *puzzlement,* and *statement*. Across conditions *perception* (50 times, M = 1.0, SD = 1.3) was the most frequently verbalized function. The categories *insight* and *strategy* were frequent as well (42 and 37 times respectively) (see Table 9.11).

category	raw frequency	mean	standard deviation
perception	50	1.00	1.34
insight	42	0.84	1.33
strategy	37	0.74	1.19
statement	16	0.32	0.62
puzzlement	14	0.28	0.57
all	160	3.20	3.70

Table 9.11: Results frequency of categories describing the function of constraints in think aloud protocols; sorted from most to least frequent.

Differences between conditions could be observed regarding the dependent variables *perception* and *statement*. Participants in the *verbal and visual goal* condition mentioned constraints in *statements* (10 times, M = 0.6, SD = 0.7) more frequently than participants in the other conditions (see Figure 9.22 and Table 9.33 in the appendix). Participants in the *verbal goal* condition referred to constraints in *perception* frequently (22 times, M = 1.4, SD = 1.3). *Perception* was less frequently verbalized in the *verbal and visual goal* condition than in the other conditions. However, those differences in distribution did not reach statistical significance, χ^2 (8, N = 160) = 12.3, p = .14.

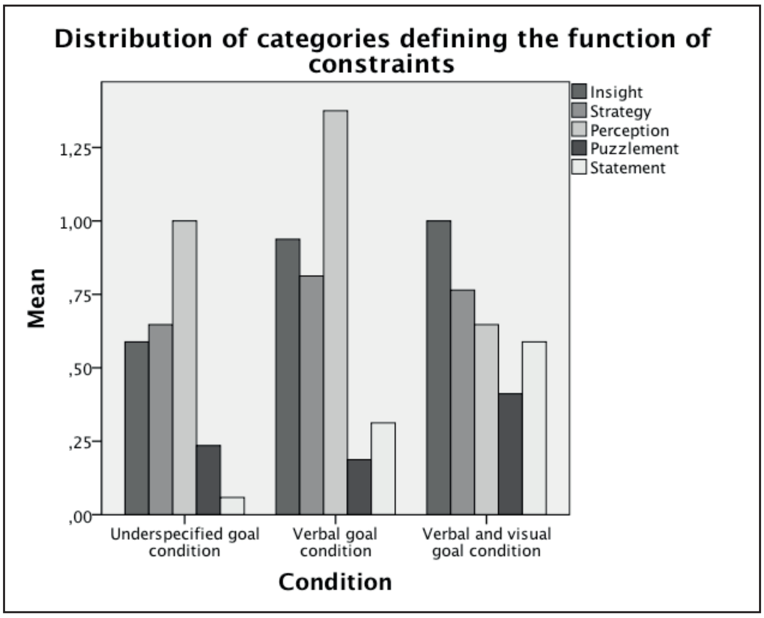


Figure 9.22: Bar plot displaying the mean raw frequency of categories describing the function of constraints across conditions in think aloud protocols.

Concerning the distinction between perception and usage of constraints, no differences were observed between conditions. Participants generally tended to mention constraints at the level of interpretation rather than at the level of sensing (see Table 9.12 for more detail).

condition	class of perception at the level of sensing	class of usage at the level of interpretation	others	sum
underspecified goal	3 (11.11%)	22 (81.48%)	2 (7.41%)	27 (100%)
verbal goal	4 (13.79%)	25 (86.21%)	2 (6.90%)	29 (100%)
verbal and visual goal	6 (10.91%)	44 (80.00%)	5 (9.09%)	55 (100%)

Table 9.12: Raw frequency and percentage of constraints mentioned in the different conditions in retrospective reports distinguishing between *class of perception*, *class of interpretation*, and *others*.

9.5.3.1.2 Retrospective report

As in the previous sub-section, the first part focuses on the distribution of the categories describing the nature of constraints and the second part presents results on the distribution of categories describing the function of constraints.

Across retrospective reports *bevel* (25 times, $M = 0.5$, $SD = 0.7$) was the most frequently reported constraint. *Supporting system* (19 times, $M = 0.4$, $SD = 0.7$) and *borehole* (17 times, $M = 0.3$, $SD = 0.7$) were almost equally frequently reported in retrospective reports (see Table 9.13).

category	raw frequency	mean	standard deviation
bevel	25	0.50	0.74
supporting system	19	0.38	0.67
borehole	17	0.34	0.66
roof	14	0.28	0.46
groove	8	0.16	0.37
picture	8	0.36	0.42
analogy	7	0.14	0.35
others	7	0.14	0.40
no screwing	3	0.06	0.24
connection	3	0.06	0.31
all	111		

Table 9.13: Results frequency of categories describing the nature of constraints in retrospective reports; sorted from most to least frequent.

Although trends marking differences between conditions with regard to the categories were not statistically significant, $L\chi^2$ (16, N = 103) = 19.5, p = .24[281], they are reported as this study wants to highlight differences that might be significant if a larger data set is investigated. Participants in the *verbal and visual goal* condition referred to *bevels* (12 times, M = 0.71, SD = 0.69) and the *supporting system* (10 times, M = 0.6, SD = 0.9) more frequently than participants in the other two conditions (see Figure 9.23). Participants in the *underspecified goal* condition referred to *grooves* least often (1 time, M = 0.1, SD = 0.2) (see Table 9.34 in the appendix).

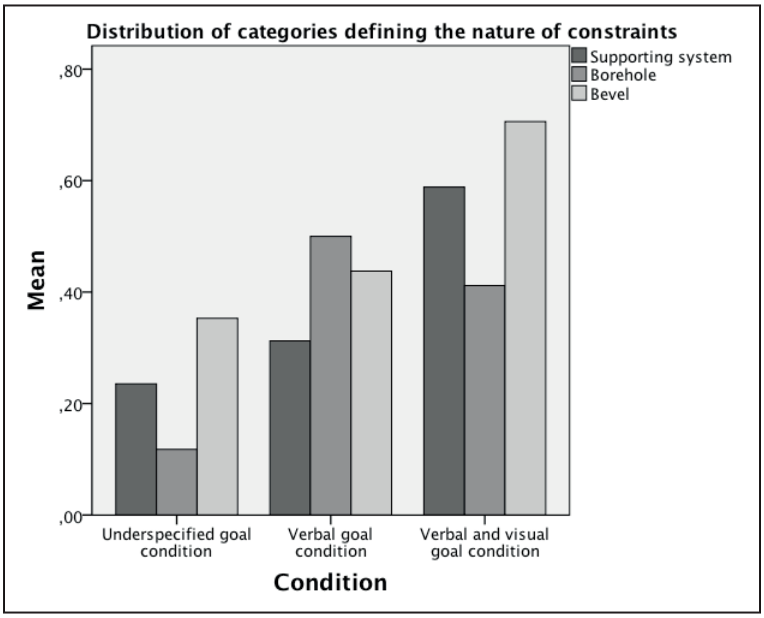


Figure 9.23: Bar plot displaying the distribution of the three most frequent categories defining the nature of constraints between conditions in retrospective reports.

So far the task specific nature of constraints was investigated. Returning to Norman's (2002) more general classification of constraints, the following distribution in retrospective reports could be identified. Participants reporting their activities during assembly did not show a clear preference for one category of general constraints (see Figure 9.24). They mentioned *logical constraints* most frequently (34 cases, M = 0.68, SD = 1.02) but the difference was only marginal as compared to *physical* (25 cases, M = 0.50, SD = 0.74) and *semantic constraints* (20 cases, M = 0.40, SD = 0.70) (see Table 9.35 in the appendix).

[281]Again, the category *picture* is excluded from the statistical test because it can only be mentioned by participants in the *verbal and visual goal* condition hence there needs to be a significant effect there. But this significant effect is not representative for considering differences between categories of constraints that can be (potentially) observed in all conditions.

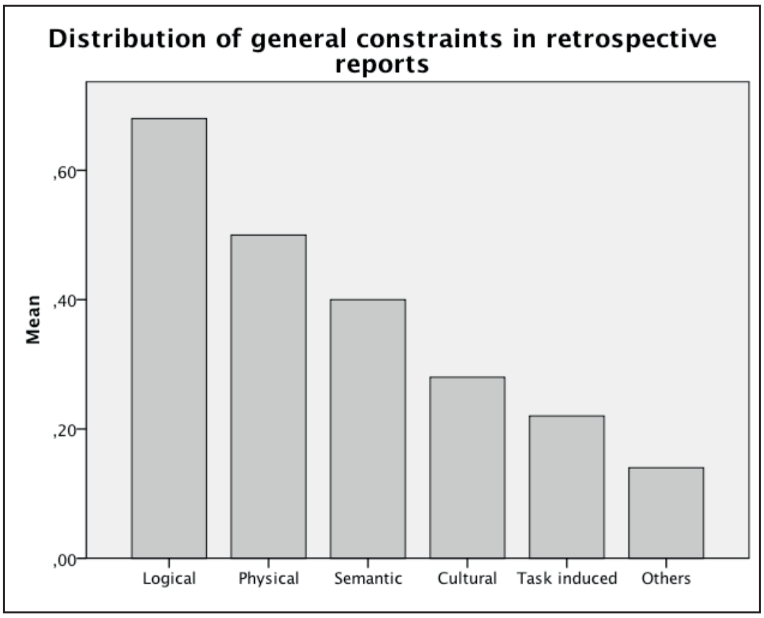


Figure 9.24: Bar plot displaying the distribution of general constraints in retrospective reports; mean raw frequency.

- Function in retrospective reports

The 111 constraints that were identified could be classified into seven categories describing their function, namely *description*, *perception*, *distinguish*, *function assignment*, *insight*, *puzzlement*, and *strategy*. Considering their general distribution across retrospective reports, *function assignment* was most frequent (44 times, M = 0.9, SD = 0.9). The functions *strategy* (23 times, M = 0.5, SD = 0.8) and *insight* (13 times, M = 0.3, SD = 0.5) ranked second and third (see Table 9.14 for an overview).

category	raw frequency	mean	standard deviation
function assignment	44	0.88	0.90
strategy	23	0.46	0.79
insight	13	0.26	0.49
description	10	0.20	0.45
distinguish	9	0.18	0.48
puzzlement	9	0.18	0.44
perception	3	0.06	0.31
all	111	2.20	2.07

Table 9.14: Results frequency of categories describing the function of constraints in retrospective reports; sorted from most to least frequent.

Differences between conditions could be observed although they did not reach statistically significance, $L\chi^2$ (12, N = 111) = 19.8, p = .07. Participants in the *verbal and visual goal* condition mentioned strategic use of constraints more frequently than all other participants (15 times, M = 0.9, SD = 1.0) (see Figure 9.25). Participants in the *underspecified goal* condition reported insights by means of constraints least often (1 time, M = 0.1, SD = 0.2). Function assignment was most frequently reported by participants in the *verbal and visual goal* condition (19 times, M = 1.1, SD = 0.9) and least frequently by participants in the *verbal goal* condition (9 times, M = 0.6, SD = 0.8) (see Table 9.37 in the appendix).

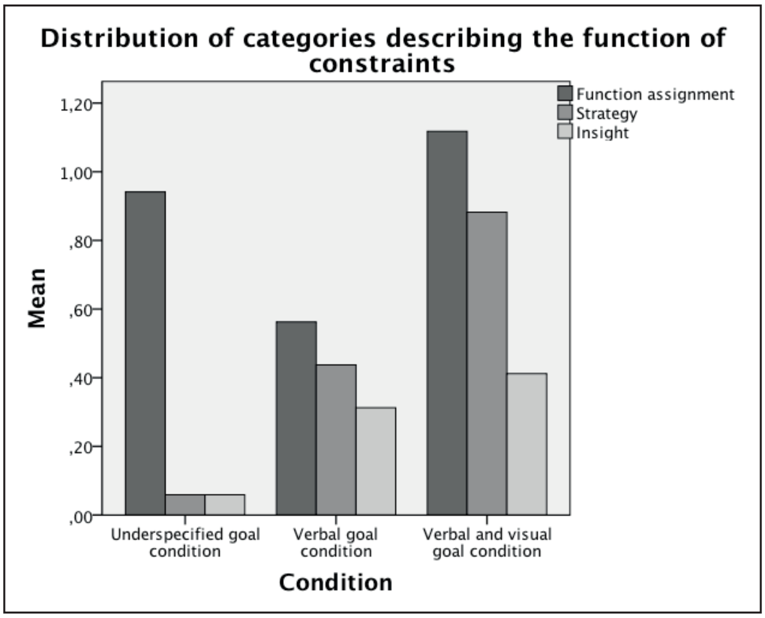


Figure 9.25: Bar plot displaying the distribution of categories describing the nature of constraints identified in retrospective reports, within conditions; mean raw frequency.

- <u>Think aloud and retrospective reports concerning function</u>

The analysis of the distribution of function categories, with regard to the level of sensing and the level of interpretation, revealed that think aloud protocols contained information about constraints at the level of sensing in one third of the cases. Information about constraints at the level of interpretation were verbalized in almost 60% of the cases. This was different in retrospective reports. Participants remembered clearly more information about constraints at the level of interpretation than at the level of sensing (see Table 9.15).

data set	class of perception at the level of sensing	class of usage at the level of interpretation	others	sum
think aloud	50 (31.25%)	95 (59.34%)	3 (1.88%)	160 (100%)
retrospective reports	13 (11.71%)	89 (80.18%)	9 (8.11%)	111 (100%)

Table 9.15: Raw frequency and percentage of constraints mentioned in think aloud protocols and retrospective reports distinguishing between *class of perception*, *class of interpretation*, and *others*.

9.5.3.1.3 Instruction

This sub-section reports findings on the analysis of nature and function of constraints mentioned in instructions. The first part focuses on the distribution of the categories describing the nature of constraints and the second part presents results on the distribution of categories describing the function of constraints.

Out of those ten categories that were identified in think aloud protocols five were adapted for defining constraints identified in instructions, namely *supporting system*, *groove*, *borehole*, *bevel*, and *connection*. The results clearly revealed a preference for referring to the *boreholes* as constraints (73 times, M = 4.6, SD = 4.9). A one-tailed t-test revealed that *boreholes* were mentioned significantly more frequently than expected, t (15) = 3.8, p = .002. *Bevel*, ranking second, was less frequently mentioned (33 times, M = 2.1, SD = 3.0) followed by *groove* which ranked third (23 times, M = 1.4, SD = 1.3) (see Figure 9.26 and Table 9.16). A one-tailed t-test revealed that *grooves* were mentioned significantly less frequently than expected, t (15) = 4.4, p = .001. The same was observed for *bevels*, t (15) = 2.8, p = .01.

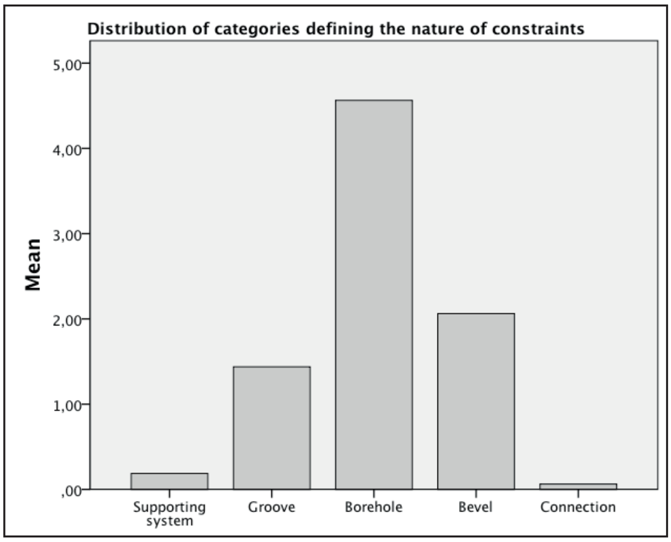


Figure 9.26: Bar plot displaying the distribution of categories describing the nature of constraints identified in instructions; mean raw frequency.

category	raw frequency	mean	standard deviation
borehole	73	4.56	4.86
bevel	33	2.06	2.95
groove	23	1.44	1.31
supporting system	3	0.10	0.40
connection	1	0.06	0.25
all	136	8.50	5.72

Table 9.16: Results frequency of categories describing the nature of constraints in instructions; sorted from most to least frequent.

So far the categories defining constraints in task specific terms have been presented. Taking Norman's (2002) classification into consideration, instructors showed the same tendency as participants thinking aloud during assembly. Instructors mentioned *semantic constraints* most frequently (73 cases, M = 4.6, SD = 4.8) (for more detail see Table 9.38 in the appendix). *Logical constraints* were referred to least often (26 cases, M = 1.6, SD = 1.2). Instructors did not refer to *cultural* and *task induced* constraints (see Figure 9.27).

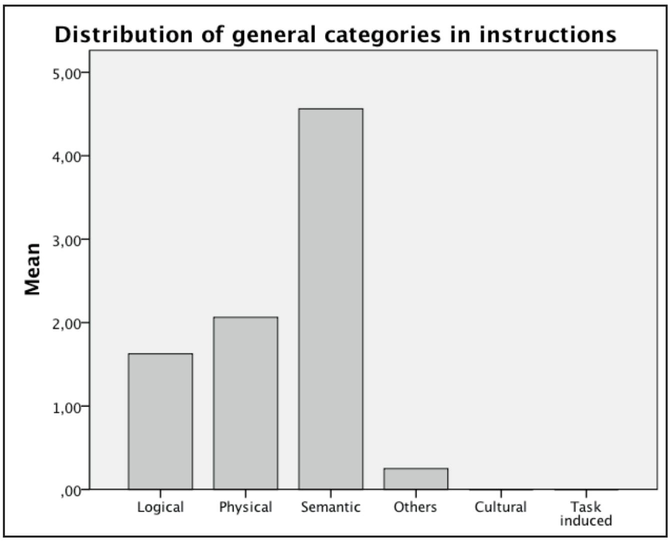


Figure 9.27: Bar plot displaying the distribution of general constraints in instructions; mean raw frequency.

The function of the identified constraints could be described by seven categories, i.e. *description*, *distinguish*, *function assignment*, *orientation*, *placement*, *strategy*, and *comment on self*. The categories *placement* (43 times, M = 2.7, SD = 2.9) and *description* (40 times, M = 2.5, SD = 1.9) were the most frequent ones. *Orientation* ranked third with 30 times (M = 1.9, SD = 2.3) (see Figure 9.28 and Table 9.17).

category	raw frequency	mean	standard deviation
placement	43	2.69	2.89
description	40	2.50	1.86
orientation	30	1.88	2.28
distinguish	9	0.56	1.03
function assignment	6	0.38	0.62
strategy	5	0.31	0.70
comment on self	3	0.19	0.40
all	136	8.50	5.72

Table 9.17: Results frequency of categories describing the function of constraints in instructions; sorted from most to least frequent.

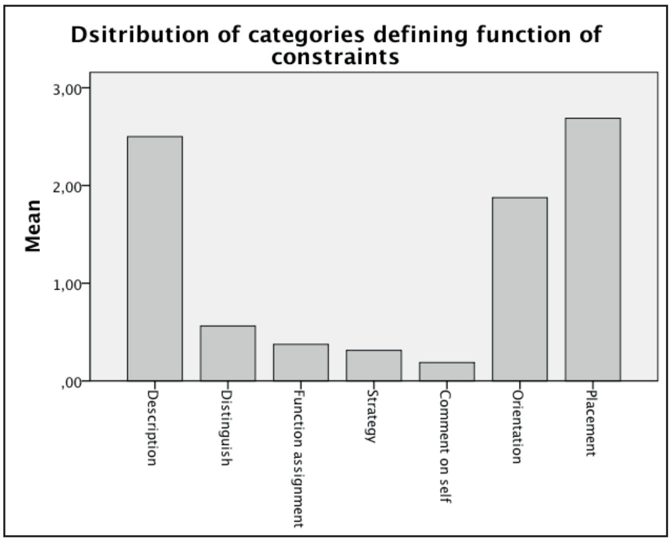


Figure 9.28: Bar plot displaying the distribution of categories describing the function of constraints identified in instructions; mean raw frequency.

9.5.3.1.4 Between verbal reports

The comparison of the three most frequently mentioned categories describing the task specific nature of constraints showed that think aloud protocols and instructions had the same ranking, i.e. *borehole*, *bevel, and groove* although with different frequencies. Retrospective reports, on the other hand, showed a different ranking, namely *bevel*, *supporting system*, and *borehole*. The difference between these rankings was statistically significant, $\chi^2(4, N = 282) = 11.2, p = .03$. The standardized residuals revealed a main effect of kind of verbal report on the dependent variable *bevel*. Instructors referred to *bevels* significantly more frequently (33 times, M = 2.06) (see Table 9.18) than participants thinking aloud during assembly (30 times, M = 0.6) or recalling their activities after assembly (25 times, M = 0.5), $p < .01$ (see Table 9.40 in the appendix for standardized residuals).

kind of verbal report	rank 1	rank 2	rank 3
think aloud	borehole (52 cases, M = 1.04)	bevel (30 cases, M = 0.60)	groove (21 cases, M = 0.42)
retrospective report	bevel (25 cases, M = 0.50)	supporting system (19 cases, M = 0.38)	borehole (17 cases, M = 0.34)
instruction	borehole (73 cases, M = 4.56)	bevel (33 cases, M = 2.06)	groove (23 cases, M = 1.44)

Table 9.18: Summary of the three most frequent categories describing the nature of constraints in the investigated kinds of verbal reports.

The distribution of general constraints within verbal reports showed that think aloud protocols were similar to instructions in first rank whereas retrospective reports show a different trend (see Table 9.19). Second and third rank were quite close in think aloud data and retrospective reports thus the interpretation was not as strong as the difference between second and third rank in instructions. Taking this observation into account, the distribution of general constraints was similar between think aloud protocols and instructions but highly significantly different to retrospective reports, $\chi^2(4, N = 347) = 22.8, p = .00$. Participants reporting their activities used *semantic constraints* significantly less frequently than the other participants, $p < .01$. Furthermore, instructors mentioned *logical constraint* significantly more often than the other participants, $p < .01$ (see Table 9.41 in the appendix for standardized residuals).

kind of verbal report	rank 1	rank 2	rank 3
think aloud	semantic (72 cases, M = 1.44)	logical (34 cases, M = 0.68)	physical (30 cases, M = 0.60)
retrospective report	logical (34 cases, M = 0.68)	physical (25 cases, M = 0.50)	semantic (20 cases, M = 0.40)
instruction	semantic (73 cases, M = 4.56)	physical (33 cases, M = 2.06)	logical (26 cases, M = 1.63)

Table 9.19: Summary of the three most frequent categories describing the general nature of constraints in the investigated kinds of verbal reports.

The comparison of the three most frequently identified functions of constraints showed that *insight* and *strategy* were among those three, both in think aloud protocols and retrospective reports but at different places within the ranking (see Table 9.20). The categories identified in instructions were different. However, this impression holds only at first glance because in instructions the two categories *placement* and *orientation* were individual categories whereas those were part of the category *strategy* in the other two kinds of verbal reports. This categorization was based on the observation that the number of *placements* of objects and references to constraints for means of *orientation* were frequent in instructions and needed to be analyzed separately. However, in order to compare the different kinds of investigated verbal reports, it was necessary to define one consistent category, thus *placement* and *orientation* were subsumed in the category *strategy* (column *instruction combined* in Table 9.20).

The analysis of the resulting raw frequencies (see Table 9.21) revealed a highly significant effect of kind of verbal report on the dependent variables *insight* and *strategy*, $\chi^2(2, N = 193) = 55.7, p = .00$. The analysis of the standardized residuals highlighted that the high frequency of *insights* in think aloud protocols contributed to this main effect, $p < .001$ (see Table 9.42 in the appendix for standardized residuals). Additionally, instructors and assemblers thinking aloud referred to using constraints strategically significantly more often ($p < .01$ in both cases) than participants reporting their activities.

kind of verbal report	rank 1	rank 2	rank 3
think aloud	perception (50 cases, M = 1.00)	insight (42 cases, M = 0.84)	strategy (37 cases, M = 0.74)
retrospective report	function assignment (44 cases, M = 0.88)	strategy (23 cases, M = 0.46)	insight (13 cases, M = 0.26)
instruction	placement (43 cases, M = 2.69)	description (40 cases, M = 2.50)	orientation (30 cases, M = 1.88)
instruction combined	strategy (78 cases, M = 4.88)	description (40 cases, M = 2.50)	distinguish (9 cases, M = 0.56)

Table 9.20: Summary of the three most frequent categories describing the function of constraints in the investigated kinds of verbal reports and additionally the distribution of categories if the category constraint is defined exactly the same in instructions as in the other two reports.

kind of verbal report	insight	strategy
think aloud	42	37
instruction	0	78
retrospective report	13	23

Table 9.21: Raw frequency of categories describing functions of constraints between investigated kinds of verbal reports.

9.5.3.1.5 Investigating the nature of constraints within functions

After identifying the different categories describing the nature of constraints it was possible to investigate which constraints frequently served which function. This question was investigated by eliciting the distribution of categories on the nature of constraints within each functional category. Only those categories that were identified in at least two of the verbal reports can be statistically analyzed (see Table 9.43 in the appendix) therefore a table with raw frequencies is provided in the appendix (see Table 9.44). For a better overview Table 9.22 presents only a selection of those constraints that were ranked first to third most frequent in one of the categories.

Boreholes were most frequently used strategically (M = 12.0). This trend got more marked when observing the results of the category *strategy combined*; here *boreholes* were used strategically most often (M = 28.7). Although *bevels* were also mentioned to be used strategically, this was done less frequently (M = 6.3). The analysis revealed that *bevels* fostered insights (M = 9.0) most frequently. *Grooves* and *boreholes* supported insights (M = 5.5 and M = 5.0 respectively). *Boreholes* (M = 9.5) and *grooves* (M = 6.0) were the constraints that were perceived most frequently (see Table 9.22).

Participants mentioned to be puzzled about the *supporting system* in the majority of cases (M = 4.0) but less often about *boreholes* (M = 2.5). Interestingly, *boreholes*, *bevels*, and *grooves* were described to a comparable amount. If participants made use of constraints to distinguish between objects, they most frequently referred to the *bevel* (M = 5.0) and less frequently to the *supporting system* (M = 2.5). If constraints were used in assigning function to objects or when function was assigned to constraints, this related to *bevels* (M = 6.0) as well as to *grooves* (M = 5.5) in most cases (see Table 9.22).

	supporting system	groove	borehole	bevel	analogy	roof
insight	0	5.5 (3.54)	5.0 (5.66)	9.0 (0.66)	0.5 (0.71)	0
strategy	0.67 (0.58)	0.33 (0.58)	12.0 (8.55)	3.0 (2.0)	1.33 (2.31)	0
percep-tion	4.0 (4.24)	6.0 (8.49)	9.5 (12.02)	4.0 (5.66)	0	0
puzzle-ment	4.0 (1.41)	0	2.5 (2.12)	0	0	0
descrip-tion	0	6.0 (8.49)	7.0 (8.49)	6.5 (7.78)	0.5 (0.71)	0
distingu-ish	2.50 (3.54)	0.5 (0.71)	0.5 (0.71)	5.0 (2.83)	0.5 (0.71)	0
function assign-ment	2.50 (3.54)	5.5 (0.71)	1.0 (1.41)	6.0 (8.49)	0	7.0 (9.90)
strategy combined	1.33 (0.58)	1.67 (2.08)	28.67 (22.50)	6.33 (4.16)	1.33 (2.31)	0

Table 9.22: Overview frequency of selected constraints within functions; mean and standard deviation.

A chi-square test was run to investigate the frequency of references to *boreholes*, *grooves*, and *bevels* within the different categories defining the function of constraints. The test revealed that the nature of constraints had a highly significant effect on the dependent variable function of constraints, $L\chi^2$ (12, N = 226) = 82.0, p = .00. *Bevels* were significantly more frequent in the category *distinguish* than *boreholes* and *grooves*, p < .01 (see Table 9.45 in the appendix for standardized residuals). *Boreholes* were significantly less frequent in *function assignment* than the other two types of constraints, p < .01. However, *boreholes* were significantly more often referred to in the category *strategy* than the other two kinds of constraints, $p < .001$. Additionally, *grooves* were referred to significantly less frequently, $p < .01$.

9.5.3.2 Occurrence of constraints

9.5.3.2.1 Within the problem solving process

The distribution of references to constraints within the different problem solving processes revealed that constraints were most frequently mentioned in *hypotheses* (41.5%) and *descriptions of object features* (36.5%) (see Table 9.23).

problem solving process	raw frequency constraints mentioned	percent constraints mentioned
hypothesis	66	41.51
comment on object features	58	36.48
action	16	10.06
description of mental state	12	7.55
negative evaluation	2	1.26
positive evaluation	2	1.26
comment on self	1	< 1.0
metalevel comment	2	1.26
sum	159[282]	100.00

Table 9.23: Frequency of constraints within the different problem solving processes; raw frequency and percentage.

9.5.3.2.2 Within the overall assembly process

Dividing the assembly process into quarters, the distribution of references to constraints between the four intervals revealed that constraints were most frequently mentioned in the third quarter of the assembly (58 times, $M = 1.2$, $SD = 2.1$). The frequency of constraints was comparable in the other three quarters (see Table 9.24).

quarter	raw frequency	mean	standard deviation
first	33	0.66	1.22
second	37	0.74	0.88
third	58	1.16	2.10
fourth	32	0.64	1.27

Table 9.24: Frequency of constraints in each of the quarters of the assembly process, raw frequency, mean, and standard deviation.

[282] In one case a constraint is mentioned at the end of the protocol thus it is excluded from the calculation here.

A chi-square test revealed a significant effect of condition on the frequency of references to constraints in the third quarter, $\chi^2(6, N = 160) = 17.2$, $p = .01$. Participants in the *verbal goal* condition mentioned constraints in the third quarter of the assembly significantly more often (M = 1.9, SD = 3.2) than participants in the other conditions, $p < .05$ (see Table 9.46 in the appendix for standardized residuals). Participants in the *verbal and visual goal* condition referred to constraints significantly less often (M = 0.7, SD = 0.9) than expected in the third quarter, $p < .05$.

As a trend it could be observed that participants in the *verbal and visual goal* condition mentioned constraints in the first quarter more often than participants in the other conditions. Furthermore, participants in the *underspecified goal* condition referred to constraints in the fourth quarter least frequently (see Table 9.25).

condition	1st quarter	2nd quarter	3rd quarter	4th quarter
underspecified goal	10 (M = 0.59, SD = 1.18)	13 (M = 0.76, SD = 1.62)	16 (M = 0.94, SD = 1.64)	5 (M = 0.29, SD = 0.85)
verbal goal	6 (M = 0.38, SD = 0.62)	10 (M = 0.63, SD = 0.89)	30 (M = 1.88, SD = 3.16)	12 (M = 0.75, SD = 1.53)
verbal and visual goal	17 (M = 1.0, SD = 1.62)	14 (M = 0.82, SD = 1.01)	12 (M = 0.71, SD = 0.85)	15 (M = 0.88, SD = 1.36)
all	33 (M = 0.66, SD = 1.22)	37 (M = 0.74, SD = 0.88)	58 (M = 1.16, SD = 2.10)	32 (M = 0.64, SD = 1.27)

Table 9.25: Frequency of constraints within the four quarters of the assembly process, raw frequency, mean, and standard deviation.

Besides the general frequency of constraints within quarters of the assembly process, the distribution of the identified functions within quarters was also interesting especially if differences between conditions could be observed. The reports of noticing constraints (*perception*) decreased from first to fourth quarter in the *verbal goal* condition (see Table 9.26).

Participants in the *verbal and visual goal* condition used constraints in their function as fostering *insights* (41.7%) most frequently in the third quarter. They did so more often than participants in the *underspecified goal* condition (12.5%) and the *verbal goal* condition (14.3%). However, these result did not reach statistical significance, $L\chi^2(6, N = 42) = 11.3$, $p = .08$.

Participants in the *verbal and visual goal* condition used constraints strategically less frequent in the third quarter (16.7%) than participants in the *underspecified goal* condition (43.8%) and participants in the *verbal goal* condition (25.7%).

Although the observed difference did not reach statistical significance the trend was marked, Lχ^2 (6, N = 38) = 12.0, p = .06.

condition [283]	quarter	insight	strategy	perce-ption	puzzle-ment	statement
A	1	4 (40%)	0	5 (50%)	1 (10%)	0
A	2	4 (26.7%)	3 (20%)	5 (33.3%)	2 (13.3%)	1 (6.7%)
A	3	2 (12.5%)	7 (43.8%)	6 (37.5%)	1 (6.3%)	0
A	4	0	2 (28.6%)	6 (71.4%)	0	0
B	1	2 (33.3%)	0	3 (50%)	0	1 (16.7%)
B	2	2 (20%)	2 (20%)	4 (40%)	0	2 (20%)
B	3	5 (14.3%)	9 (25.7%)	12 (34.3%)	3 (8.6%)	1 (2.9%)
B	4	6 (50%)	2 (16.7%)	3 (50%)	0	1 (8.3%)
C	1	6 (35.3%)	2 (11.8%)	2 (11.8%)	2 (11.8%)	5 (29.4%)
C	2	2 (13.3%)	4 (26.7%)	5 (33.3%)	1 (6.7%)	3 (20.0%)
C	3	5 (41.7%)	2 (16.7%)	2 (16.7%)	3 (25.0%)	0
C	4	4 (26.7%)	5 (33.3%)	2 (13.3%)	1 (6.7%)	3 (20.0%)

Table 9.26: Distribution of categories defining the function of constraints within the four quarters of the assembly process, raw frequency and percentage.

9.5.3.3 Answers to interview questions

This section focuses on the answers to the three guided questions that were posed at the end of the experimental session. The results for each of the guided questions is presented across conditions because no statistical effect of condition on answers was observed, Lχ^2 (10, N = 99) = 6.1, p = .80.

The majority of participants stated to have noticed the groove (43 cases, M = 0.9, SD = 0.4) (see Figure 9.29). Only few participants did not notice it (3 cases, M = 0.1, SD = 0.2). Out of those who noticed the groove almost equally many reported to have used it as a guide (18 cases, M = 0.4, SD = 0.5) and not to have used it (14 cases, M = 0.3, SD = 0.5). Some participants stated that they were irritated by the grooves (7 cases, M = 0.1, SD = 0.4) and few made partial use of them (3 cases, M = 0.1, SD = 0.2) (see Table 9.48 in the appendix).

[283] The conditions are abbreviated for a better overview: *A* stands for *underspecified goal* condition, *B* represents *verbal goal* condition, and *C* represents *verbal and visual goal* condition.

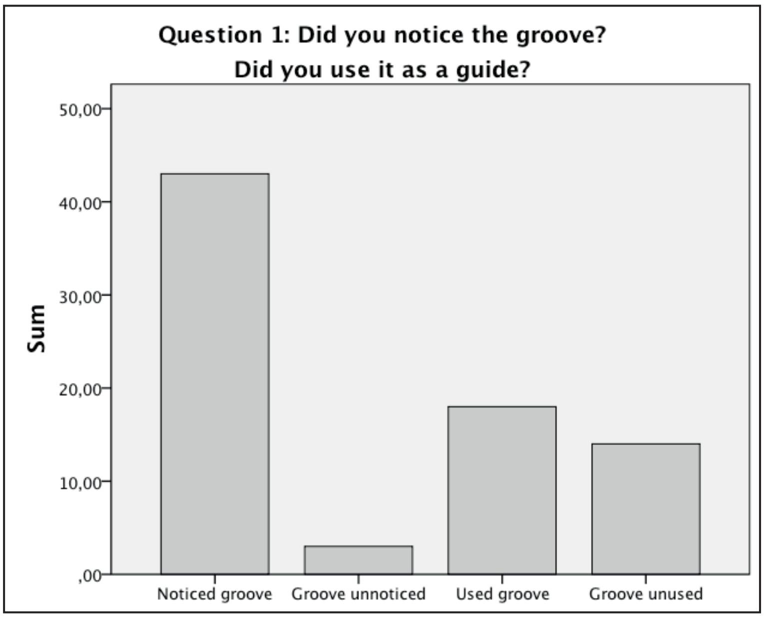


Figure 9.29: Bar plot displaying the answers to the first interview question.

All participants stated that they noticed the *boreholes* (46 cases, M = 0.9, SD = 0.3) (see Figure 9.30). Interestingly, almost half of the interviewees stated that they used it (21 cases, M = 0.4, SD = 0.5) whereas equally many stated that they disregarded them (20 cases, M = 0.4, SD = 0.5). Five participants stated that they used the boreholes partially (M = 0.1, SD = 0.3) (see Table 9.49 in the appendix).

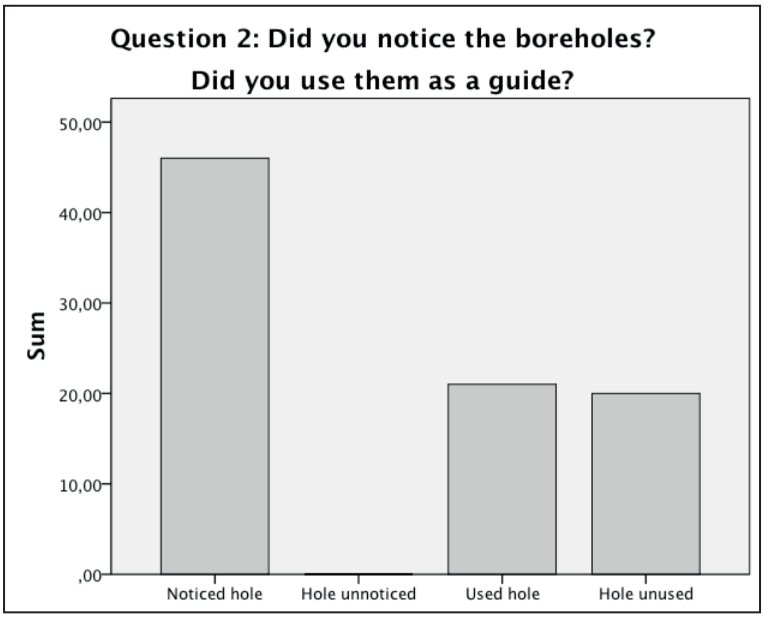


Figure 9.30: Bar plot displaying the answers to the second interview question.

Half of the interviewees stated that they noticed the label 'Selecta Spielzeug' during the assembly (29 cases, M = 0.6, SD = 0.5) (see Figure 9.31). From those participants equally many reported to have used the label as a guide (13 cases, M = 0.3, SD = 0.4) and to have ignored it. The remaining three participants stated that they did not find it helpful to have noticed the label (M = 0.1, SD = 0.2) (see Table 9.50 in the appendix).

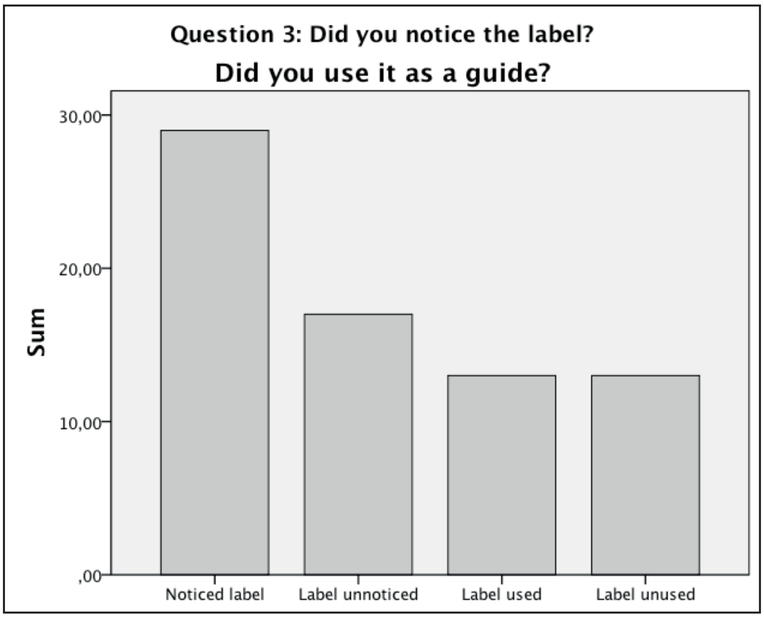


Figure 9.31: Bar plot displaying the answers to the third interview question.

9.6 Discussion

The results that were presented in the previous sections will be discussed in the light of the research questions that were posed concerning the nature and function of constraints in unaided object assembly.

The first research question aimed at identifying the nature of the identified constraints. In the analysis two kinds were distinguished, namely the task specific nature of constraints, i.e. the way in which they were expressed in the specific assembly context, and the general nature of constraints. For the description of the general nature, Norman's (2002) classification of constraints was adopted and extended if necessary. Eight constraints were identified prior to the analysis based on general assumptions about constraints as devices that help to reduce the number of possibilities regarding the actions that can be performed in a specific context (Norman 2002; Kirsh 2009). Those constraints were either physical properties of the provided wooden objects (i.e. *bevels*, *boreholes*, little half-ball like wooden pieces (*supporting system*), and *grooves*), the object itself (i.e. the *roof* and the *picture*), or object features (i.e. the label 'Selecta Spielzeug'). Furthermore, one constraint was stated explicitly in the instructions, namely that *no screws* can be used. The assignment of these expected task specific constraints to Norman's (2002) proposed categories of *physical* (bevel), *semantic* (boreholes), *logical* (groove, supporting system), and *cultural* (label, roof) constraints suggested that this classification could be adopted if one further category was introduced, namely *task induced* constraints (no screws, picture)[284].

The analysis of think aloud protocols revealed that *boreholes* were the most frequently mentioned constraint followed by *bevel* and *groove* in second and third position. This predominance of references to boreholes was unexpected, whereas the second and third position confirmed the expectations. Interestingly, participants in the *underspecified goal* condition referred to *grooves* significantly more often than participants in the other conditions. The grooves are not very salient but they can be noticed by close inspection of the objects. Those findings suggest that participants who needed to create a sensible structure from scratch, i.e. without any prior information, were very attentive to object features. This attentiveness can be interpreted as an expression of their effort to find hints to guide their assembly.

Those participants who were provided with no or few prior information (*verbal goal* condition) mentioned *bevels* significantly more frequently than participants who were provided with an external model (*verbal and visual goal* condition). Since *bevels* are very salient object features, it is very unlikely that participants in the *verbal and visual goal* condition simply did not notice the bevels. It is more likely that prior information has an influence on the verbalization of perception. Participants who saw the picture may know the bevels' function at the moment of perception, i.e. those objects need to be placed on the second story. This recognition happened so fast and unconsciously that it was not verbalized. However, participants who were looking for hints concerning the nature and function of objects noticed this constraint and tried to extract relevant information from this observation. This would be a conscious mental process and thus they verbalized noticing the bevel at the moment of perception.

The distribution of constraints in retrospective reports revealed the expected ranking, i.e. *bevels* were most frequent, the *supporting system* ranked second, and the *boreholes* were mentioned in third position. Interestingly, participants in the *underspecified goal* condition reported using *grooves* least frequently of all participants. Although this difference did not reach statistical significance, it is the reverse trend to that observed in think aloud protocols. This discrepancy might be the result of a memory effect. If the function of a *groove* is culturally well known, then noticing the *groove* and its usage is forgotten after the assembly because it is

[284] For the reasons for grouping the task specific constraints into these general categories see section 9.3.

not a prominent thought during the assembly. The analysis of the answers to question one, which asked participants if they noticed the *grooves* and if they used them in their assembly, supports this assumption. The answers revealed that the majority of participants noticed the groove (93,5%) and half of them reported to have used them as a guide at least partially (45,7%). These figures are in stark contrast to the observation that grooves accounted for only 13.12% of constraints mentioned in think aloud protocols and for 7.20% in retrospective reports. This comparison highlights that noticing and even using a constraint did not guarantee that it was verbalized. The observation that participants could recall noticing and using a specific constraint if it was pointed out to them, supports the assumption that the mental processes associated with this constraint were unconscious but that they were stored in memory.

The distribution of general constraints across think aloud protocols revealed that semantic constraints were much more frequent than logical and physical constraints. This effect can be attributed to the high frequency of references to *boreholes*. Those account for 32.5% of all mentioned constraints in think aloud protocols and 22.5% in retrospective reports. This unexpected finding was supported by the answers to question two in which all participants stated that they noticed the *boreholes*. More than half of all interviewees (56.5%) reported that they used the *boreholes* at least partially. These findings highlight that although *boreholes* were not important for succeeding in the dollhouse assembly, they were frequently used even more often than their mentioning in verbal protocols suggests. Comparing these findings to the ones on mention of *grooves* in verbal reports, it becomes evident that *boreholes* were mentioned and recalled more frequently in verbal reports. The frequent reference to *boreholes* suggests two readings, either *boreholes* were unexpected for participants or that they were assigned a specific meaningful function. The second explanation will be reconsidered and discussed in reference to research question four. The results on the distribution of categories describing the general nature of constraints in retrospective reports revealed that logical constraints were most frequent. This observation confirmed prior expectations because the *supporting system*, which contributed most to the result, was expected to be a salient feature that was highly likely to be mentioned.

In sum, the answers to the guided questions revealed that both *boreholes* and the *grooves* were more frequently noticed than verbalized. However, they also revealed that those constraints were more frequently noticed than used. This finding stresses that noticing a constraint does not imply using it. It suggests that after noticing a constraint its meaning needs to be apparent to the problem solver because otherwise it will be disregarded. This explanation is in line with previous findings that highlighted that recognition and meaningfulness of constraints depends on the individual, his experiences, and his knowledge (Kirsh 1995). Furthermore, the observation that participants noticed and used constraints more frequently than they reported them highlights the importance to analyze different data sets for drawing valid conclusions.

As a further step in the analysis, the adequacy of Norman's (2002) classification needs to be evaluated. This can be achieved by comparing the percentage of constraints that can be classified by Norman's (2002) classification to the percentage of constraints that cannot be classified by his categories. The additional category of *task induced* constraints is excluded from the analysis because it is a category of constraints that is specific to the context of an experimental setting. The results (see Table 9.27) clearly indicate that the great majority (96.9%) of task specific constraints can be classified in general terms using Norman's (2002) categories. Thus it can be concluded that Norman's (2002) classification can be adopted for describing the general nature of constraints used in an assembly scenario.

data set	constraints classified by Norman's (2002) classification	constraints not classified by Norman's (2002) classification	task induced constraints
think aloud	144	1	15
retrospective report	93	7	11
instruction	132	4	0
sum	369 (96.85%)	12 (3.15%)	26

Table 9.27: Number of constraints; raw frequency.

Research question 2 aims at drawing conclusions about the use of constraints with respect to the temporal dimension of the assembly process. Two temporal dimensions were distinguished in the analysis. First, the overall assembly process was considered, i.e. from starting the assembly to the finishing statement. Second, the problem solving process was analyzed in terms of problem solving processes (for more detail see chapter 6). In terms of problem solving processes, constraints were expected to be frequently mentioned in the process introduced as *description of object features* because this category contains all references to object features such as color or shape on a descriptive level. Additionally, constraints were expected to be mentioned in *hypotheses* because this process category contains ideas about functions of objects or object parts within the assembly. As constraints were expected to provide hints regarding the function of objects and limit their possible use, they were very likely to be verbalized in *hypotheses*. The analysis revealed that the distribution of constraints within problem solving episodes was as expected; i.e. constraints were most frequently verbalized in *hypotheses* (41.5%) and *description of object features* (36.5%).

Based on Steffensen's (forthcoming) postulatation that perception happens during action and action happens while the environment is perceived, constraints were expected to be equally distributed across the entire assembly process. It was further assumed that this trend could be affected by the amount of prior information. The results were unexpected as they revealed an even distribution from first to second quarter but a peak in third quarter. The data for participants in the *verbal goal* condition contributed most to the observed effect. Those participants mentioned significantly more constraints in third quarter than all other participants. Additionally, participants in the *verbal and visual goal* condition referred to significantly less constraints in the third quarter. As pointed out in the section on analysis, references to constraints may signal two mental states of the speaker, i.e. either he/she notices the constraint at the time of speaking or he/she uses it again after previously noticing it. The analysis of the distribution of the categories defining the function of constraints within each of the four quarters allows to distinguish between these two options. If participants referred to a constraint in the function of *perceiving*, this constraint can be assumed to encode the status 'not noticed and mentioned before'. If participants referred to the constraint in any other function, it is more likely that the constraint was previously noticed and mentioned already. An overview on the identified functions of constraints is presented later in this discussion because for the current purpose of interpreting the findings on the overall distribution of constraints, it is sufficient to distinguish between noticing and using constraints.

Considering the frequency of the category *perception* across the assembly process verbalized by participants in the *underspecified goal* condition, the results suggest that those participants noticed constraints to a comparable degree throughout the assembly process with a peak in the fourth quarter. Different explanations seem plausible regarding this observation. Either participants did not succeed in conceptualizing and arranging the objects in the first half of the assembly time, thus they started observing the given objects at a more detailed level in the second half of the assembly. Or following Klix's (1976) argument, participants noticed more constraints because they interacted with the objects for some time already.

This interaction can be assumed to make constraints more salient. Additionally, constraints can be expected to gain meaning for the assembler as a mental representation of the goal structure evolves.

If Klix's (1976) argument was the sole explanation, the same trend should be observed in the other two conditions as well. However, participants who were provided with specific prior information on the goal object (*verbal goal* and *verbal and visual goal* condition) noticed less constraints towards the end. First, this finding highlights that the results cannot be explained by Klix's (1976) approach entirely. Second, the results observed in the *verbal and visual goal condition* confirm the assumption that participants who were provided with much prior information noticed constraints early within the assembly. Participants who were told about the general nature of the goal structure (*verbal goal* condition) started with a generic mental representation that can be assumed to get more specific throughout the assembly. Assuming that this representation evolves in the first half of the assembly process, the decreasing trend of noticing constraints along with the high frequency of references to constraints in the third quarter suggests that previously perceived constraints were integrated into the new representation. Additionally, newly encountered constraints may be conceptualized more easily and are thus more frequently verbalized. The comparison of the frequency of the category *perception* (34.3%) to that of categories encoding the use of constraints (42.9% of the times[285]) in the third quarter supports this interpretation.

In this line of thought, the results obtained in the *verbal and visual goal* condition suggest that participants who were provided with a finished mental representation of the goal structure, integrated and interpreted constraints early in the assembly process which left fewer constraints to be discovered in the second half. Based on this evidence, the results highlight an influence of prior information on the time at which constraints were noticed and evaluated as meaningful. This interpretation suggests that Steffensen's account (2013:199) could be extended by including the

[285] This percentage has been calculated by adding the percentage of the categories strategy, insight, and statement as reported in Table 9.26. The category statement is considered although it includes general statements which do not encode active use of constraints because constraints are used to draw conclusions that may influence further thoughts and actions. The category puzzlement is not considered because objects may leave the participant puzzled even if he notices it a second time. This question can be tackled by annotating references to constraints as either *new*, i.e. not mentioned before, or *repetition*, i.e. it has been referred to before. This even more fine-grained annotation was not done so far because the approach adapted here provides enough insights regarding the research questions that have been posed. However, further analyses with even more specific research questions are possible.

problem solver's knowledge state as one factor into his notion of the "embodied and emotional action-perception cycles".

Research question 3 focused on the influence of the discovery of constraints on the problem solving process. Prior to the analysis, it was expected that *boreholes* could be used as hints as to where objects should be placed on the boards. Additionally, participants were assumed to use salient physical properties such as the *bevel* and the *supporting system* to distinguish between objects. The analysis revealed numerous functions in think aloud protocols and retrospective reports. Some functions were identified across verbal reports whereas others were unique to one kind. In addressing the function of constraints in the previous part of the discussion, two general classes were distinguished, i.e. 'class of perception' and 'class of usage'. The class of perception includes the category *perception* in which participants simply report noticing a constraint. This category could be identified in both data sets. The class of usage subsumes three categories of function in think aloud protocols, i.e. *insights*, *strategy*, and *statements*. The last category implies usage of constraints implicitly because constraints were used to draw general conclusions about objects or the assembly process. These conclusions were likely to influence further thoughts and actions. Sorting objects was among the strategic functions of constraints. This kind of strategic use can be interpreted as one form of active rearrangement of the environment to facilitate understanding and remembrance (Kirsh 1995; Kirsh 1996). Although it can be assumed that the focus was probably more on understanding than on remembrance in the analyzed scenario.

The analysis revealed that the majority of functions (i.e. four out of five) that were described for think aloud protocols were also identified in retrospective reports, namely *insight*, *puzzlement*, *strategy*, and *perception*. However, statements about constraints were unique to think aloud protocols. Additionally, three new categories were identified in retrospective reports, i.e. *description*, *distinguish*, and *function assignment*.

Comparing the distribution of categories defining the function of constraints between conditions, an influence of amount of prior information could be observed. Participants who were provided with few prior information, mentioned perceiving constraints more often than participants who were provided with much prior information (*verbal and visual goal* condition). Those participants tended to state facts about constraints. This finding suggests that constraints are integrated into an existing mental representation. This integration allows for conclusions about them, stated as a fact. Additionally, the findings highlighted that participants who were provided with much prior information, tended to use constraints at the level of interpretation, whereas participants who were provided with no specific information reported constraints at the level of sensing.

Furthermore, the frequency analysis of those two categories within each kind of verbal report highlighted differences between verbal reports. More specifically, participants thinking aloud mentioned perceiving constraints most frequently, whereas *function assignment* was the most frequently used category in retrospective reports. The categories *insight* and *strategy* ranked second and third in both verbal reports but in reverse order. This finding suggests that perception was most important when the assembler was engaged in the task and noticing constraints fostered understanding. In retrospective reports the assigned function to an object was prominent. This impression is strengthened by comparing references to constraints within the *class of perception* to that within the *class of usage*. The comparison highlighted that participants thinking aloud referred to constraints within the *class of usage* more than half of all times but they also referred to the *class of perception* in one third of all cases. Participants who reported their actions and thoughts during the assembly, in contrast, mentioned constraints within the *class of usage* in 80.3% of the cases. Constraints within the *class of perception* were only mentioned in 11.7% of the cases. These results highlight the different character of verbal reports and their complementary nature.

Interestingly, participants who were provided with no specific goal information (*underspecified goal* condition) reported using constraints at the level of interpretation to a comparable amount as participants who started their experiment with much prior information (*verbal and visual goal* condition). This finding suggests that reporting function assignment in retrospective reports was based on information stored in long-term memory. In order to use constraints strategically, the assembler needs to know about the actual or potential function of that constraint. The results in think aloud protocols suggest that participants who were provided with an external representation of the goal structure, were more likely to understand the function of a constraint and could thus use them strategically to facilitate the assembly process. It can be argued that the constraints and their function were well remembered in retrospective reports due to the facilitation affect. Participants who needed to create the goal structure without any prior information are likely to remember the function for the same reason, i.e. because knowing the function helped them to proceed. This makes it an important detail to store in long term-memory.

By combining the findings on the nature with those on the function of constraints, it was highlighted that some constraints were perceived as specifically well suited to serve a specific function. In order to draw general conclusions, the largest possible data set needs to be analyzed. Therefore, the analysis was based on the findings in think aloud protocols, retrospective reports, and instructions. The results revealed that each constraint served different functions, e.g. *bevels* were frequently used in *function assignment* but they were also referred to when participants reported distinguishing between objects. However, some constraints were specifically prominent in one functional category, such as the cultural constraint (*roof*) that was specifically helpful in assigning function to the goal object by raising expectations concerning its general nature. The finding that *grooves* frequently served *function assignment* confirmed the expectation that cultural knowledge about the use of constraints influenced the assembly process. More surprising and unexpected were the findings on the frequent strategic use of *boreholes*. Prior to the analysis, it was expected that participants would not pay attention to the *boreholes* because as there were no screws they did not have a function within this specific assembly context. The unexpectedly high frequency suggests that participants pay attention to constraints even if they are not crucial for the successful completion of the task. The findings support the assumption that clues, provided by the environment, are used if they facilitate the problem solving process by reducing the number of possibilities (Kirsh 2010).

By analyzing the nature and function of constraints in instructions, the importance of individual constraints was highlighted because instructors were expected to use constraints to facilitate the assembly process for their partner. Based on the previous observations, instructors were expected to use *boreholes* strategically, for indicating the spatial layout for example. Additionally, instructors were expected to refer to the *supporting system* and the *bevel* for distinguishing between objects.

The results highlighted that instructors referred to constraints more frequently (M = 8.50) than participants thinking aloud (M = 3.20) and reporting their activities (M = 2.22). Moreover, instructors mentioned the same constraints that were previously identified. This finding suggests that instructors used their own experiences as assemblers to facilitate the assembly process for their partners. The analysis revealed that instructors used constraints strategically for distinguishing between objects, for assigning function, or for describing objects. However, constraints were most frequently used to coordinate the placement of objects and to describe their orientation. This observation confirms the expectation that *boreholes* were used for indicating the spatial layout of the assembly structure and individual objects. Unexpectedly, the salient feature *supporting system* was hardly referred to at all and seldom to distinguish between objects. Instructors rather used *bevels* to point out differences between objects. If constraints were used to assign function to objects, which was rarely done, instructors always referred to *grooves*. At the level of description, instructors mentioned *grooves*, *boreholes*, and b*evels.* The observation that they mentioned those constraints to a comparable degree suggests that *bevels* were frequently described but its function was not made explicit.

In sum, the qualitative analysis highlighted a number of constraints that were assigned different functions in unaided object assembly. Moreover, the quantitative analysis confirmed the expectation that constraints are verbalized at different frequencies. However, not the most salient constraints were mentioned most frequently. The analysis rather highlighted that constraints were mentioned that were evaluated as meaningful because they served a specific function within the assembly or give off information as to possible object configurations or actions that can be performed. This finding can be generalized by proposing two factors that influence the active use of constraints; first, the meaningfulness that can be assigned to a constraint and second the importance that is assigned to a constraint by the assembler. In general, the analysis illustrated how additional insights can be gained by analyzing different kinds of verbal reports and how trends that show in one kind of verbal report can be supported and strengthened by trends observed in other kinds of verbal reports.

Given that this analysis, at least to my knowledge, is the first empirical study that investigates the nature and use of constraints the conclusions are tentative but as outlined in the discussion, the findings in different data sets make clear suggestions. Besides contributing first insights on the nature and function of constraints, this analysis presents one approach to the methodological challenges posed by analyzing everyday problem solving activities. Encouraged by the finding that Norman's (2002) categorization can be adopted to describe the general nature of the identified constraints, the presented methodology seems well suited to be adopted to analyze the use of constraints in other problem solving experiments as well.

9.7 Appendix

9.7.1 Nature and function of constraints in think aloud protocols

category	underspecified goal condition	verbal goal condition	verbal and visual goal condition
label	3 (M = 0.18, SD = 0.73)	1 (M = 0.06, SD = 0.25)	4 (M = 0.24, SD = 0.75)
supporting system	3 (M = 0.18, SD = 0.39)	4 (M = 0.25, SD = 0.58)	6 (M = 0.35, SD = 0.86)
groove	11 (M = 0.64, SD = 1.17)	3 (M = 0.19, SD = 0.54)	7 (M = 0.41, SD = 1.00)
picture	0	0	11 (M = 0.65, SD = 0.86)
borehole	16 (M = 0.94, SD = 2.01)	20 (M = 1.25, SD = 2.21)	16 (M = 0.94, SD = 1.56)
bevel	2 (M = 0.12, SD = 0.33)	19 (M = 1.19, SD = 2.32)	9 (M = 0.53, SD = 1.07)
connector	4 (M = 0.24, SD = 0.56)	4 (M = 0.24, SD = 0.58)	4 (M = 0.24, SD = 0.56)
connection	3 (M = 0.18, SD = 0.73)	4 (M = 0.25, SD = 0.68)	4 (M = 0.24, SD = 0.56)
no screwing	1 (M = 0.06, SD = 0.25)	1 (M = 0.19, SD = 0.54)	0
others	1 (M = 0.06, SD = 0.24)	0	0
all	44 (M = 2.59, SD = 3.14)	58 (M = 3.36, SD = 4.40)	58 (M = 3.41, SD = 3.66)

Table 9.28: Raw frequency, mean, and standard deviation of categories describing the nature of constraints in think aloud protocols.

category nature of constraints	underspecified goal condition	verbal goal condition	verbal and visual goal condition
label	0.4	-1.2	0.9
supporting system	-0.4	-0.5	0.9
groove	1.9	-1.8	-0.1
borehole	0.2	-0.1	-0.1
bevel	-2.3	2.1	-0.2
connector	0.2	-0.3	0.1
connection	0.4	0.5	-1.0
no screwing	-0.2	1.2	-1.1
others	1.3	-0.6	-0.6

Table 9.29: Standardized residuals for distribution of categories describing nature of constraints between conditions in think aloud protocols.

category	raw frequency	mean	standard deviation
semantic constraint	72	1.44	2.08

logical constraint	34	0.68	1.15
physical constraint	30	0.60	1.50
task induced constraint	15	0.30	0.65
cultural constraint	8	0.16	0.62
others	1	0.02	0.14
all	160	3.20	3.70

Table 9.30: Results frequency of categories describing the general nature of constraints in think aloud protocols; sorted from most to least frequent.

category	underspecified goal condition	verbal goal condition	verbal and visual goal condition
logical	14 (M = 0.82, SD = 1.33)	7 (M = 0.44, SD = 0.89)	13 (M = 0.76, SD = 1.20)
physical	2 (M = 0.12, SD = 0.33)	19 (M = 1.19, SD = 2.32)	9 (M = 0.53, SD = 1.07)
semantic	23 (M = 1.35, SD = 2.00)	28 (M = 1.75, SD = 2.43)	21 (M = 1.24, SD = 11.86)
cultural	3 (M = 0.18, SD = 0.73)	1 (M = 0.06, SD = 0.25)	4 (M = 0.24, SD = 0.75)
task induced	1 (M = 0.06, SD = 0.24)	3 (M = 0.19, SD = 0.55)	11 (M = 0.65, SD = 0.86)
others	1 (M = 0.06, SD = 0.24)	0	0
all	44 (M = 2.59, SD = 3.14)	58 (M = 3.36, SD = 4.40)	58 (M = 3.41, SD = 3.66)

Table 9.31: Raw frequency, mean, and standard deviation of categories describing the general nature of constraints in think aloud protocols.

category	definition	examples
insight	a) understanding the function of a feature b) assigning function to an object c) no connectors are provided	a) „ach deswegen hat der auch so ne Einkerbung" [2] b) „scheint mir so ne Art Verbindungsteil zu sein" [10] c) „es gibt wohl keine Stecker" [10]
strategy	a) recognizing and making use of object features such as holes or groove b) using world knowledge to search for things that may be useful	a) „dass die auf den Löchern drauf sind" [14]; „die haben hier extra solche Einkerbungen / damit ich da reinstecken kann" [23] b) „ich bräuchte irgendwelche Bolzen oder so" [24]; „wo finde ich Verbindungsteile?" [20]
perception	a) object features are perceived b) objects and its assigned function are perceived	a) „und dann hier dieses auch mit so ner Einkerbung" [8] b) „und oh hier ist ne schöne Steckverbindung über" [17]

puzzlement	asking oneself about the function of a perceived feature or placement of an object	„diese Noppen sollen die unten stehen?“ [10]; „und wozu sind diese Löcher hier?“ [44]
statement	a) stating a fact b) evaluative statement about an object referring to the memorized picture c) repetition of task instructions	a) „auf dem Bild sah das gar nicht so schwer aus“ [14] b) „aber so ähnlich wie auf dem Bild sieht das aus“ [27] c) „ohne Schrauben“ [31]

Table 9.32: Classification of functions identified in think aloud protocols; examples in the German original.

category	underspecified goal condition	verbal goal condition	verbal and visual goal condition
insight	10 (M = 0.59, SD = 0.87)	15 (M = 0.94, SD = 1.77)	17 (M = 1.00, SD = 1.27)
strategy	11 (M = 0.65, SD = 1.27)	13 (M = 0.81, SD = 1.22)	13 (M = 0.76, SD = 1.15)
perception	17 (M = 1.00, SD = 1.66)	22 (M = 1.38, SD = 1.31)	11 (M = 0.65, SD = 0.93)
puzzlement	4 (M = 0.24, SD = 0.44)	3 (M = 0.19, SD = 0.54)	7 (M = 0.41, SD = 0.71)
statement	1 (M = 0.06, SD = 0.24)	5 (M = 0.31, SD = 0.70)	10 (M = 0.59, SD = 0.71)
all	44 (M = 2.59, SD = 3.14)	58 (M = 3.63, SD = 4.40)	58 (M = 3.41, SD = 3.66)

Table 9.33: Raw frequency, mean, and standard deviation of categories describing the function of constraints in think aloud protocols.

9.7.2 Nature and function of constraints in retrospective reports

category	underspecified goal condition	verbal goal condition	verbal and visual goal condition
supporting system	4 (M = 0.24, SD = 0.56)	5 (M = 0.31, SD = 0.48)	10 (M = 0.59, SD = 0.87)
groove	1 (M = 0.06, SD = 0.24)	2 (M = 0.13, SD = 0.34)	5 (M = 0.29, SD = 0.47)
picture	0	0	6 (M = 0.35, SD = 0.49)
borehole	2 (M = 0.12, SD = 0.33)	8 (M = 0.50, SD = 0.82)	7 (M = 0.41, SD = 0.71)
bevel	6 (M = 0.35, SD = 0.79)	7 (M = 0.44, SD = 0.73)	12 (M = 0.71, SD = 0.69)
connection	2 (M = 0.12, SD = 0.49)	0	1 (M = 0.06, SD = 0.24)
no screwing	0	2 (M = 0.13, SD = 0.34)	1 (M = 0.06, SD = 0.24)
others	1 (M = 0.06, SD = 0.25)	1 (M = 0.06, SD = 0.25)	5 (M = 0.29, SD = 0.59)
roof	8 (M = 0.47, SD = 0.51)	2 (M = 0.13, SD = 0.34)	4 (M = 0.24, SD = 0.44)
analogy	1 (M = 0.06, SD = 0.25)	2 (M = 0.13, SD = 0.34)	4 (M = 0.24, SD = 0.44)
all	27 (M = 1.59, SD = 1.50)	29 (M = 1.81, SD = 1.94)	55 (M = 3.24, SD = 2.39)

Table 9.34: Raw frequency, mean, and standard deviation of categories describing the nature of constraints in retrospective reports.

category	underspecified goal condition	verbal goal condition	verbal and visual goal condition
logical	6 (M = 0.35, SD = 0.61)	9 (M = 0.56, SD = 0.73)	19 (M = 1.12, SD = 1.41)
physical	6 (M = 0.35, SD = 0.79)	7 (M = 0.44, SD = 0.73)	12 (M = 0.71, SD = 0.69)
semantic	4 (M = 0.24, SD = 0.56)	8 (M = 0.50, SD = 0.82)	8 (M = 0.47, SD = 0.72)
cultural	8 (M = 0.47, SD = 0.51)	2 (M = 0.13, SD = 0.34)	4 (M = 0.24, SD = 0.44)
task induced	2 (M = 0.12, SD = 0.49)	2 (M = 0.13, SD = 0.34)	7 (M = 0.41, SD = 0.51)
others	1 (M = 0.06, SD = 0.24)	1 (M = 0.06, SD = 0.24)	5 (M = 0.29, SD = 0.59)
all	27 (M = 1.59, SD = 1.50)	29 (M = 1.81, SD = 1.94)	55 (M = 3.24, SD = 2.39)

Table 9.35: Raw frequency, mean, and standard deviation of categories describing the general nature of constraints in think aloud protocols.

category	definition	examples
description	a) object features are described b) general thoughts about observed features are reported	a) „mir ist aufgefallen, dass hier diese Halbkugeln dran sind“ [16] b) „dann hab ich mir Gedanken darüber gemacht wies wohl ist ohne schrauben äh ohne Schrauben klar zu kommen weil ja die Löcher ähm in den Bodenteilen vorhanden sind“ [31]
perception	features are reported to have been observed	„weil ich ja auch gesehen habe, dass es da irgendwie ähm Löcher für Schrauben gibt“ [21] „ich hatte zuvor aber erstmal die Ecken erkannt“ [27]
distinguish	object features are used to distinguish between objects	„dann fiel mir auf, dass dann äh einige Platten so abgeschrägt sind und die dachte ich ähm und die anderen gerade“ [25] „danach hab ich dann diese äh ohne Schrägen Dinger eingebaut“ [31]
function assignment	a) function is assigned to objects by means of constraints b) function of constraints is explained	a) „zum Beispiel dieses rote Dach hat mich an n Dach erinnert“ [10]; „also mir ist halt relativ schnell aufgefallen, dass es halt hier Kanten so Schnittkanten hat, dass es halt zum Dach gehören muss“ [18] b) „man kann sie also schlecht wenn mans dieses Verbindungsstück dazwischenstecken will irgendwie anders anordnen“ [19]; „bin ich ja relativ schnell drauf gekommen, dass die schrägen Seiten nach oben müssen damit das Dach halten und

		dass das Ganze zweitsöckig ist" [23]
insight	conclusions about objects are based on perceived constraints	„bin dann auf die Idee gekommen dass die teile die ich unten eingebaut hab wahrscheinlich nach oben kommen weil die angeschrägt sind so wie das dass und weil die Form dann wieder ineinander passte" [47]; „allerdings waren die Wände da nicht außen abgeflacht und äh das Dach hätte nicht raufgepasst also muss-ten logischerweise die Wände mit den Kugeln auf den Säulen nach unten" [59]
puzzlement	constraints cannot be conceptualized and leave the assembler puzzled	„war die ganze Zeit irritiert von diesen Ku-geln hier weil ich immer dachte so die sind so schön abgerundet die müssen ja irgend-wo oben ähm wahrscheinlich sein" [20]; „und ich hab mich zwischendurch gefragt warum die einen hier rund sind und die anderen nicht" [22]; „äh Verwirrung hat bei mir gestiftet, dass mache solchen Pinökel oben drauf haben andere nicht" [50]
strategy	strategic use of constraints by a) consciously disregarding them b) using them for orientation e.g. in putting & arranging objects c) using them to sort objects	a) „hab dann irgendwie auch das Bild son bisschen aus dem Kopf gelegt" [20] b) „dann hab ich äh mich an den Kugeln an diesen Kugeln oben drauf orientier und hab gedacht ok das ist dann da-rauf können die nicht stehen dann weiß ich wo oben und unten ist" [24]; „die zusammengesteckt, so dass die Löcher passen" [26]; „ kam irgend-wann der Durchbruch irgendwas ach ja genau ich hab das geordnet nach den Löchern und hab geguckt wo was reinpasst" [32] c) „genau dann ham wir halt versucht schrägen Wände und die äh Kugeln zu trennen" [16]; „und als nächstes hat-te ich mich an diesen Schrägen die sind mir als nächstes aufgefallen und dann hab ich gedacht alles klar das muss irgendwie mit dem Dach zusam-menhängen hab dann eben sortiert in ähm die Teile die vom Obergeschoss sein müssen und vom unteren" [24]

Table 9.36: Categories of functions identified in retrospective reports: definitions and examples (German original wording).

category	underspecified goal	verbal goal condition	verbal and visual goal

	condition		condition
description	1 (M = 0.06, SD = 0.24)	4 (M = 0.25, SD = 0.45)	5 (M = 0.29, SD = 0.59)
perception	2 (M = 0.12, SD = 0.49)	0	1 (M = 0.06, SD = 0.24)
distinguish	4 (M = 0.24, SD = 0.56)	2 (M = 0.13, SD = 0.34)	3 (M = 0.18, SD = 0.53)
function assignment	16 (M = 0.94, SD = 0.90)	9 (M = 0.56, SD = 0.81)	19 (M = 1.12, SD = 0.62)
insight	1 (M = 0.06, SD = 0.24)	5 (M = 0.31, SD = 0.48)	7 (M = 0.41, SD = 0.62)
puzzlement	2 (M = 0.12, SD = 0.33)	2 (M = 0.13, SD = 0.34)	5 (M = 0.29, SD = 0.59)
strategy	1 (M = 0.06, SD = 0.24)	7 (M = 0.44, SD = 0.73)	15 (M = 0.88, SD = 0.99)
all	27 (M = 1.59, SD = 1.50)	29 (M = 1.81, SD = 1.94)	55 (M = 3.24, SD = 2.39)

Table 9.37: Raw frequency, mean, and standard deviation of categories describing the function of constraints in retrospective reports.

9.7.3 Nature and function of constraints in instructions

category	raw frequency	mean	standard deviation
logical constraint	26	1.63	1.15
physical constraint	33	2.06	2.95
semantic constraint	73	4.56	4.82
cultural constraint	0		
task induced constraint	0		
others	4	0.25	0.77
all	136	8.5	5.72

Table 9.38: Frequency of categories defining constraints in general terms in instructions; raw frequency, mean, and standard deviation.

category	definition	examples
description	object features are pointed out the current assembly status is described	„wie du siehst haben die hier so eine Rille, eine Fuge“ [2] „hier sind noch keine Schrägen eingebaut“ [3]
distinguish	Features are highlighted to distinguish the object in focus from others. This is done by means of comparison or contrast.	„im Prinzip all die Elemente, die wir gerad schon hatten nur mit Schrägen dran“ [3]
function assignment	Highlighting or mentioning the function that a feature serves in the whole assembly.	„in diese Einkerbungen kommen die beiden Holz kommen die beiden äh Rückbauelemente“ [3]

orientation	Object features are used to explain the orientation of an object.	„dass du auf der rechten Seite zwei Löcher vor dir hast“ [2]
placement	Object features are used to describe where and how to position an object in relation to others.	„genau auf diese drei Löcher“ [2]; „und schieben die Bretter links und rechts von der Fräsnaht in die blauen Stangen rein“ [6]
comment about self	Comment about instructor’s own assembly process.	„und bei den Schrägen habe ich mich ständig vertan“ [5]; „ich hab noch darauf geachtet, dass diese Löcher unten sind an dem Brett“ [5]
strategy	Point out object features and explain their strategic use for the assembly process.	„dass wir die Bohrlöcher beachten“ [20]

Table 9.39: Classification of functions identified in instructions, examples in German.

9.7.4 Nature and function of constraints between verbal reports

kind of verbal report	borehole	bevel	groove
think aloud	0	-0.4	0.5
instruction	1.0	-1.1	-0.2
retrospective report	-1.6	2.4	-0.4

Table 9.40: Standardized residuals for distribution of categories describing nature of constraints between investigated kinds of verbal reports.

kind of verbal report	semantic constraints	logical constraints	physical constraints
think aloud	0.9	-0.5	-0.8
instruction	-2.9	2.7	1.1
retrospective report	1.3	-1.6	-0.1

Table 9.41: Standardized residuals for distribution of categories describing the general nature of constraints between investigated kinds of verbal reports.

kind of verbal report	insight	strategy
think aloud	4.1	-2.6
instruction comb.	-4.7	3.0
retrospective report	0.9	-0.5

Table 9.42: Standardized residuals for distribution of categories describing functions of constraints between investigated kinds of verbal reports.

9.7.5 Investigating the nature of constraints within functions

	supporting system	groove	borehole	bevel	analogy	roof	label	picture	connector	connection
insight	0	5.5 (3.54)	5.0 (5.66)	9.0 (0.66)	0.5 (0.71)	0	0.5 (0.71)	0.5 (0.71)	3.5 (4.95)	1.0 (1.41)
strategy	0.67 (0.58)	0.33 (0.58)	12.0 (8.55)	3.0 (2.0)	1.33 (2.31)	0	0.67 (1.15)	1.0 (1.0)	1.0 (1.73)	0.67 (1.15)
perception	4.0 (4.24)	6.0 (8.49)	9.5 (12.02)	4.0 (5.66)	0	0	2.0 (2.83)	0	0	1.5 (2.12)
puzzlement	4.0 (1.41)	0	2.5 (2.12)	0	0	0	0.5 (0.71)	0	1.0 (1.41)	0.5 (0.71)
description	0	6.0 (8.49)	7.0 (8.49)	6.5 (7.78)	0.5 (0.71)	0	0	1.5 (2.12)	0	0.5 (0.71)
distinguish	2.50 (3.54)	0.5 (0.71)	0.5 (0.71)	5.0 (2.83)	0.5 (0.71)	0	0	0	0	0
function assignment	2.50 (3.54)	5.5 (0.71)	1.0 (1.41)	6.0 (8.49)	0	7.0 (9.90)	0	0.5 (0.71)	0	1.0 (1.41)
strategy combined	1.33 (0.58)	1.67 (2.08)	28.67 (22.50)	6.33 (4.16)	1.33 (2.31)	0	0.67 (1.15)	1.0 (1.0)	1.0 (1.73)	1.0 (1.0)

Table 9.43: Overview frequency of constraints within functions; mean and standard deviation.

	supporting	groove	borehole	bevel	analogy	roof	label	picture	connector	connection
insight	1	11	10	18	1	0	1	1	7	2
strategy	2	1	36	9	4	0	2	3	3	2
perception	8	12	19	8	0	0	4	0	0	3
puzzlement	8	0	5	2	0	0	1	2	2	1
description	2	12	14	13	1	0	0	3	0	1
distinguish	5	1	1	10	1	0	0	0	0	0

function assignment	5	11	2	12	0	0	0	1	0	2
strategy combined	2	4	54	11	0	0	0	0	0	1

Table 9.44: Overview frequency of constraints within functions; raw frequency.

nature of constraints	insight	perception	puzzlement	description	distinguish	function assignment	strategy (combined)
groove	0.7	1.1	-0.8	1.1	-1.0	2.3	-2,9
borehole	-1.7	0.4	-1.2	-0.8	-1.9	-2.7	4.2
bevel	1.4	-1.4	2.0	0	3.0	1.3	-2.5

Table 9.45: Standardized residuals for distribution of categories describing the nature of constraints within categories describing the function of constraints.

9.7.6 Frequency of constraints across assembly process

condition	1st quarter	2nd quarter	3rd quarter	4th quarter
underspecified goal	0.3	0.9	0	-1.3
verbal goal	-1.7	-0.9	2.0	0.1
verbal and visual goal	1.5	0.2	-2.0	1.0

Table 9.46: Standardized residuals for frequency of constraints within the four quarters of the assembly process.

condition	1st half	2nd half
underspecified goal	23 (52.27%)	21 (47.73%)
verbal goal	16 (27.59%)	42 (72.41%)
verbal and visual goal	31 (53.45%)	27 (46.55%)
all	70 (43.75%)	90 (56.25%)

Table 9.47: Frequency of constraints within the two parts of the assembly process, raw frequency and percentage.

9.7.7 Answers to guided questions

answers	raw frequency	mean	standard deviation
noticed groove	43	0.86	0.35
groove unnoticed	3	0.06	0.24

used groove	18	0.36	0.48
groove unused	14	0.28	0.45
groove partially used	3	0.06	0.24
groove irritated	7	0.14	0.35
uncertain	1	0.02	0.14

Table 9.48: Answers to the first interview question, i.e. "Did you notice this groove? And did you use it as a guide?"; raw frequency, mean, and standard deviation.

answers	raw frequency	mean	standard deviation
noticed borehole	46	0.92	0.27
borehole unnoticed	0	0	0
used borehole	21	0.42	0.50
borehole unused	20	0.40	0.50
borehole partially used	5	0.10	0.30

Table 9.49: Answers to the second interview question, i.e. "Did you notice the boreholes? And did you use them as a guide?"; raw frequency, mean, and standard deviation.

answers	raw frequency	mean	standard deviation
noticed label	29	0.58	0.45
label unnoticed	17	0.34	0.48
used label	13	0.26	0.44
label unused	13	0.26	0.44
label not helpful	3	0.06	0.24

Table 9.50: Answers to the third interview question, i.e. "Did you notice this label? And did you use it as a guide?"; raw frequency, mean, and standard deviation.

10 Conclusion

10.1 Introduction

This thesis investigated problem solving processes in unaided object assembly and their linguistic representation. In the experimental studies participants were faced with the task of assembling a two-story dollhouse. As they were not provided with a manual, they needed to figure out the function of each of the 13 wooden objects and their arrangement by themselves.

Studying performance on assembly tasks is scientifically interesting because people engage in assembly tasks, especially furniture assembly, quite frequently (Richardson 2007). In this field of research, most researchers tended to focus either on manual design with the aim to create better manuals (e.g. Marcus et al. 1996; Maes et al. 2004; Tversky et al. 2009; Daniel & Tversky 2012) or on general procedures (e.g. Helander & Willén 2003; Richardson et al. 2004). Understanding of general procedures is specifically interesting with regard to the automation of assembly steps. A few months ago, for example, computer scientists demonstrated that two robots can jointly assemble a coffee table sold at the IKEA furniture store

(Knepper et al. 2013). This table consisted of five wooden objects; those were four legs and the tabletop, i.e. a square board in descriptive terms. The robots did not use the instructions but figured out how objects needed to be arranged and fastened by themselves; the computer script simply contained general operations that are necessary in an assembly task.

This thesis extended previous research on manuals and general procedures by investigating mental processes that are involved when humans solve an assembly task. Mental processes were studied by eliciting verbalizations of thought during assembly performance, i.e. think aloud protocols, and analyzing those verbalizations concerning their content as well as their linguistic structure, i.e. verb phrases, noun phrase, and discourse markers. These protocols were supplemented by other kinds of verbal reports and performance measures, such as assembly success and time. The aim of this extended approach was two-fold. First, it was intended to shed light on people's performance by studying which problem solving processes are verbalized and identify structural patterns describing their occurrence. Second, the linguistic analysis aimed at describing features of problem solving processes to highlight systematic recurring patterns. Methodologically, these aims were approached by combining content-based *protocol analysis* with *Cognitive Discourse Analysis* that focuses on the investigation of linguistic features in verbal data.

10.2 Main findings

Think aloud protocols are widely used in problem solving research (e.g. de Groot 1969; Newell & Simon 1972; Roth 1985; Bartl & Dörner 1998). It is assumed that verbalizations in think aloud protocols represent traces of problem solving processes that are generated while solving the task (e.g. Ericsson & Simon 1993). Traditionally, problem solving research focused on the investigation of logic based and well-defined problem solving tasks, such as chess or the Tower of Hanoi. More recent studies investigate problem solving behavior on real world problems such as navigation through an unfamiliar environment (e.g. Tenbrink et al. 2011) or collectively solving a problem in an office environment (e.g. Steffensen 2013). Similar to the approach presented in this thesis, the analysis of think aloud protocols or transcripts of conversations was supplemented by analyses of other kinds of verbal reports and further performance measures in these studies.

The use of think aloud protocols raises the controversially debated question of reactivity, i.e. the influence of verbalization on task performance. Studies investigating this kind of reactivity yielded different results (e.g. Ahlum-Heath & Di Vesta 1986; Deffner 1989; Short et al. 1993; Schooler et al. 1994; Gilhooly et al. 2010). However, it is generally agreed that the wording of the instructions and the nature of the task determine if there is an effect and in which direction, i.e. positive or negative. Therefore, researchers are encouraged to collect a silent control group to compare their performance to that of participants thinking aloud. In the presented experiments, the comparison of the mean assembly times recorded for the silent control group to those recorded for the group assembling while verbalizing their thoughts revealed that participants thinking aloud were significantly faster. This finding suggests that verbalization of thoughts facilitated assembly performance. However, verbalization did not affect assembly success. This observation suggests that verbalization speeded up performance but it did not influence performance on the conceptual level.

10.2.1 Problem solving processes in unaided object assembly and their linguistic representation

The analysis of think aloud protocols revealed seven problem solving processes, out of which five are well described in the literature, namely *hypothesis*, *action*, *evaluation*, *dead end*, and *fresh start*. Additionally, two categories were identified that are not described in traditional accounts, i.e. *description of mental state* and *comment on object features*.

In the present problem context, *hypothesis* was the most frequently verbalized process. *Hypotheses* concerned different aspects of the assembly process. These aspects were related either to the level of conceptualization of objects, the level of planning, or to the level of physical manipulation. At the level of conceptualization, assemblers proposed functions of objects, such as "presumably those are the floors and those over here are the walls" [16]. This specific aspect of explicit function assignment was investigated in more detail. The linguistic analysis revealed two different types of explicit function assignment that express different levels of certainty. Function was either assigned with high certainty by stating "those **are** the floors" (*direct mapping*) or with a tentative character as in "those **look like** walls" (*representational mapping*). In *direct mapping phrases* the verb 'sein' (be) was most frequent indicating that the assigned nature is assumed as given. In some cases, participants used markers such as modal verbs or modal adverbs, as 'presumably' in the example above, to weaken the proposal. In *representational mapping phrases* verbs such as ‘aussehen wie’ (look like) expressed function assignment based on comparisons or verbs such as ‘nehmen als’ (use as) highlighted that the speaker’s decision was related to the current situation.

At the level of planning, future actions were specified as in „ok dann ähm kommt das natürlich alles aufs zweite“ (“ok then uhm this goes on the second”) [16]. In this example, ‘second’ refers to the second story and the speaker recognizes that some objects need to be placed on the second story. At the level of physical manipulation, hypotheses expressed possible configurations of objects or their position, e.g. “probably this way in order for the walls to stand like this” [16]. The linguistic analysis revealed that *hypotheses* were commonly marked by verbs of *being and having*, such as ‘be’, ‘look like’, and ‘stand’ in the given examples. Verbs were most frequently marked by the present tense and indicative mood. In some cases, the subjunctive mood was used to indicate the tentative character of the formulated assumption or proposal. This character was emphasized by the modal verb ‘können’ (could), modal adverbs that signal tentativeness (e.g. ‘vermutlich’ (presumably), ‘wahrscheinlich’ (probably)), and adverbs that mark indefiniteness, such as ‘einige’ (some), ‘irgendwie’ (somehow), ‘irgendwo’ (somewhere), and ‘irgendwas’ (something).

The second most frequently observed process was that of *evaluation*. This superordinate category comprises two categories that can be distinguished based on the conclusions that are drawn, namely positive evaluations and negative evaluations. In the given context, different aspects of the assembly task were evaluated. First, participants evaluated their understanding of the task or more specifically of the structure that they were instructed to assemble, e.g. “suddenly everything becomes clearer” [27] in the sense of ‘now everything makes sense’. Second, assemblers evaluated the possibility that objects could be inserted or fastened, e.g. “yes they fit in here” [27], and that the structure would be stable, e.g. “now it does not bounce that much anymore” [27]. Third, assemblers evaluated sub-assemblies, e.g. “and then we have assembled the basement already” [58].

As illustrated by the examples, evaluations were commonly marked by verbs of *being and having* in the present tense. This highlights that assemblers evaluated current states. Positive evaluations frequently contained the discourse marker ‘so’ (so) signaling confirmation of previous expectations (Erben 1972) as well as the ending of sub-processes (Tenbrink 2008). This finding strengthens the interpretation that participants focused on the present situation and manipulated objects in the physical world rather than mentally simulating possible actions. Additionally, affirmative particles such as ‘gut’ (good) along with ‘sehr gut’ (very good), and the comparative ‘besser’ (better) were identified in numerous positive evaluations. Negative evaluations frequently contained negations such as ‘nein’ (no) or ‘keine’ (none) and adverbs signaling uncertainty; ‘irgendwie’ (somehow) was specifically prominent.

Besides hypotheses and evaluations, *actions* were frequently verbalized. The majority of actions started at the moment of speaking. Only few instances of planning were identified. Planning phrases indicated that sub-processes needed to be performed. In the example “thus I will work on the second story now” [24] a plan was verbalized that involved numerous actions although they were not specifically spelled out at that moment. Immediate actions were expressed by phrases such as “this I put in the edge now” [27]. As expected, the linguistic analysis highlighted that *actions* were marked by verbs of *doing and happening*, such as ‘work’ or ‘put’ in the present tense and indicative mood as shown in the examples. The future tense was rarely used in the investigated assembly context. Additionally, actions were marked by temporal markers that stress that the present situation is the reference frame, e.g. ‘jetzt’ (now).

Dead ends and *fresh starts* were rarely observed in the analyzed data. This could be an effect of the experimental design because participants were allowed to move objects freely and all actions could be redone if they proofed to be wrong. Some protocols contained utterances in which helplessness was expressed, e.g. “by now I am ready to give up” [32]. In some cases, assemblers verbalized reasons for this dead end state, e.g. “and at the moment I see I do not know how to proceed” [54] or “I do not know where to start” [37]. In some *fresh starts* participants explicitly verbalized the decision to start over again, e.g. “ok maybe I start over again” [62] or “back again” [59].

In sum, think aloud protocols contain much information on hypotheses concerning function and position of objects, evaluations of achieved sates as well as actions that are performed. These are components of the general ‘search and test’ procedure as proposed by Newell and Simon (1972). The combination of existing information along with the generation of new information in hypotheses expresses the verbalization of the search process through the problem space. The verbalization of actions resembles the execution process within the ‘test’ procedure that is completed by the evaluation process. Processes that were less frequently observed, such as *dead end* and *fresh start*, resemble alternative ending or beginning sequences of a ‘search and test’ procedure. However, it seems that these were rarely verbalized because of the given reversible nature of the task.

In addition to these processes, two processes were frequently verbalized that were not described in the literature before. First, the content-based analysis revealed numerous instances in which assemblers described their mental state, e.g. “I am curious about the parts that are in the box here” [10] or “a little difficult (...) I find it a little difficult because I don’t if these parts shall be up or down” [18]. As these examples illustrate, *descriptions of mental state* contain statements about the assembler’s consciousness. The linguistic analysis revealed that those statements were frequently marked by verbs of *sensing* such as ‘think’, ‘believe’, and ‘hope’ in present tense and indicative mood. The frequent use of the personal pronoun ‘I’ signals that the speaker refers to himself/herself as the source of information as in “right now **I** am asking myself ...” [10] or “what is written there **I** do not know” [18]. As the first example illustrates, emotions were also expressed in *descriptions of mental state*, in this example the assembler refers to curiosity.

Second, assemblers also mentioned object features such as color, shape, or other physical properties, e.g. “this looks like an edge” [27] or “so uh there are these half like balls up here” [16] that were based on perceptual information. Sometimes object features were commented on by stating that a specific feature that was previously encountered was missing, e.g. “surprisingly there is no such notch here” [25]. Some objects were expected in the present context based on prior knowledge or because of physical properties but those were not provided, e.g. “apparently there are no connecting devices” [16]. This comment on a missing object implies that the assembler expected to find connecting devices of some sort. *Comments on object features* were frequently marked by verbs of *being and having* in present tense and indicative mood supporting the assumption that participants verbalized noticing features at the moment of perception or shortly after.

These additional processes highlight that participants followed the instruction to verbalize whatever came to their minds even feelings that are raised. This suggests that participants did not select information that is not verbalized in public normally, such as turning to oneself for an answer. This impression is strengthened by those cases in which participants admitted that they felt overwhelmed and did not know how to proceed, i.e. in *dead end states*. The verbalization of object features suggests that participants verbalized perceptual information when perceiving it. Moreover, the analysis of process sequences highlighted that such comments were frequently followed by a *hypothesis*. This pattern suggests that participants integrated these features into the existing mental representation and used them as cues for processes such as function assignment. This aspect is further investigated in the analysis of constraints that will be reviewed in the next subsection.

Besides identifying problem solving processes and describing their linguistic structure, the annotated data was analyzed to highlight recurring sequences of processes. The analysis highlighted traditionally assumed sequences of processes that are generally described as a 'search and test' procedure (e.g. Newell & Simon 1972). Moreover, the analysis revealed that the expected sequence of *hypothesis* ⟶ *action* ⟶ *evaluation* was very frequent. As outlined above, hypotheses represent expressions of the search process in which potential arrangements and conceptualizations of objects are generated, i.e. referred to as 'new states' in Newell and Simon (1972). The verbalization of actions reflects the instantiation of these new states. Combined with the evaluation of the created state, this execution of actions represents the proposed test procedure. Furthermore, it was possible to extract common four process sequences, such as *hypothesis* ⟶ *action* ⟶ *evaluation* ⟶ *hypothesis*. This sequence suggests that the closing of one 'search and test' procedure was followed by a new 'search' process. This sequence represents the cyclic nature of problem solving processes that is proposed in the literature (e.g. de Groot 1969; Dörner 1987).

However, it needs to be stressed that a great variability of process sequences was observed. Besides these commonly proposed sequences, the analysis revealed that processes were repeated before new processes were verbalized, e.g. *hypothesis* ⟶ *action* ⟶ *action* ⟶ *hypothesis*. Due to this great variability, it was difficult to assess the prominence of individual combinations because frequency of individual combinations decreased drastically by including one more process into the analysis, i.e. analyzing sequences containing three instead of two processes. This observation illustrates the difficulties that are encountered if processes are analyzed in more than one protocol. However, it also suggests that there may be more variety regarding the structure of the problem solving process than previously assumed. This interpretation can be supported by findings reported by Wedman et al. (1996). Their study of processes verbalized during analogical problem solving highlighted that expected processes were missing and they did not offer a generalization of the identified patterns although this was one of the proposed research aims. Although, Wedman et al. (1996) do not explicitly discuss this missing result, it can be suspected that the observed variance did not allow for valid abstractions.

An alternative explanation is to be found in the nature of the analyzed data set, i.e. think aloud protocols. As is generally acknowledged (e.g. Ericsson & Simon 1993; van Someren et al. 1994), think aloud protocols can never be complete in the sense that all thoughts and therefore all mental processes are reported. Thus, it is possible that some processes were not verbalized. However, it can be argued that this effect was compensated for by the large number of processes that was analyzed, i.e. 2,756 processes. Nonetheless, it could be possible that all 50 participants systematically did not verbalize some processes. However, this possibility seems unlikely given the finding that all problem solving processes that are reported in the literature were identified.

In sum, this review of the major findings highlights that problem solving processes that are commonly assumed in the literature (e.g. Newell & Simon 1972; Dörner 1989) can be identified in an assembly task as well. This observation suggests that solving a real world problem involves the same basic problem solving processes as solving a logic based problem. The linguistic analysis revealed systematic patterns of verb type (modal as well as general), modal adverbs, temporal markers, personal pronouns, and discourse markers for those processes that were most frequently verbalized. Importantly, these patterns are observed in natural language that was unconstrained and can therefore be expected to reflect participants' thoughts as closely as possible. Furthermore, these patterns emerged in the analysis of a large data set, i.e. 50 think aloud protocols, and could be related to previous studies, hence it is proposed that these linguistic features represent general process markers.

10.2.2 On functional conceptualization of objects and limiting possible actions by constraints

One important step within problem solving is the functional conceptualization of objects. Defining the function of an object within the goal structure resembles a new state within the problem space. These conceptualizations are tested and if they are evaluated to be useful and adequate, they are adopted. So far, conceptualizations of objects were studied as phenomena of addressee design, i.e. in communicative settings (e.g. Clark & Wilkes-Gibbs 1986; von Stutterheim et al. 1993; Mangold-Allwinn et al. 1995; Rieser 1995). In these experimental designs, objects need to be identified in order to assemble larger structures or to solve matching tasks successfully. Success is only possible if speaker and addressee agree on one reference to an object. Therefore, the reference represents the agreed and shared conceptualization (for more detail see chapter 8). In experiments in which think aloud protocols are elicited, low communicative intention is assumed because no addressee is present or explicitly introduced. Thus, it is proposed that object references in think aloud protocols express the speaker's current mental representation of the referred object.

The analysis of noun phrases revealed that participants generally tended to refer to objects by domain unspecific nominal references, such as 'Ding' (thing) or 'Teil' (part). However, one object was frequently referred to by a house specific reference, namely the roof part, i.e. 'Dach'. This result highlights that the objects themselves, except for the roof part, were not prototypical examples of house specific parts, such as wall, floor, or ceiling. Overall, participants referred to half of all objects by domain specific terminology at least once during the assembly. This finding suggests that objects were conceptualized within the assembly context but not as frequently as expected. These observations provide a static view on conceptualization throughout the entire assembly process.

The analysis of referential change, in contrast, provides a dynamic view. The qualitative analysis of referential change from first to fifth reference revealed five categories that described this change. These categories were adopted from literature on the referential production in communicative settings and refined for the specific framework of think aloud protocols. The analysis revealed that references were either *reduced* (e.g. 'Ding' (thing) to 'das' (it)), *repeated*, *rephrased* within the same frame of reference (e.g. 'Ding' (thing) to 'Teil' (part)), *reframed* from unspecific to specific (e.g. 'Ding' (thing) to 'Wand' (wall)), or previously mentioned references were *reused*.

It is proposed that these categories indicate different changes in conceptualization of objects. Whereas *repeating*, *rephrasing*, and *reusing* signal persistence of conceptualization, *reducing* signals a change in familiarity from unfamiliar within the given context to familiar and in focus. *Reframing* signals change of conceptualization by means of function assignment. A frequency analysis highlighted that participants tended to reduce, repeat, or rephrase subsequent references. The assemblers' preference to repeat subsequent references signals that conceptualizations tended to be maintained. The observation that references were reduced from first to second mention most frequently highlights that familiarity was established early in the assembly process. Moreover, Horton and Gerrig (2002) pointed out that consistency of expression is a principle of speech economy. Interpreting this phenomenon from the speaker's internal production perspective, Bock (1986:379) argued that people used the same structure repeatedly in cases in which production of new structures would result in hesitations and pauses.

Fukumura and van Gompel (2012) stated more explicitly that "repeating the recently mentioned word is usually easier than producing a word that has not been mentioned" (Fukumura & van Gompel 2012:1305). These observations support the assumption that assemblers did not tailor their utterances for an imagined addressee but relied on their own discourse model throughout the assembly. Furthermore, the findings suggest that participants were very engaged in solving the assembly task and did not try to find different, possibly more adequate references to the provided objects.

Concerning the methodological level, these findings stress that object conceptualizations can be investigated by means of linguistic analyses of referential form. However, different aspects such as referential form and referential change need to be combined to highlight and interpret general trends, such as maintenance of referential form.

Besides studying conceptualization of objects, it was investigated which object and features of objects guided participants concerning this conceptualization as well as location and arrangement of objects. Researchers working in the framework of *situated cognition* pointed out that problem solvers use constraints to limit the selection of possible operations within the problem space (e.g. Norman 2002; Kirsh 2010; Steffensen 2013). The analysis of constraints mentioned in the collected data highlighted nine constraints; those were mentioned at different frequencies. On the descriptive level, constraints referred to physical properties of the provided wooden objects (i.e. *bevels*, *boreholes,* little half-ball like wooden pieces (*supporting system*), *grooves*), the object itself (i.e. the *roof* and the *picture*), or object features (i.e. the *label* 'Selecta Spielzeug'). Furthermore, one constraint was stated explicitly in the instructions, namely that *no screws* can be used.

The frequency analysis revealed that constraints were mentioned at different frequencies based on their nature as well as with regard to the kind of verbal report that was analyzed. Participants thinking aloud mentioned *boreholes* significantly more frequently than all other constraints. *Bevels* and *grooves* ranked second and third. Participants who recalled their thoughts, in contrast, mentioned *bevels* most frequently. The *supporting system* ranked second and the *boreholes* were mentioned in third position. The reversed ranking of *boreholes* could be explained by the observation that participants tended to mention everything that came to their mind in think aloud protocols. *Boreholes* were salient for two reasons. First, they were perceptually salient because they created a specific spatial layout on the boards. Second, they could be expected to be functionally salient because they are generally important in an assembly task as providing the fastening points, i.e. where the screws need to be inserted. Based on these considerations, it can be argued that *boreholes* were mentioned when they were perceived but as they did not have their expected, conventionalized function in the given context, they were forgotten and thus not reported in retrospective reports. This interpretation is supported by the answers to the guided question at the end of the experimental session. Those revealed that the great majority of participants noticed the *boreholes* but only half of them used them in some way. In the same line of thought, the finding that *grooves* were rarely mentioned in retrospective reports might be the result of a memory effect. If the function of a *groove* is culturally well known, then noticing the *groove* and its usage is forgotten after the assembly because it is not a prominent thought during the assembly. Concerning the methodological level, these findings highlight that different kinds of verbal reports need to be elicited to gain information at different levels, such as immediate perception and memory. Moreover, future studies should include a second measure such as eye movements to support the suggested interpretations.

Besides describing the nature of the identified constraints on a qualitative level and their general distribution on a quantitative level, the function of constraints was studied. Concerning the unexpected prominence of *boreholes* it was specifically interesting to show which function *boreholes* served. The analysis revealed that *boreholes* were reported to be used strategically significantly more often than *bevels* and *grooves*. Strategic use either referred to orientation regarding placement, e.g. “stuck them together in a way that the holes fit” [26], or consciously disregarding constraints “then I put the picture out of my mind a little” [20]. The use of constraints to understand the spatial layout supports the assumption that constraints were used to restrict the possibilities of object placement and arrangement. If *bevels* were used strategically, they were most often used to distinguish between objects as reported in “right then I tried to separate between these beveled walls and uh balls” [16]. Furthermore, bevels were reported to foster insights, as in the following example “but the walls were not beveled and the roof would not have fit on top therefore it was logical that the walls with the balls on the columns needed to go down” [59]. The frequency analysis highlighted that *bevels* and *grooves* fostered function assignment significantly more frequently than *boreholes*. This finding suggests that *bevels* and *grooves* could be used according to their conventionalized function thus fostering insights that were related to background knowledge, whereas the function of *boreholes* needed to be defined for the specific context.

To summarize, the most unexpected finding was revealed by combining qualitative content-based analysis and frequency analysis of the three verbal reports. These analyses highlighted that assemblers did not mention constraints, i.e. object features, that are most salient but rather those that were evaluated as meaningful in the given task context. Moreover, by summarizing the findings on the function of constraints, two conceptual levels could be distinguished. At one level, constraints were interpreted, as in insights, function assignment, and strategic use. If constraints were interpreted, they influenced the problem solving process by restricting possible moves and suggesting the ones that were most likely to be successful. At the other level, constraints were only perceived and described. In this case, noticing the constraints did not contribute to the search process at that moment.

10.2.3 Influence of prior information on problem solving processes

The analyses revealed that prior information had significant effects on different aspects of the assembly process. On the performance level, prior information had an influence on assembly success and the distribution of process categories. On the linguistic level, an effect of prior information was observed on the use of pronominal forms, the kind of referential change throughout the assembly, and representation of explicit function assignment. Furthermore, prior information had an effect on the kind of constraints that were mentioned in think aloud protocols.

Regarding assembly success, all participants, theoretically, had the same chance to assemble the original dollhouse because they could have used the boreholes that marked the spatial layout on the boards and that corresponded to boreholes on the object parts. Additionally, some physical object features constrained possible actions and could have guided the assembly as outlined in the previous section. The analysis revealed that participants who saw the picture before the assembly were more successful in assembling the original dollhouse than participants who did not see the picture. This finding suggests that an external model provided better and more reliable clues about objects than the objects themselves.

However, the analysis of assembly success in terms of the goal information that was provided in the instruction highlighted that object features were also noticed and informative. The analysis revealed that participants who were told to assemble a two-story dollhouse were most successful. This finding suggests that knowing about the general nature of the object was sufficient to succeed in assembling a two-story house. Theoretically, these participants could rely on their general knowledge about a two-story house and the features of the given objects. The observation that a comparable number of participants who did not know anything about the goal object succeeded in assembling a two-story house strongly suggests that object features were salient enough to be recognized and that they provided meaningful guidance. Based on the observation that participants who saw the picture succeeded less frequently in reaching the formulated goal state, it can be assumed that the picture influenced participants' strategies. Apparently, these participants tried to match the given objects to the memorized picture and it seems as if they did not use object features as a guide in cases in which this matching was unsuccessful. Participants who did not have an external representation relied on the given objects to provide information regarding possible configurations. Therefore, they seem to pay more attention to object features.

This assumption is supported by a significant difference between assembly conditions regarding the frequency of the process category *comment on object features*. Participants, who needed to create a model of the goal object by themselves, verbalized comments on object features significantly more frequently than participants who saw the picture. Additionally, those assemblers referred to the *groove*, a physical property that was not immediately salient, significantly more frequently. If these observations are combined with the findings on assembly success, it indicates that participants who needed to construct the goal object from scratch, were more attentive to object features as they were looking for clues about the nature of a possible goal object. Nonetheless, a second interpretation needs to be considered; this difference may not reflect the frequency of noticing but rather the frequency of verbalizing what is noticed. Assuming that participants who know about the goal object, either verbally only or visually as well, can conceptualize object features faster because they understand their function shortly after noticing them, these thoughts were not verbalized because the act of perception was not in focus of attention. This interpretation is supported by the finding that assemblers who held a mental representation of the goal structure mentioned *bevels* less frequently than the other assemblers did. *Bevels* were salient object features but if assemblers remembered the picture well, they recognized that these objects belong to the second floor. The process of function assignment might have been so fast that it did not reach the necessary level of attention to be verbalized.

The possibility of missing information needs to be considered as one option in all analyses of think aloud protocols since there is no way of knowing which information is not verbalized. However, alternative interpretations can be supported by providing evidence from analyses of different aspects of the phenomenon, from analyses of the same aspect in different verbal reports, or by other performance measures. Therefore, findings on the distribution of problem solving processes will be considered here as well.

The frequency analysis of processes revealed that participants who were not provided with any prior information, verbalized significantly more *evaluations* and *actions* than participants in the other conditions. This finding suggests that these participants tested their hypotheses more frequently than participants who held a mental representation of the goal object earlier in the assembly process. This strengthens the interpretation that these participants were more attentive to features and changes resulting from their actions. The difficulties in assigning function to objects were mirrored in the significantly lower frequency of domain specific nominal references as compared to the other participants. It can be assumed that this difficulty of function assignment resulted in an enhanced focus of attention to object features. The focus on the overall assembly structure, rather than on individual parts, was also expressed in functional terms that were explicitly assigned to objects. Participants without prior information assigned structure-based terminology, such as 'Haus' (house) and 'Etage' (story) significantly more frequently than participants with prior information.

The reported findings from different perspectives of the analysis suggest that prior information influences how much attention participants pay to the individual objects, i.e. how closely they inspect them. Participants who are provided with an external representation of the goal structure tend to pay less attention to individual object features than participants who need to create a sensible goal structure from the presented objects by themselves. Moreover, the task to construct a goal structure is linguistically reflected in the frequency of referring to structure-based terminology as compared to part-based terminology when explicitly assigning function to individual objects.

10.2.4 Influence of communicative intention on problem solving processes

As expected, the analyses revealed significant effects of communicative intention on the linguistic level of object conceptualization and mention of constraints. Concerning object conceptualization as expressed in noun phrases, the analysis highlighted that instructors tended to refer to objects by noun phrases with domain unspecific nouns from first to fourth reference. This was markedly different from assemblers who tended to refer to objects by pronouns significantly more often in first and second reference. The instructors' tendency to refer to objects on the descriptive, non-functional level supports the assumption that the individual objects did not suggest a prototypical function. If objects had suggested a prototypical function, it can be assumed that instructors would have used it to facilitate identification. The fact that they maintained an unfamiliar status throughout the assembly strengthens this impression by suggesting that even after repeated introduction the object cannot be easily identified.

One possible strategy to compensate for a missing conventional referential term is the process of negotiation, i.e. offering a reference and wait for the addressee's confirmation. However, this is only possible in cases of direct interaction, as reported in Rieser (1995) and von Stutterheim et al. (1993), for example. If negotiation is not possible, the speaker may offer a conceptualization by explicitly assigning a reference and presume it as given thereafter. Contrary to prior expectations, the analysis revealed that explicit function was not significantly more frequent in instructions than in self-assembly. However, on the descriptive level, one further category of mapping phrases was identified in addition to the two strategies of *direct mapping* and *representational mapping* that were identified in think aloud protocols. Instructors were observed to *reframe* objects within one utterance, e.g. „die ähm Teile mit den langen Holzplatten dran also die hinteren Wände“ (“those uhm parts with the long wooden pieces well the walls in the back“) [B8]. In this example, the descriptive but unspecific reference ‘parts with the long wooden pieces’ is reframed to the specific reference ‘walls’ highlighting its function within in the goal structure. Since this strategy occurred exclusively in instructions in the elicited data set and similar examples were reported in two studies on assembly instructions (Rieser 1996; Daniel & Tversky 2012), it is assumed that it is a communicative strategy. Since only few cases of explicit mapping were observed in instructions, more data is needed to draw general conclusions about the implications of this specific kind of function assignment.

The analysis of the static view on referential form was supplemented by the analysis of maintenance and change of referential form throughout the instructions. This analysis of referential change highlighted that instructors *repeated* or *rephrased* references most frequently. The finding that references were rephrased could imply that instructors forgot the initial reference thus choosing another one that signals the same state of unfamiliarity. These results highlight that instructors who were familiar with the objects before the instruction considered their addressee's point of view when choosing a referential term.

Combined with the previously discussed findings, the results suggest that instructors adapted to the addressee's unfamiliar knowledge state and that they focused on completing the task rather than conceptualizing the individual parts therein. This might be an effect of the task that focused explicitly on instructing the addressee on the assembly while this person is performing the same actions. Thus, the manipulation of the current state to reach the goal state was in focus rather than learning the assembly for subsequent assemblies. This interpretation is supported by findings reported by Maes et al. (2004), who compared the frequency of referential overspecification in written instructions of participants who were told that the reader would perform the action while reading the instruction to those written for an addressee who needed to learn the procedure, i.e. here set an alarm clock. The results revealed the general trend that participants who were told to write an instruction for a reader who performed the procedure while reading the instruction included fewer detail in their referential phrases concerning shape, size, and location of objects than those in the other condition.

The analysis of constraints revealed that communicative intention also had an effect on their use. In general terms, the analysis of the nature of constraints mentioned in think aloud protocols and in instructions revealed that the same constraints were mentioned. In both verbal data sets, *boreholes* and *bevels* were the constraints that were frequently mentioned. However, there were differences with regard to the function of *boreholes*. Instructors used *boreholes* predominantly to coordinate the placement of objects or to describe the spatial layout. Assemblers working for themselves, in contrast, tended to verbalize the perception of these features and sometimes described insights that they gained by noticing them. This difference can be interpreted in two ways. On the one hand, it highlights that the same constraints may serve different functions. Although *boreholes* seem uninformative in the assembly scenario when no screws are provided, they are assigned a strategic function when there is a need to coordinate actions. On the other hand, it may as well be possible that instructors discovered the function of boreholes during their own silent assembly and used it during the instruction by verbalizing their own strategy. Combined with previous findings, it might be possible that the use of constraints specifically facilitates coordination in a context in which identification of and reference to objects is difficult because of a lack of shared terminology.

At the methodological level, the reported diversity of findings on problem solving processes involved in unaided assembly and their linguistic representation highlights the range of insights that can be gained by combining *protocol analysis* with a systematic linguistic analysis. The following section takes a critical view on the applicability of *Cognitive Discourse Analysis* and the contribution of this thesis to testing and refining the methodology. The thesis closes by outlining possible future directions.

10.3 Assessment of methodology

This section discusses the general advantages of *Cognitive Discourse Analysis* (CODA) highlighting contributions of this thesis to strengthen them. Moreover, the limitations that were faced in this thesis are discussed and some suggestions are made on how they can be addressed in future studies.

One characteristic of CODA is the application of discourse analytical frameworks to the specific discourse type of verbal reports that do not involve a specific addressee. This thesis, specifically, highlighted that well-established frameworks can be extended for a systematic investigation of phenomena in think aloud protocols. In the analysis of referential form and change, for example, frameworks established in communicative paradigms were adopted for the analysis of object references in think aloud protocols. This thesis contributed a theoretical proposal on the different levels of speaker – addressee interactivity and engagement as well as a systematic investigation of differences concerning object reference and functional use of constraints. The comparison between references in think aloud protocols and instructions contributed new empirical support for the basic assumption that think aloud protocols contain self-directed speech as opposed to addressee-oriented speech. The analysis of linguistic patterns associated with self-directed and addressee-directed speech highlighted discourse task specific differences as well as conceptual differences.

This latter contribution is also important with regard to one of the disadvantages of applying *Cognitive Discourse Analysis*. Researchers who apply this methodology need expertise and practice to distinguish between discourse task specific aspects and those aspects that are very likely to reflect mental processes. A comprehensive review of previous findings along with a detailed description of the investigated tasks and discourse types can facilitate the application of the methodology. Tenbrink (e.g. 2008; 2010) as well as Tenbrink and Gralla (2009) provide first overviews, which should be extended by the findings summarized in the theoretical introduction to CODA and the empirical findings obtained in this thesis. This future work needs to be emphasized because different researchers expressed interest in applying the methodology as has been experienced in discussions after a number of my talks and in workshops by Thora Tenbrink.

Cognitive Discourse Analysis extends the kind of information that can be gained by *protocol analysis* by highlighting linguistic patterns (Tenbrink 2010). It is acknowledged that verbal protocols capture information that reaches the speaker's attention (e.g. Ericsson & Simon 1993; van Someren et al. 1994). However, by studying the linguistic representation of this information and contrasting it with other possible linguistic representations, unconscious choices can be highlighted. If these findings are contrasted with findings of the same discourse type, e.g. retrospective reports, on different tasks, correlations between linguistic representations and mental processes can be identified. This thesis contributed new evidence for this postulation by providing insights on the correlation between linguistic features and mental processes across different discourse types on one task. The application to a problem solving task that was not investigated by CODA before highlights the general applicability of the methodology.

Tenbrink (e.g. 2008; 2010) repeatedly stresses that *Cognitive Discourse Analysis* can be applied to systematically analyze natural and unconstrained data in real-world problem solving tasks. The results reported in this thesis support this proposition because problem solving processes that were described in the literature could be identified highlighting their characteristic linguistic representation. Furthermore, as the qualitative analysis described the individual processes at a more fine-grained level, processes were highlighted that were not described before. This finding is interesting given that Tenbrink and Gralla (2009) suggested that the identification and description of differences in thought processes and their externalization in diverse discourse tasks could inspire theories on modeling cognitive processes. The findings reported in this thesis provide some suggestions regarding a model of human problem solving that is not as well-organized as assumed by traditional information processing theories and observed in problems that have a clearly defined inherent structure, such as chess. As this observation is also reported by Steffensen (2013), the analysis should be extended to other real-life problem solving tasks in future work; more specific suggestions will be made in the next section.

However, the application of CODA also suffers the disadvantage of high labor intensity that is specific to verbal protocol analysis. The analysis requires data transcription and if participants are free to say everything that comes to their mind, these reports can be quite long. Additionally, the coding procedure is time consuming in CODA because numerous aspects need to be coded to systematically investigate linguistic patterns and identify relevant aspects. This thesis contributes an in-depth analysis of object references, verb type and form, as well as on nature and function of constraints that could be used as a starting point for new studies. The reported analyses are based on natural, unconstrained data; therefore, it can be argued that they cover a wide range of possibilities and can thus be applied in the investigation of other discourse tasks as well.

To conclude, the application of CODA revealed new insights on the nature of problem solving processes involved in unaided object assembly and highlighted characteristic linguistic patterns for most of them. One of the main contributions of this thesis to the refinement and extension of this methodology is the systematic analysis of selected aspects of the linguistic structure of the elicited data along with the theoretical framework that is adopted in these analyses. There is still much work to be done to tackle the outlined limitations of the methodology; some suggestions for further work on them will be made in the following section.

10.4 Directions of future research

The inspiring discussions after my talks along with encouraging and constructive feedback on selected aspects of this thesis suggest that the presented approach is well received in the cognitive science and cognitive linguistics community and should be pursued in future work. The following two main directions of future research are suggested; i.e. focusing on the refinement of the methodology to address its limitations and focusing on the application of the proposed framework on different tasks.

The work on this thesis highlighted that future research should aim at reducing the labor that needs to be invested in preparing the data for analysis and selecting which aspects are informative with regard to the posed research questions. Concerning the preparation of the data, transcriptions are very time consuming. One of the findings in a web-based study on strategies employed in analogical problem solving (Gralla et al. 2012) was that planning protocols contained information that is typical of think aloud protocols, i.e. precise descriptions of the actions along with temporal markers. These linguistic features together with fast reaction times on solving the actual task suggested that participants actually solved the task when asked to provide a plan and entered the solution afterwards. If this effect could be replicated in another study, written planning protocols might be used instead of verbal protocols in some studies.

Regarding the selection of informative aspects, new directions are explored in an on-going collaboration with Sven Brüssow. Based on the findings obtained in this thesis, we assumed that the linguistic structure highlights the nature of mental processes in problem solving. We reasoned that it might be possible to identify processes based on a restricted number of words already if the linguistic structure of the target processes is well defined. Thus, we tested how much information can be maintained by the elicitation of n-grams of verbal reports. In first analyses, we selected n-grams that contained between three and five units and coded those according to the process categories that were described in this thesis. Generally, it showed that trigrams were more informative than bigrams as they provided important information, compare, for example, „macht keinen" (makes no) to „macht keinen Sinn" (makes no sense). In order to assess if this procedure is sufficient to highlight the task specific nature of protocols, we compared think aloud protocols collected on two distinct tasks, i.e. the self-assembly data and data on a scheduling task (for more detail see chapter 7). The qualitative analysis of the automated collocation protocol analysis reliably revealed the task specific problem structure. Additionally, it highlighted general problem solving processes, such as *evaluations* and *hypotheses*. So far, this work is in a preliminary state but encouraged by our findings and feedback on a presentation of this approach[286] that Sven Brüssow gave in Freiburg, we believe that this approach might be a reasonable extension of the work presented so far.

Concerning the application of the proposed framework to a different experimental setting, it would be interesting to test if the proposed technique for highlighting sequences of problem solving processes in think aloud protocols, works for different tasks as well. This is specifically interesting regarding the observation that problem solving as verbalized while working on real-world problems seems to be less structured than previously described for logic-based problems. The findings in this thesis suggest that this might be the case. Furthermore, the observation that hypotheses are very frequently used in the recorded data suggests that the notion of 'solution probing' (Cowley & Nash 2013; Steffensen 2013) might be one successful strategy in problem solving that has not been described before. Steffensen (2013) observed that the participants formulated numerous hypotheses regarding possible solutions. Those hypotheses were not evaluated right away but numerous were verbalized and only evaluated later in the process, finally leading the way to a satisfactory solution. Similarly, Dörner (2006) proposed that

[286] The presentation was entitled "Do measures of word association help facilitate the analysis of verbal think-aloud protocols?" and was presented in a colloquium at Freiburg University in November 2012.

successful problem solvers engaged in a dialogue posing questions and formulating statements and hypotheses. The finding that participants working on the assembly task silently took longer to complete the task than participants thinking aloud supports the idea that inner dialogue and hypothesis testing are an important part in problem solving.

The investigation of this phenomenon could be started by re-coding existing data sets applying the proposed framework before designing a new experiment or by running a new experiment right away. If existing data sets are re-coded, it is important that the think aloud protocols were elicited on a similar task, i.e. ideally an assembly task, and that a silent control group was tested. Since these requirements are very specific, it is more likely that a new study needs to be designed. The design of a future study should encourage hypothesis generation and testing. Based on the reported findings and those reported in the literature (e.g. Wedman et al. 1996; Steffensen 2013), it is assumed that unstructured tasks that involve the need to conceptualize objects in a specific context are specifically suited to encourage hypothesis generation. Furthermore, a future design needs to include a silent control condition to test whether there is an effect of verbalization and if so in which direction, i.e. facilitating or hindering performance.

Future application of the framework should also incorporate different performance measures for triangulation. Generally, the extension of the presented approach by recordings of eye movements and fixation points would be of great interest. Based on assumptions about the relation between fixation time and rates, and mental processes (e.g. Knoblich et al. 2005; Henderson et al. 2007; Land 2007), converging evidence could be established for the relation between verbalized thoughts and mental processes. More specifically, it would be a reasonable extension to the proposed analysis of use and function of constraints for two reasons. First, new technologies (e.g. Tobii Eyetracking[287] or SMI Eyetracking glasses[288]) allow participants to move almost unconstrained while working on a task. This new development would allow to record participants natural behavior and speech while engaged in solving a problem. Second, fixation points and durations are specifically interesting to evaluate if participants noticed a constraint and how they referred to it if at all. The combination of the linguistic analysis, as outlined in this thesis, and eye movement data could help to find more evidence for the suggestion that

[287] On Tobii Eyetracking see http://www.tobii.com/de/eye-tracking-research/germany/research/akademische-forschung/; 20.06.2013.

[288] On SMI Eyetracking glasses see http://www.eyetracking-glasses.com/products/eyetracking-glasses/technology/; 20.06.2013.

participants verbalize object features that are meaningful to them and thus reach attention whereas features that are noticed but not meaningful are not verbalized. Moreover, it might be possible to describe the relation between the duration of fixation on a feature and the likelihood to be mentioned in retrospective reports.

In sum, these possible future directions highlight that this thesis provides a solid base regarding the framework of investigating problem solving in unaided object assembly that should be extended in a number of ways in the work to come. Moreover, the discussion of the methodology highlighted how current limitations could be addressed in future studies.

References

Agrawala, M., Phan, D., Heiser, J., Haymaker, J., Klingner, J., Hanrahan, P., & Tversky, B. (2003). Designing effective step-by-step assembly instructions. *SIGGRAPH 03: ACM SIGGRAPH 2003 Papers*, 828-837.

Agre, P. & Chapman, D. (1987). Pengi: An implementation of a theory of activity. *Proceedings of the Sixth National Conference on Artificial Intelligence*, American Association for Artificial Intelligence (pp.268-272). Seattle: Kaufmann.

Ahlhum-Heath, M.E. & Di Vesta, F.J. (1986). The effect of conscious controlled verbalization of a cognitive strategy on transfer in problem solving. *Memory & Cognition, 14,* 281-285.

Ainsworth, S. & Loizou, A. (2003). The effects of self-explaining when learning with text or diagram. *Cognitive Science, 27*, 669-681.

Anderson, J.R. (2005). *Cognitive Psychology and its implications.* 6th *edition*. New York, NY: Worth Publishers.

Andonova, E., Tenbrink, T., & Coventry, K. (2010). Function and context affect spatial information packaging at multiple levels. *Psychonomic Bulletin & Review, 17*, 575-580.

Andonova, E., Tenbrink, T., & Coventry, K. (2008). Spatial description, Function and Context. In Love, B.C., McRae, K., & Sloutsky, V.M. (Eds.). *Proceedings of the 30th Annual Conference of the Cognitive Science Society.* (pp.199-204). Austin, TX: Cognitive Science Society.

Arnold, J.E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes, 23*, 495-527.

Auer, J.C.P. (1984). Referential problems in conversation. *Journal of Pragmatics,* 8, 627-648.

Bartl, C. & Dörner, D. (1998). Sprachlos beim Denken – Zum Einfluß von Sprache auf die Problemlöse- und Gedächtnisleistung bei der Bearbeitung eines nicht-sprachlichen Problems. *Sprache & Kognition, 17*, 224-238.

Bateman, J.A. (2008). Systemic Functional Linguistics and the Notion of Linguistic Structure: Unanswered Questions, New Possibilities. In Webster, J.J. (Ed.). *Meaning in Context. Implementing Intelligent Applications of Language Studies*. (pp.24-58). London, England: Continuum.

Bateman, J.A. (1989). Dynamic systemic-functional grammar: a new frontier. *Word, 40*, 263-287.

Bateman, J.A. (1988). Aspects of clause politeness in Japanese: An extended inquiry semantic treatment. *Proceedings of the 26th International Conference in Computational Linguistics*, 147-154.

Bäuerle, R. (1991). Verben der propositionalen Einstellung. In Stechow, A. von, & Wunderlich, D. (Hrsg.). *Semantik. Ein funktionales Handbuch der zeitgenösischen Forschung. Semantics*. (pp.709-722). Berlin: de Gruyter.

Bavelas, J.B., Coates, L., & Johnson, T. (2000). Listeners as Co-Narrators. *Journal of Personality and Social Psychology, 79*, 941-952.

Bégoin-Augereau, S. & Caron-Pargue, J. (2003). Linguistic criteria for demarcation and hierarchical organization of episodes in a problem solving task. *Proceedings of the Fifth Conference of the International Society for the Study of Argumentation*, 81-87. Amsterdam: Sic Sat.

Bégoin-Augereau, S. & Caron-Pargue, J. (2009). Linguistic markers of decision processes in a problem solving task. *Cognitive Systems Research*, 10, 102-123.

Bégoin-Augereau, S. & Caron-Pargue, J. (2010). Modified decision processes marked by linguistic forms in a problem solving task. *Cognitive Systems Research, 11*, 260-286.

Beller, S. & Bender, A. (2010). *Allgemeine Psycholgie – Denken und Sprache*. Göttingen, Germany: Hogrefe.

Betsch, T., Funke, J., & Plessner, H. (2010). *Allgemeine Psycholgie für Bachelor: Denken – Urteilen, Entscheiden, Problemlösen. Lesen, Hören, Lernen im Web*. Wiesbaden, Germany: C.H. Beck.

Beyer, R. & Gerlach, R. (2011). *Sprache und Denken*. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Bilalić, M., McLeod, P., & Gobet, F. (2008). Inflexibility of experts – Reality or myth? Quantifying the Einstellung effect in chess master. *Cognitive Psychology*, 56, 73-102.

Bise, V. (2008). *Problemlösen im Dialog mit sich selbst: dialogische Strukturen im inneren Sprechen beim Problemlösen. Eine explorative Studie nach der Methode des Lauten Denkens*. Marburg, Germany: Tectum-Verlag.

Blackman, H.S. (1988). The use of Think Aloud Verbal Protocols for the identification of Mental Models. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting October 1988*, 872-874.

Boroditsky, L. (no date). How Language Shapes Thought. (http://fora.tv/2010/10/26/Lera_Boroditsky_How_Language_Shapes_Thought; 25.05.2012).

Bosch, P. & Umbach, C. (2007). Reference Determination for Demonstrative Pronouns. In Bittner, D. & Gargarina, N. (Eds). *Intersentential Pronominal Reference in Child and Adult Language, ZAS Papers in Linguistics No 48*. (pp. 39-51). Zentrum für Allgemeine Sprachwissenschaft (ZAS).

Bransford, J.D. & Johnson, M.K. (1972). Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.

Bredel, U. & Töpler, C. (2007). Verb. In Hoffmann, L. (Ed.). *Deutsche Wortarten*. (pp.823-901). Berlin, Germany: de Gruyter.

Brinck, I. (2012). *The many ways of referential behaviour*. Talk at the Conference on Interactivity, Language and Communication, Odense 12-14 September 2012.

Brinck, I. (2003). The objects of attention: Causes and targets. *Behavioral and Brain Sciences, 26*, 287-288.

Brinck, I. (2001). Attention and the Evolution of Intentional Communication. *Pragmatics and Cognition, 9*, 259-277.

Brinkmann, H. (1971). *Die deutsche Sprache. Gestalt und Leistung. 2. überarbeitete & erweiterte Aufl.* Düsseldorf, Germany: Schwann.

Brockhaus Enzyklopädie. Eintrag "Denken" in Munzinger Online/Brockhaus – Enzyklopädie in 30 Bänden. 21. Auflage. Aktualisiert mit Artikeln aus der Brockhaus-Redaktion. (http://www. munzinger.de/document/12005024709; 27.04.2013).

Brown, R. (1958). How shall a thing be called? *Psychological Review, 65*, 14-21.

Brüssow, S., Gralla, L., Tenbrink, T., Holt, D., & Funke, J. (in preparation). Identifying traces of mental simulation in think aloud protocols: a combined approach.

Bühler, A. (1989). Semantik kognitiver Sätze und mentale Repräsentationen. In Falkenberg, G. (Ed.). *Wissen, Wahrnehmen, Glauben. Epistemtische Ausdrücke und propositionale Einstellungen*. (pp.9-25). Tübingen, Germany: Niemeyer.

Butulussi, E. (1991). *Studien zur Valenz kognitiver Verben im Deutschen und Neugriechischen*. Tübingen, Germany: Niemeyer.

Caffarel, A. (2010). Systemic functional grammar and the study of meaning. In Heine, B. & Narrog, H. (Eds.). *The Oxford Handbook of Linguistic Analysis*. (pp.797-825). New York, NY: Oxford University Press.

Caffarel, A., Martin, J.R., & Matthiessen, C.M.I.M. (Eds.) (2004). *Language Typology – A functional perspective*. Amsterdam, Netherlands: John Benjamins.

Carbonell, J.G. (1986). Derivational analogy: A theory of reconstructive problem solving and expertise acquisition. In Michalski, R.S., Carbonell, J.G., & Mitchell, T.M. (Eds.). *Machine learning: An artificial intelligence approach – Volume II*. (pp.371-393). Los Altos, CA: Kauffmann.

Carbonell, J.G. (1983). Learning by analogy. Formulating and generalizing plans from past experiences. In Michalski, R.S., Carbonell, J.G., & Mitchell, T.M. (Eds.). *Machine learning – An artificial intelligence approach*. (pp.173-161). Berlin, Germany: Springer.

Caron-Pargue, J. & Caron, J. (1991). Psychopragmatics vs. sociopragmatics: the function of pragmatic markers in think-aloud protocols. In Verscheuren, J. (Ed.). *Pragmatics at issue – Selected papers of the International Pragmatics Conference Antwerp, Aug 17-22 1987, Vol. 1*, 29-36. Amsterdam, Netherlands: John Benjamins.

Caron, J. (1996). Linguistic markers and cognitive operations. In Caron-Pargue, J., & Gillis, S. (Eds.). *Verbal production and problem solving. Antwerp Papers in Linguistics*. (pp. 11-28). Wilrijk, Belgium: Univ. Antwerp.

Caron, J. & Caron-Pargue, J. (1987). Towards a psycholinguistic approach to argumentative operators: The ‚thinking aloud' procedure. In van Emeren, F.H., Grootendorst, R., Blair, J.A., & Willard, C.A. (Eds.). *Argumentation: Perspectives and Approaches. Proceedings of the Conference on Argumentation 1986*. (pp.170-178). Dodrecht-HoHand, USA: Foris Publications.

Chi, M.T.H. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *The Journal of Learning Sciences, 6*, 271-315.

Chi, M.T.H. (1996). Constructing Self-Explanations and Scaffolded Explanations in Tutoring. *Applied Cognitive Psychology, 10*, 33-49.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science, 13*, 145-182.

Chrysikou, E.G. & Weisberg, R.W. (2005). Following the Wrong Footsteps: Fixation Effects of Pictorial Examples in a Design Problem-Solving Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1134-1148.

Clark, A. (2008). *Supersizing the mind. Embodiment, Action, and Cognitive Extension*. Oxford, UK: Oxford University Press.

Clark, H.H. (2005). Coordinating with each other in a material world. *Discourse Studies, 7*, 507-525.

Clark, H.H. (2004). Pragmatics of Language Performance. In Horn, L.R., & Ward, G. (Eds.). *Handbook of pragmatics*. (pp.365-382). Oxford, UK: Blackwell.

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

Clark, H.H. (2003). Pointing and Placing. In Kita, S. (Ed.). *Pointing. Where language, culture, and cognition meet*. (pp.243-268). Hillsdale ,NJ: Erlbaum.

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

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

Cokely, E.T. & Kelley, C.M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making, 4*, 20-33.

Cornish, F. (2001). 'Modal' that as determiner and pronoun: the primacy of the cognitive-interactive dimension. *English Language and Linguistics, 5*, 297-315.

Cowley, S. & Nash, L. (2013). Language, interactivity and solution probing: repetition without repetition. *Adaptive Behavior, 21*, 187-198.

Crutcher, R.J. (1994). Telling What We Know: The Use of Verbal Report Methodologies in Psychological Research. *Psychological Science, 5*, 241-244.

Culioli, A. (1995). *Cognition and Representation in Linguistic Theory*. Amsterdam, Netherlands: Benjamins.

Daniel, M.-P. & Tversky, B. (2012). How to put things together. *Cognitive Processing, 13*, 303-319.

Deffner, G. (1989). Interaktion zwischen Lautem Denken, Bearbeitungsstrategien und Aufgabenmerkmalen? Eine experimentelle Prüfung des Modells von Ericsson und Simon. *Sprache & Kognition, 8*, 98-111.

DeSouza, M.L., DaSilveira, A., & Gomes, W.B. (2008). Verbalized Inner Speech and the Expressiveness of Self-Consciousness. *Qualitative Research in Psychology, 5*, 154-170.

Deutsches Wortschatzportal (Uni Leipzig: http://wortschatz.uni-leipzig.de/; 24.10.2012).

Dickson, J., McLennan, J., & Omodei, M.M. (2000). Effect of Concurrent Verbalization on a time-critical, Dynamic Decision-Making Task. *The Journal of General Psychology, 127*, 217-228.

Diewald, G. (1996). *Die Modalverben im Deutschen. Grammatikalisierung und Polyfunktionalität*. Tübingen, Germany: Niemeyer.

Dijk, T.A. van (2006). *Discourse and Context. A sociocognitive Approach.* Cambridge, UK: Cambridge University Press.

Dijk, T.A. van (2000). Cognitive Discourse Analysis. An introduction. (http://www.discour-ses.org/UnpublishedArticles/cogn-dis-anal.htm; 21.04.13).

Dijk, T.A. van & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, USA: Academic Press.

Dixon, P., Zimmerman, C., & Nearey, S. (1997). Prior experience and complex procedures. *Memory & Cognition, 25*, 381-394.

Dixon, P. (1987a). Actions and procedural directions. In Tomlin, R.S. (Ed.). *Coherence and grounding in discourse*. (pp.69-89). Amsterdam, Netherlands: John Benjamins.

Dixon, P. (1987b). The Structure of Mental Plans for Following Directions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 18-26.

Dixon, P. (1982). Plans and Written Directions for Complex Tasks. *Journal of Verbal Learning and Verbal Behavior, 21*, 70-84.

Dominowski, R.L. (1998). Verbalization and Problem Solving. In Hacker, D.J., Dunlosky, J., & Graesser, A.C. (Eds.). *Metacognition in educational theory and practice*. (pp.25-45). Mahwah, NJ: Erlbaum.

Dörner, D. (2006). Sprache und Denken. In Funke, J. (Ed.). *Enzyklopädie der Psychologie. Denken und Problemlösen*. (pp.619-646). Göttingen, Germany: Hogrefe.

Dörner, D. (1987). *Problemlösen als Informationsverarbeitung. 3. Aufl.* Stuttgart, Germany: Kohlhammer.

Dörner, D., Kreuzig, H.W., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern, Switzerland: Verlag Hans Huber.

Duncker, Karl (1974). *Zur Psychologie des produktiven Denkens – dritter Neudruck*. Berlin, Germany: Springer.

Eichinger, L.M. (1989). *Raum und Zeit im Verbwortschatz des Deutschen. Eine velanzgrammatische Studie*. Tübingen, Germany: Niemeyer.

Eisenberg, P. (2006). *Der Satz – Grundriss der deutschen Grammatik. 3.Auflage*. Stuttgart, Germany: Metzler.

Engel, U. (2002). *Kurze Grammatik der deutschen Sprache*. München, Germany: Iudicium.

Engel, U. (1996). *Deutsche Grammatik. 3.korrigierte Auflage*. Heidelberg, Germany: Groos.

Epstein, R. (2011). The distal demonstrative as discourse marker in *Beowulf*. *English Language and Linguistics, 15(1)*, 113-135.

Erben, J. (1972). *Deutsche Grammatik. Ein Abriss*. Müchen, Germany: Max Huber.

Ericsson, K.A. & Fox, M.C. (2011). Thinking Aloud Is Not a Form of Introspection but a Qualitatively Different Methodology: Reply to Schooler (2011). *Psychological Bulletin, 137*, 351-354.

Ericsson, K.A. (2006). Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts' Performance in Representative Tasks. In Ericsson, K.A., Charness, N., Feltovich, P.J., & Hoffmann, R.R. (Eds.). *The Cambridge Handbook of Expertise and Expert Performance*. (pp. 223-241). Cambridge, UK: Cambridge University Press.

Ericsson, K.A. (1999). Protocol Analysis. In Betchel, W., & Graham, G. (Eds.). *A Companion to Cognitive Science*. (pp. 425-432). Oxford, UK: Blackwell.

Ericsson, K.A., Delaney, P.F., Weaver, G., & Mahadevan, R. (2004). Uncovering the structure of a memorist's superior "basic" memory capacity. *Cognitive Psychology, 49*, 191-237.

Ericsson, K.A. & Kintsch, W. (1995). Long-Term Working Memory. *Psychological Review, 102*, 211-245.

Ericsson, K.A. & Simon, H.A. (1993). *Protocol Analysis. Verbal Reports as Data. Revised version*. Cambridge, MA: MIT Press.

Ericsson, K.A. & Cructher, R.J. (1991). Introspection and verbal reports on cognitive processes – Two approaches to the study of thinking: a response to Howe. *New Ideas in Psychology, 9*, 57-71.

Ericsson, K.A. & Simon, H.A. (1984). *Protocol Analysis. Verbal Reports as Data*. Cambridge, MA: MIT Press.

Ericsson, K.A. & Simon, H.A. (1980). Verbal Reports as Data. *Psychological Review, 87*, 215-251.

Ertel, S. & Bloemer, W-D. (1975). Affirmation and Negation as Constructive Action. *Psychological Research, 37*, 335-342.

Ertel, S. (1972). Erkenntnis und Dogmatismus. *Psychologische Rundschau, 23*, 241-269.

Evans, V. & Green, M. (2011). *Cognitive Linguistics. An Introduction*. Edinburgh, UK: Edinburgh University Press.

Farrington-Flint, L. & Wood, C. (2007). The Role of Lexical Analogies in Beginning Reading: Insights From Children's Self-Reports. *Journal of Experimental Psychology, 99*, 326-338.

Fauconnier, G. (1999). Methods and generalizations. In Janssen, T., & Redeker, G. (Eds.). *Cognitive Linguistics: Foundations, Scope, and Methodology*. (pp. 95-127). Berlin, Germany: de Gruyter.

Fawcett, R. (2005). *Invitation to Systemic Functional Linguistics: the Cardiff Grammar as an extension and simplification of Halliday's Systemic Functional Grammar*. 2nd *edition*. (http://www.cf.ac.uk/encap/fontaine/cardiffgrammar/cardiffgrammar.htm; 13.11.2012).

Fox, M.C., Ericsson, K.A., & Best, R. (2011). Do Procedures for Verbal Reporting of Thinking Have to Be Reactive? A Meta-Analysis and Recommendations for Best Reporting Methods. *Psychological Bulletin, 137*, 316-344.

Frankenstein, J., Büchner, S.J., Tenbrink, T., & Hölscher, C. (2010). Influence of geometry and objects on local route choices for wayfinding. In Hölscher, C., Shipley, T.F., Belardinelli, M.O., Bateman, J., & Newcombe, N.S. (Eds.) (2010). *Spatial Cognition VII – International Conference, Spatial Cognition 2010, Mt. Hood/Portland, OR, USA, August 15-19, 2010: Proceedings.* (pp.41-53). Berlin, Germany: Springer.

Franzen, U. & Merz, F. (1976). Einfluß des Verbalisierens auf die Leistung bei Intelligenzprüfungen: Neue Untersuchungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 8*, 117-134.

Fukumura, K. & van Gompel, R.P.G. (2012). Producing Pronouns and Definite Noun Phrases: Do Speakers Use the Adressee's Discourse Model? *Cognitive Science, 36*, 1289-1311.

Funke, J. & Spering, M. (2006). Methoden der Denk- und Problemlöseforschung. In Funke, J. (Ed.). *Enzyklopädie der Psychologie. Denken und Problemlösen*. (pp.647-744). Göttingen, Germany: Hogrefe.

Funke, J. (2003). *Problemlösendes Denken*. Stuttgart, Germany: Kohlhammer.

Ganier, F. (2004). Factors Affecting the processing of Procedural Instructions: Implications for Document Design. *IEEE Transactions on Professional Communication, 47*, 15-26.

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

Garrod, S.C. & Sanford, A.J. (1982). The mental representation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics, 1*, 21-41.

Gentner, D., Holyoak, K.J., & Kokinov, B.N. (2001). Introduction. In Gentner, D., Holyoak, K.J., & Kokinov, B.N. (Eds.). *The Analogical Mind – Perspectives from Cognitive Science*. Cambridge, MA: MIT Press.

Gentner, D. & Stevens, A.L. (1983). *Mental models*. Hilsdale, NJ: Erlbaum.

Gibbs, R.W. & Colston, H.L. (2006). Figurative language. In Traxler, W., & Gernsbacher, M. (Eds.). *Handbook of Psycholinguistics. 2nd edition.* (pp. 835-862) .Oxford, UK: Elsevier.

Gick, M.L. & Holyoak, K.J. (1980). Analogical Problem Solving. *Cognitive Psychology, 12*, 306-355.

Gilhooly, K.J., Fioratou, E., & Henretty, N. (2010). Verbalization and problem solving: Insight and spatial factors. *British Journal of Psychology, 101*, 81-93.

Gilhooly, K.J., Fioratou, E., Anthony, S.H., & Wynn, V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology, 98*, 611-625.

Glenberg, A.M. & Langston, W.E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory & Language, 31*, 129-151.

Glinz, H. (1973). *Die innere Form des Deutschen. Eine neue deutsche Grammatik. 6. durchgesehene Aufl.* Bern, Switzerland: Francke.

Goffman, E. (1981). *Forms of Talk*. Oxford, UK: Blackwell.

Gog, T. van, Paas, F., van Merriënboer, J.J.G., & Witte, P. (2005). Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology: Applied, 11*, 237-244.

Gralla, L. & Tenbrink, T. (2012). Verbalization of problem solving processes in unaided object assembly. Poster presentation at the 11th International Conference on Cognitive Modeling, April 13-15 2012, Berlin, Germany.

Gralla, L., Tenbrink, T., Siebers, M., & Schmid, U. (2012). Analogical Problem Solving: Insights from Verbal Reports. In Miyake, N., Peebles, D., & Cooper, R. (eds.). *Proceedings of the 33rd Annual Conference of the Cognitive Science Sociecty, Sapporo, Japan, 1.-4.8.2012.* (pp.396-401). Mahwah, NJ : Lawrence Erlbaum.

Green, A.J.F. (1998). *Using verbal protocols in language testing research: A handbook*. Cambridge, UK: Cambridge University Press.

Greeno, J.G. (1978). Natures of Problem-Solving Abilities. In Estes, W.K. (ed.). *Handbook of Learning and Cognitive Processes. Volume 5: Human Information Processing*. (pp.239-270). Hillsdale, NJ: Lawrence Erlbaum.

Greeno, J.G. (1974). Hobbits and Orcs: Acquisition of a sequential concept. *Cognitive Psychology, 6*, 270-292.

Groot, A.D. de (1969). *Methodology – Foundations in inference and research in behavioral sciences*. The Hague, Netherlands: Mouton.

Groot, A.D. de (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.

Gugerty, L. & Rodes, W. (2007). A Cognitive Model of Strategies for Cardinal Direction Judgment. *Spatial Cognition and Computation, 7*, 179-212.

Gundel, J.K., Bassene, M., Gordon, B., Humnick, L., & Khalfaoui (2010). Testing predictions of the Giveness Hierarchy framework: A crosslinguistic investigation. *Journal of Pragmatics, 42*, 1770-1785.

Gundel, J.K., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the form of referring expressions in discourse. *Language, 69*, 274-307.

Habel, C. (1988). Prozedurale Aspekte der Wegplanung und Wegbeschreibung. In Schnelle, H. & Rickheit, G. (eds.). *Sprache in Mensch und Computer: kognitive and neuronale Sprachverarbeitung*. (pp.107-133). Opladen, Germany: Westdeutscher Verlag.

Halliday, M.A.K. (2009a). Methods – techniques – problems. In Halliday, M.A.K. & Webster, J.J. (Eds.). *Continuum Companion to Systemic Functional Linguistics*. (pp. 59-86). London, UK: Continuum.

Halliday, M.A.K. (2009b). *The Essential Halliday. edited by Webster, J.J.* London, UK: Continuum.

Halliday, M.A.K. (2004). *An introduction to functional grammar. 3rd revised version by Matthiessen, C.M.I.M.* London, UK: Hodder.

Halliday, M.A.K. (2002). Modes of meaning and modes of expression: types of grammatical structure and their determination by different semantic functions (1979). In Webster, J. (Ed.). *On Grammar. Volume 1 in the collected works of M.A.K. Halliday*. (pp.169-218). London, UK: Continuum.

Halliday, M.A.K. (1985). *An Introduction to Functional Grammar*. London, UK: Edward Arnold.

Halliday, M.A.K. (1970). Functional diversity in language as seen from a consideration of modality and mood in English. *Foundations of Language, 6*, 322-361.

Halliday, M.A.K. and Matthiessen, C.M.I.M. (2002). *Construing experience through meaning. A language-based approach to cognition*. London, UK: Continuum.

Halmari, H. (1994). On Accessibility and Coreference. *Nordic Journal of Linguistics, 17*, 35-59.

Hayes-Roth, B. (1980). Flexibility in executing strategies. N-1170-ONR. The Official Naval Research. Rand, Santa Monica, California.

Hayes-Roth, B. & Hayes-Roth, F. (1979). A Cognitive Model of Planning. *Cognitive Science, 3*, 275-310.

Hayes, A.F. & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures, 1*, 77-89.

Heiser, J., Phan, D., Agrawala, M., Tversky, B., & Hanrahan, P. (2004). Identification and validation of cognitive design principles for automated generation of assembly instructions. *Proceedings of the working conference on advanced visual interface*. May 25-28, 2004, Gallipoli (Lecce), Italy.

Heiser, J., Tversky, B., Agrawala, M., & Hanrahan, P. (2003). Cognitive Design Principles for Visualization: Revealing and Instantiating. In Alterman, R. & Kirsh, D. (Eds.). *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. (pp.545-550). Boston, USA: Cognitive Science Society.

Helander, M.G. & Willén, B.A. (2003). Design of Human Assembly (DHA). In Karwowski, W. & Marras, W.S. (Eds.). *Occupational Ergonomics: Design and Management of Work Systems*. (pp.26.1-26.17). Boca Raton, USA: CRC Press.

Hellbig, H. & Buscha, J. (2005). *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Berlin, Germany: Langenscheidt.

Heller, D., Gorman, K.S., & Tanenhaus, M.K. (2012). To Name or to Describe: Shared Knowledge Affects Referential Form. *Topics in Cognitive Science, 4*, 290-305.

Henderson, J.M., Brockmole, J.R., Castelhano, M.S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In van Gompel, R.P.G., Fischer, M.H., Murray, W.S., & Hill, R. L. (Eds.). *Eye Movements: A window on mind and brain.* (pp. 537-562). Amsterdam, Netherlands: Elsevier.

Hentschel, E. & Weydt, H. (2003). *Handbuch der deutschen Grammatik. 3. Auflage.* Berlin, Germany: Walter de Gruyter.

Hermann, T. & Deutsch, W. (1976). *Psychologie der Objektbennung*. Bern, Switzerland: Verlag Hans Huber.

Hesse, F.W. (1982). Effekte des semantischen Kontexts auf die Lösung komplexer Probleme. *Zeitschrift für Experimentelle und Angewandte Psychologie, 29*, 62-91.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.R. (1996). *Induction. Processes of Inference, Learning, and Discovery*. Cambridge, MA: MIT Press.

Hölscher, C., Tenbrink, T., & Wiener, J. (2011). Would you follow your own route description? Cognitive strategies in urban route planning. *Cognition, 121*, 228-247.

Hölscher, C., Büchner, S.J., Meilinger, T., & Strube, G. (2009). Adaptivity of wayfinding strategies in a multi-building ensemble: The effects of spatial structure, task requirements, and metric information. *Journal of Environmental Psychology, 29*, 208-219.

Horton, W.S. (2009). Memory and other limits on audience design in reference production. Paper presented that the Production of Referring Expressions Conference: Bridging the Gap between Computational and Empirical Approaches to References (PRE-CogSci), Amsterdam, 20 July 2009.

Horton, W.S. & Gerrig, R.J. (2005a). The impact of memory demands on audience design during language production. *Cognition, 96*, 127-142.

Horton, W.S. & Gerrig, R.J. (2005b). Conversational Common ground and Memory Processes in Language Production. *Discourse Processes, 40*, 1-35.

Horton, W.S. & Gerrig, R.J. (2002). Speakers' experience and audience design: knowing *when* and knowing *how* to adjust utterances to the addressee. *Journal of Memory and Language, 47*, 589-606.

Hussy, W. (1993). *Denken und Problemlösen*. Stuttgart, Germany: Kohlhammer.

Hutchins, E. (2011). Enculturating the Supersized Mind. *Philosophical Studies, 152*, 437-446.

Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of Pragmatics, 37*, 1555-1577.

Hutchins, E. (1999). Cognitive Artifacts. In Wilson, R.A., & Keil, F.C. (Eds.). *The MIT Encyclopedia of the Cognitive Sciences*. (pp.126-128). Cambridge, MA: MIT Press.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

IfcBuilding Element (http://www.iai-tech.org/ifc/IFC2x4/alpha/html/index.htm; 8.12.2009).

Johnson-Laird, P.N. (2004). The history of mental models. In Manktelow, K. & Chung, M.C. (Eds.). *Psychology of Reasoning: Theoretical and Historical Perspectives*. (pp.179-212). NY: Psychology Press.

Johnson-Laird, P.N. (1998). Mental Models. In Posner, M.I. (Ed.). *Foundations in Cognitive Science*. (pp.469-499). Cambridge, MA: MIT Press.

Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge, MA: MIT Press.

Johnson, D.M. (1955). *The psychology of thought and judgment*. New York, NY: Harper.

Kaplan, C.A. & Simon, H.A. (1990). In Search of Insight. *Cognitive Psychology, 22*, 374-419.

Kibrik, A.A. (2011). *Reference in Discourse*. Oxford, UK: Oxford University Press.

Kibrik, A.A. (2001). Cognitive Discourse Analysis: Some results. In Enikő, N. (Ed.). *Cognition in language use: Selected papers from the 7th International Pragmatics Conference. Vol. 1.* (pp.164-180). Antwerp, Belgium: International Pragmatics Association.

Kiefer, M., Baratelli, S., & Mangold-Allwinn, R. (1993). *Kognition und Kommunikation: Ein integrativer Ansatz zur multiplen Determination der lexiaklischen Spezifität der Objektklassenbezeichnung*. Universität Mannheim: Lehrstuhl Psychology III.

Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review, 95*, 163-182.

Kintsch, W. & Greeno, J.G. (1985). Understanding and Solving Word Arithmetic Problems. *Psychological Review, 92*, 109-129.

Kirsh, D. (2010). Thinking with external representations. *AI & Society, 25*,441-454.

Kirsh, D. (2009). Problem Solving and Situated Cognition. In Robbinson, P. & Ayede, M. (Eds.). *The Cambridge Handbook of Situated Cognition*. (pp. 264-306). Cambridge, UK: Cambridge University Press.

Kirsh, D. (2008). Distributed Cognition: A methodological note. In Dror, I.E. & Harnad, S. (Eds.). *Cognition Distributed. How cognitive technology extends our minds*. (pp.57-69). Amsterdam, Netherlands: John Benjamins.

Kirsh, D. (1995). The intelligent use of space. *Artificial intelligence, 73*, 31-68.

Kitchener, K.S. (1983). Cognition, Metacognition, and Epistemic Cognition – A Three-Level Model of Cognitive Processing. *Human Development, 26*, 222-232.

Klippel, A., Tenbrink, T., & Montello, D.R. (2013). The role of structure and function in the conceptualization of direction. In Dimitrova-Vulchanova, M., & van der Zee, E. (Eds.). *Motion Encoding in Spatial Language*. (pp.102-119). Oxford, UK: Oxford University Press.

Klix, F. (1976). Über Grundstrukturen und Funktionsprinzipien kognitiver Prozesse. (Aspekte eines Zugangs zu Analyse und Synthese geistiger Leistungen). In Klix, F. (Ed.). *Psychologische Beiträge zur Analyse kognitiver Prozesse*. (pp.9-56). Berlin, Germany: Verlag der Wissenschaften.

Knauff, M. & Wolf, A.G. (2010). Complex cognition: the science of human reasoning, problem-solving, and decision making. (Editorial). *Cognitive Processing, 11*, 99-102.

Knepper, T.A., Layton, T., Romanishin, J., & Rus, D. (2013). IkeaBot: An Autonomous Multi-Robot Coordinated Furniture Assembly System. *Proceedings of the IEEE International Conference on Robotics and Atomation (ICRA). Karlsruhe, Germany, May 2013*.

Knoblich, G., Öllinger, M., & Spivey, M.J. (2005). Tracking the eyes to obtain insights into insight problem solving. In Underwood, G. (Ed.). *Cognitive Processes in Eye Guidance*. (pp.355-375). Oxford, UK: Oxford University Press.

Knoblich, G., Ohlsson, S., & Raney, G.E. (2001). An eye movement study of insight problem solving. *Memory & Cognition, 29*, 1000-1009.

Knoblich, G. & Ohlsson, S. (1999). Constraint and Chunk Decomposition in Insight Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*,1534-1555.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint Relaxation and Chunk Decomposition in Insight Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1534-1555.

Knoblich, G. & Rhenius, D. (1995). Zur Reaktivität Lauten Denkens beim komplexen Problemlösen. *Zeitschrift für experimentelle Psychologie, 42*, 419-454.

Krahmer, E. & Ummelen, N. (2004). Thinking about thinking aloud: a comparison of two verbal protocols for usability testing. *IEEE Transactions on Professional Communication, 47*, 105-117.

Krauss, R.M. & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Psychology and Social Psychology*, 4(3), 343-346.

Krippendorff, K. (2004). *Content Analysis. An Introduction to Its Methodology. 2nd edition.* Thousand Oaks, USA: SAGE Publishing.

Kucan, L. & Beck, I.L. (1997). Thinking Aloud and Reading Comprehension Research: Inquiry, Instruction, and Social Interaction. *Review of Educational Research, 67*, 271-299.

Labov, W. (1973). The boundaries of words and their meaning. In Bailey, C.-J.N. & Shuy, R.W. (Eds.). *New ways of Analyzing Variation in English*. (pp. 340-373). Washington, USA: Georgtown University Press.

Land, M.F. (2007). Fixation Strategies During Active Behaviour: A Brief History. In van Gompel, R.P.G., Fischer, M.H., Murray, W.S., & Hill, R. L. (Eds.). *Eye Movements: A window on mind and brain.* (pp. 75-95). Amsterdam, Netherlands: Elsevier.

Landis, J.R. & Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*, 159-174.

Langacker, R.W. (1999). Assessing the cognitive linguistic enterprise. In Janssen, T. & Redeker, G. (Eds.). *Cognitive Linguistics: Foundations, Scope, and Methodology*. (pp.13-59). Berlin, Germany: de Gruyter.

Latzel, S. (1978). *Die Verben "wissen", "kennen" und "können". Eine Bedeutungs- und Gebrauchsbescheibung mit Übungen für das Fach 'Deutsch als Fremdsprache'*. München, Germany: Druckerei Kemmler.

Laucht, M. (1982). *Kontexteffekte bei der Objektbennung*. Frankfurt a.M, Germany.: Peter Lang.

Lee, D. (2001). *Cognitive Linguistics. An introduction*. Oxford, UK: Oxford University Press.

Lenzen, M. (2002). *Natürlich und künstliche Intelligenz. Einführung in die Kognitions-wissenschaft*. Frankfurt a.M., Germany: Campus Verlag.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.

Levelt, W.J.M. (1981). The speaker's linearization problem. *Philosophical Transaction of the Royal Society London, 295*, 305-315.

Lombard, M., Snyder-Duch, J., & Bracken, C.C. (2002). Content Analysis in Mass Commun-ication. Assessment and Reporting of Intercoder Reliability. *Human Communication Research, 28*, 587-604.

Lovelace, K.L., Hegarty, M., & Montello, D.R. (1999). Elements of good route directions in familiar and unfamiliar environments. In Freska, C. & Mark, D.M. (Eds.). *Proceeding COSIT '99 Proceedings of the International Conference on Spatial information theory: Cognitive and Computational Foundations of Geographic Information Science*. (pp.65-82). Berlin, Germany: Springer.

Maclay, H. & Osgood, C.E. (1959). Hesitation Phenomena in Spontaneous English Speech. *Word. Journal of the Linguistic Circle of New York, 15*, 19-44.

Maes, A. (1999). What can production data teach us about the effectiveness of instructive text design variables? *Document Design, 1*, 170-182.

Maes, A., Arts, A., & Noordman, L. (2004). Reference Management in Instructive Discourse. *Discourse Processes, 37*, 117-144.

Maes, A. & Lenting, H. (1999). How to put the instructive space into words? *IEEE Transactions on Professional Communication, 42*, 100-113.

Malt, B., Sloman, S.A., & Gennari, S. (2003). Speaking versus Thinking about Objects and Actions. In Gentner, D. & Goldin-Meadow, S. (Eds.). *Languages in Mind – Advances in the Study of Language and Thought*. (pp.81-111). Cambridge, MA: MIT Press.

Malt, B., Sloman, S.A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus Naming: Similarity and the Linguistic Categorization of Artifacts. *Journal of Memory and Language, 40*, 230-262.

Mangold-Allwinn, R., Baratelli, S., Kiefer, M., & Koelbing, H.G. (1995). *Wörter für Dinge. Von flexiblen Konzepten zu variablen Benennungen*. Opladen, Germany: Westdeutscher Verlag.

Mangold-Allwinn, R., Baratelli, S., Kohlmann, U., & Koelbing, H.G. (1992). Objektbenennungen im Diskurs: Eine interdisziplinäre Untersuchung. *Kognitionswissenschaft, 3*, 1-11.

Mangold-Allwinn, R., Stutterheim, C. von, Baratelli, S., Kiefer, M., & Koelbing, H.G. (1995). *Wörter für Dinge. Von flexiblen Konzeoten zu variable Benennungen*. Opladen, Germany: Westdeutscher Verlag.

Mani, K. & Johnson-Laird, N.P. (1982). The mental representation of spatial descriptions. *Memory & Cognition, 10*, 181-187.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding Instructions. *Journal of Educational Psychology, 88(1)*, 49-63.

Martin, J.R. & Rose, D. (2003). *Working with Discourse. Meaning beyond the clause*. London, UK: Continuum.

Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two- and Four-Year-Olds Learn to Adapt Referring Expressions to Context: Effects of Distracters and Feedback on Referential Communication. *Topics in Cognitive Science, 4*, 184-201.

Matthiessen, C.M.I.M. (2009). Ideas and new directions. In Halliday, M.A.K. & Webster, J.J. (eds.). *Continuum Companion to Systemic Functional Linguistics*. (pp. 12-58). London, UK: Continuum.

McCabe, A. (2011). *An Introduction to Linguistics and Language Studies*. London, UK: Equinox.

McGeorge, P. & Burton, A.M. (1989). The effects of concurrent verbalization on performance in a dynamic systems task. *British Journal of Psychology, 80*, 455-465.

Merz, F. (1969). Der Einfluß des Verbalisierens auf die Leistung bei Intelligenzaufgaben. *Zeitschrift für Experimentelle Psychologie & Angewandte Psychologie, 16*, 114-137.

Meyer-Fujara, J. & Rieser, H. (2005). *Grammar-Based Metonomy Resolution. SFB 360 Report 2005/01*. Bielefeld, Germany: Universität.

Meyerhoff, M. (2009). Replication, transfer, and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change, 21*, 297-317.

Mietzel, G. (2006). *Wege in die Psychologie – völlig überarbeitete Neuausgabe*. Stuttgart, Germany: Klett-Cotta.

Miller, G.A. (1996). Contextuality. In Oakhill, J. & Garnham, A. (Eds.). *Mental models in cognitive Science – essays in honour of Phil Johnson-Laird*. (pp.1-18). Sussex, UK: Psychology Press.

Miller, G.A., Galanter, E., & Pribram, K.H. (1970). *Plans and the structure of behavior*. Aylesbury, UK: Hazell Watson & Viney Ltd.

Moratz, R. & Tenbrink, T. (2006). Spatial Reference in Linguistic Human-Robot Interaction: Iterative, Empirically Supported Development of a Model of Projective Relations. *Spatial Cognition, 6*, 63-106.

Neuendorf, K.A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications.

Neuman, Y. & Schwarz, B. (1998). Is self-explanation while solving problems helpful? The case of analogical problem-solving. *British Journal of Educational Psychology, 68*, 15-24.

Newell, A., Shaw, J.C., & Simon, H.A. (1960). Report on a general-problem solving program. In *Information processing: proceedings of the International Conference on Information Processing, Unesco, Paris 15-20 June 1959*. (pp.256-264). Paris, France: Unesco.

Newell, A., Shaw, J.C., & Simon, H.A. (1958). Elements of a theory of human problem solving. *Psychological Review, 65*, 151-166.

Newell, A. & Simon, H.A. (1972). *Human problem solving*. Engelwood Cliffs, NJ: Prentice-Hall.

Nisbett, R.E. & Wilson, T.D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review, 8*, 231-259.

Norman, D.A. (2002). *The design of everyday things*. New York, USA: Basic Books.

Norvick, L.R. & Morse, D.L. (2000). Folding a fish, making a mushroom: The role of diagrams in executing assembly procedures. *Memory & Cognition, 28*, 1242-1256.

O'Donnel, M. & Bateman, J. (2005). SFL in Computational Contexts: a Contemporary History. In Webster, J., Hasan, R., & Matthiessen, C. (Eds.). *Continuing Discourse on Language: A Functional Perspective*. (pp.343-382). London, UK: Equinox.

Oberbauer, K. (2012). Modelling Working Memory. In Rußwinkel, N., Drewitz, U., & van Rijn, H. (Eds.). *Proceedings of ICCM 2012, 11th International Conference on Cognitive Modeling.* (p.xii). Berlin, Germany: Universitätsverlag der TU Berlin.

Ochs, E. (1979). Transcription as Theory. In Ochs, E. & Schieffelin, B.B. (Eds.). *Developmental Pragmatics*. (pp.43-72). New York, USA: Academic Press.

Oxford English Dictionary, online version (http://oald8.oxfordlearnersdictionaries.com/dictionary; 01.06.2013).

Palmer, F.R. (1974). *The English Verb. (Reprint of the 1965 edition).* London, UK: Longman.

Palmer, S.E. (1977). Hierarchical Structure in Perceptual Representation. *Cognitive Psychology, 9*, 441-474.

Passini, R. (1981). Wayfinding: A conceptual framework. *Urban Ecology, 5*, 17-31.

Pecher, D. & Zwaan, R.A. (2005). Introduction to Grounding Cognition. The Role of Perception and Action in Memory, Language, and Thinking. In Pecher, D. & Zwaan, R.A. (Eds.). *Grounding Cognition. The Role of Perception and Action in Memory, Language, and Thinking.* (pp.1-7). Cambridge, UK: Cambridge University Press.

Petersen, U.H. (unpublished). *Bedeutung, Funktion und Form. Eine systemisch-funktionale Beschreibung des Deutschen. Unveröffentlichter Entwurf (Stand 2008)*. Kapitel 3 (die ideationelle Metafunktion), Kapitel 4 (Die logische Metafunktion: Konnexion und Projektion), Kapitel 5 (Die experientielle Metafunktion: Der Satz als Darstellung).

Petrova, S. & Solf, M. (2010). Pronominale Wiederaufnahme im ältesten Deutsch. Personal- vs. Demonstrativpronomen im Althochdeutschen. In Ziegler, A. & Braun, C. (Eds.). *Traditionen, Innovationen, Perspektiven. Band 1: Diachronie, Althochdeutsch, Mittelhochdeutsch*. (pp.339-363). Berlin, Germany: de Gruyter.

Piaget, J. (1959). *Language and Thought of the child*. London, UK: Routledge & Kegan Paul LTD.

Pickering, M.J. & Garrod, S. (2006). Alignment as the Basis of Successful Communication. *Research on Language and Communication, 4*, 203-228.

Pickering, M.J. & Garrod, S. (2004). Towards a Mechanistic Psychology of Dialogue. *Behavioral and Brain Sciences, 27*, 169-225.

Prabhu, G.V., Helander, M.G., & Shalin, V.L. (1995). Effect of Product Structure on Manual Assembly Performance. *The International Journal of Human Factors in Manufacturing, 5*, 149-161.

Prince, E.F. (1981). Toward a Taxonomy of Given-New Information. In Cole, P. (Ed.). *Radical Pragmatics*. (pp.223-255). New York, USA: Academic Press.

Reisberg, D. (2010). *Cognition. Exploring the Science of the Mind.* 4th *edition*. New York, USA: W.W. Norton & Company.

Reither, F. (1996). Denken und Sprache. In Dörner, D. & Selg, H. (Eds.). *Psychologie – Eine Einführung in ihre Grundlagen und Anwendungsfelder. 2. überarbeitet und erweiterte Auflage*. (pp.194-210). Suttgart, Germany: Kohlhammer.

Richardson, M. (2011). Assembly Complexity and the Design of Self-Assembly Products. In Karwowski, W., Soares, M.M., & Stanton, N.A. (Eds.). *Human Factors and Ergonomics in Consumer Product Design. Methods and Techniques*. (pp.187-200). Boca Raton, USA: CRC Press.

Richardson, M. (2007). Errors, accidents and self-assembly products. In Bust, P.D. (Ed.). *Contemporary Ergonomics 2007*. (pp. 305-310). London, UK: Taylor & Francis.

Richardson, M. & Jones, G. (2005). To DIY for: The characteristics that make self-assembly tasks difficult. In Bust, P.D. & McCabe, P.T. (Eds.). *Contemporary Economics 2005*. (pp. 180-184). London, UK: Taylor & Francis.

Richardson, M., Jones, G., & Torrance, M. (2004). Identifying the task variables that influence perceived object assembly complexity. *Ergonomics, 47*, 945-964.

Rieser, H. (1997). Repräsentations-Metonymie, Perspektive und Koordiantion in aufgabenorientierten Dialogen. In Umbach, C., Grabski, M., & Hörnig, R. (Eds.). *Perspektive in Sprache und Raum. Aspekte von Repräsentation und Perspektivität*. (pp.169-194). Wiesbaden, Germany: Deutscher Universitäts Verlag.

Rieser, H. (1996). *Repräsentations-Metonymie, Perspektive und Koordination in aufgaben-orientierten Dialogen. Report 96-6, SFB 360 "Situierte Künstliche Kommunikatoren"*, Universität Bielefeld.

Ritter, F.E. & Larkin, J.H. (1994). Developing process models as summaries of HCI action sequences. *Human-Computer Interaction, 9*, 345-383.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

Rosch, E. (1975). Cognitive representation of semantic categories. *Journal of Experimental Psychology: General, 104*, 193-233.

Roth, T. (1987). Erfolg bei der Bearbeitung komplexer Probleme und linguistische Merkmale des Lauten Denkens. *Sprache & Kognition, 4*, 208-220.

Roth, T. (1985). Sprachstatistisch objektivierbare Denkstilunterschiede zwischen „guten" und „schlechten" Bearbeitern komplexer Probleme. *Sprache & Kognition, 4*, 178-191.

Russo, J.E., Johnson, E.J., & Stephens, D.L. (1989). The validity of verbal protocols. *Memory & Cognition, 17*, 759-769.

Schegloff, E.A. (1979). The relevance of repair to syntax-for-conversation. In Givón, T. (ed.). *Syntax and Semantics 12: Discourse and Syntax*. (pp.261-286). New York, USA: Academic Press.

Schellhorn, S.E., Griego, J., & Schmid, U. (2007). Transformational and derivational strategies in analogical problem solving. *Cognitive Processing, 8*, 45-55.

Schiffrin, D. (1988). *Discourse markers*. Cambridge, UK: Cambridge University Press.

Schmid, H-J. (1993). *Cottage and Co., Idea, Start vs. Begin: Die Kategorisierung als Grundprinzip einer differenzierten Bedeutungsbeschreibung*. Tübingen, Germany: Niemeyer.

Schmid, U. & Carbonell, J. (1999). Empirical evidence for derivational analogy. *Proceedings der 4.Fachtagung der Gesellschaft für Kognitionswissenschaft*. (pp.116-121). Bielefeld, Germany: Bielefeld Universität.

Schober, M.F. & Brennan, S.E. (2003). Processes of Interactive Spoken Discourse: The Role of the Partner. In Graesser, A.C., Gernsbacher, M.A., & Goldman, S.R. (Eds.). *Handbook of discourse processes*. (pp.123-164). Hillsdale, NJ: Lawrence Erlbaum.

Schober, M.F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*, 1-24.

Schober, M.F. & Clark, H.H. (1989). Understanding by Addressees and Overhearers. *Cognitive Psychology, 21*, 211-232.

Schoenfeld, A.H. (1985). *Mathematical Problem Solving*. Orlando, USA: Academic Press.

Schott, F. & Ghanbari, S.A. (2012). *Bildungsstandards, Komptenzdiagnostik und Kompetenz-orientierter Unterricht zur Qualitätssicherung des Bildungswesens. Eine problemorientierte Einführung in die theoretischen Grundlagen*. Göttingen, Germany: Waxmann.

Schooler, J.W. (2011). Introspecting in the Spirit of William James: Comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin, 137*, 345-350.

Schooler, J.W., Ohlsson, S., & Brooks, K. (1993). Thoughts Beyond Words: When Language Overshadows Insight. *Journal of Experimental Psychology: General, 122*, 166-183.

Schraw, G., Dunkle, M.E., & Bendixen, L.D. (1995). Cognitive Processes in Well-defined and Ill-defined Problem Solving. *Applied Cognitive Psychology, 9*, 523-538.

Seifert, C.M. (1999). Situated Cognition and Learning. In Wilson, R.A. & Keil, F.C. (Eds.). *The MIT Encyclopedia of the Cognitive Sciences*. (pp.767-769). Cambridge, MA: MIT Press.

Short, E.J., Evans, S.W., Friebert, S.E., & Schatschneider, C.W. (1991). Thinking aloud during problem solving: facilitation effects. *Learning and Individual Differences, 3*, 109-122.

Siegler, R. & Stern, E. (1998). Conscious and unconscious strategy discoveries. *Journal of Experimental Psychology: General, 127*, 377-397.

Simon, H.A. & Newell, A. (1971). Human problem solving: the state of the theory in 1970. *American Psychologist, 26*, 145-159.

Sinha, C. (2007). Cognitive Linguistics, Psychology, and Cognitive Science. In Geeraerts, D. & Cuyckens, H. (Eds.). *The Oxford Handbook of Cognitive Linguistics*. (pp.1266-1294). Oxford, UK: Oxford University Press.

Slobin, D.I. (1996). From "Thought and Language" to "Thinking and Speaking". In Gumprez, J.J., & Levinson, S.C. (Eds.). *Rethinking linguistic relativity*. (pp.70-96). Cambridge, UK: Cambridge University Press.

Someren, M.W. van, Barnard, Y.F., & Sandberg, J.A.C. (1994). *The Think Aloud Method. A practical guide to modeling cognitive processes*. London, UK: Academic Press.

Steels, L. & Loetzsch, M. (2009). Perspective Alignment in Spatial Language. In Coventry, K.R., Tenbrink, T., & Bateman, J. (Eds.). *Spatial Language and Dialogue.* (pp.70-88). Oxford, USA: Oxford University Press.

Steffensen, S.V. (2013). Human interactivity: problem-solving, solution-probing, and verbal patterns in the wild. In Cowley, S.J., & Vallée-Tourangeau, F. (Eds.). *Cognition Beyond the Body: Interactivity and Human Thinking*. (pp.195-221). Dordrecht, Germany: Springer.

Steffensen, S.V. (2012). Care and conversing in dialogical systems. *Language Sciences, 34*, 513-531.

Steiner, E. & Teich, E. (2004). Metafunctional profile of the grammar of German. In Caffarel, A., Martin, J.R., & Matthiessen, C.M.I.M. (Eds.) (2004). *Language Typology – A functional perspective*. (pp.139-184). Amsterdam, Netherlands: John Benjamins.

Sterelny, K. (2004). Externalism, Epistemic Artefacts and the Extended Mind. In Schantz, R. (Ed.). *The Externalist Challenge*. (pp.237-254). Berlin, Germany: Walter de Gruyter.

Sternberg, R.J. & Sternberg, K. (2012). *Cognitive Psychology. 6th edition*. Belmont, USA: Wardsworth.

Stillings, N.A., Weisler, S.E., Chase, C.H., Feinstein, M.H., Garfield, J.L., & Rissland, E.L. (1995). *Cognitive Science. An Introduction. 2nd edition*. Cambridge, MA: MIT Press.

Stutterheim, C. von, Mangold-Allwinn, R., Baratelli, S., Kohlmann, U., & Kölbing, H.-G. (1993). Reference to objects in text production. In Nuyts, J. & Pederson, E. (Eds.). *Prespectives on language and conceptualization*. (pp.99-125). Amsterdam, Netherlands: John Benjamins.

Taatgen, N. (2012). The Nature and Transfer of Cognitive Skill. In Rußwinkel, N., Drewitz, U., & van Rijn, H. (Eds.). *Proceedings of ICCM 2012, 11th International Conference on Cognitive Modeling*. (p.xi). Berlin, Germany: Universitätsverlag der TU Berlin.

Talmy, L. (2000). *Toward A Cognitive Semantics. Volume I: Concept Structuring Systems*. Cambridge, MA: MIT Press.

Taylor, C. & Gibbs, G.R. (2010). How and what to code. Online QDA Web Site (http://onlineqda.hud.ac.uk/ Intro_QDA/how_what_to_code.php; 22.05.2013).

Tenbrink, T., & Seifert, I. (2011). Conceptual layers and strategies in tour planning. *Cognitive Processing, 12*, 109-125.

Tenbrink, T. (2010). CODA: Kognitive Diskursanalyse. In Stolz, T., Ruigendijk, E., & Trabant, J. (eds.). *Linguistik im Nordwesten: Beträge zum 1. Nordwestdeutschen Linguistischen Kolloquium*. (pp.117-134). Bochum, Germany: Brockmeyer.

Tenbrink, T. (2009). Identifying objects in English and German: a Constrastive Linguistic Analysis of Spatial Reference. In Coventry, K., Tenbrink, T., & Bateman, J. (Eds.). *Spatial Language and Dialogue.* (pp.104-118). Oxford, USA: Oxford University Press.

Tenbrink, T. (2008). The verbalization of cognitive processes: Thinking-aloud and retro-spective reports. In Ramm, W. & Fabricius-Hansen, C. (Eds.). *Linearisation and Segmentation in Discourse. Multidisciplinary Approaches to Discourse 2008 (MAD 08), Feb 20-23 2008, Lysebu, Oslo.* (pp.125-135). Oslo: Dept. of Literature, Area Studies and Europ. Languages, University of Oslo.

Tenbrink, T. (2005). Identifying Objects on the Basis of Spatial Contrast: An Empirical Study. In Freksa, C., Knauff, M., Krieg-Brückner, B., Nebel, B., & Barkowsky, T. (Eds.). *Spatial Cognition IV: Reasoning, Action, Interaction. International Conference Spatial Cognition 2004, Frauenchiemsee, Germany, October 2004, Proceedings.* (pp.124-146). Berlin, Germany: Springer.

Tenbrink, T., Andonova, E., & Coventry, K. (2008). Negotiating relationships in dialogue: The role of the addressee. In Ginzburg, J., Healey, P., & Sato, Y. (Eds.). *Proceedings of LONDIAL – The 12th SEMIDIAL workshop, June 2nd - June 4th, 2008, King's College, London, UK*. (pp. 201-208). London, UK: Queen Mary University.

Tenbrink, T., Bergman, E., & Konieczny, L. (2011). Wayfinding and description strategies in an unfamiliar building. *Proceedings of the 32nd Annual Conference of the Cognitive Science Sociecty. Boston, Massachussets, USA, July 20-23.2011*, 1262-1267.

Tenbrink, T. & Gralla, L. (2009). Accessing complex cognitive processes via linguistic protocol analysis. In Schmid, U., Ragni, M., & Knauff, M. (Eds.). *Proceeding of the KI 2009 Workshop „Complex Cognition" Paderborn, Germany, September 15, 2009. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 82*. (pp.1-12). Bamberg, Germany: Bamberg University.

Tenbrink, T. & Hui, S. (2007) Negotiating Spatial Goals with a Wheelchair. *Proceedings of the 8th SIGdial wokshop on Discourse and Dialogue. Antwerp, September 2007*, 103-110.

Tenbrink, T. & Wiener, J. (2009). The verbalization of multiple strategies in a variant of the traveling salesperson problem. *Cognitive Processing, 10*, 143-161.

Thagard, P. (1996). *Mind. An Introduction to Cognitive Science*. Cambridge, MA: MIT Press.

Thieroff, R. (2010). Mood in German. In Rothstein, B. & Thierhoff, R. (Eds.) *Mood in the Languages of Europe*. (pp.133-154). Amsterdam, Netherlands: John Benjamins.

Tomasello, M. (2005). *Construing A Language. A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (1998). Reference: Intending that others jointly attend. *Pragmatics & Cognition, 6*, 229-243.

Tversky, B., Heiser, J., Lee, P., and Daniel, M.P. (2009). Explanations in gesture, diagram, and word. In Coventry, K., Tenbrink, T., & Bateman, J. (Eds.). *Spatial Language and Dialogue*. (pp.119-131). Oxford, USA: Oxford University Press.

Tversky, B., Agrawala, M., Heiser, J., Lee, P.U., Hanrahan, P., Phan, D., Stolte, C., & Daniele, M.-P. (2007). Cognitive design principles for generating visualizations. In Allen, G. (Ed.). *Applied spatial cognition: From research to cognitive technology*. (pp. 53-73). Mahwah, NJ: Erlbaum.

Tversky, B. & Hemenway, K. (1984). Objects, Parts, and Categories. *Journal of Experimental Psychologie. General, 113*, 69-197.

Ungerer, F. & Schmid, H.-J. (2006). *An introduction to cognitive linguistics. 2nd edition.* London, UK: Pearson Education Limited.

Vallée-Tourangeau, F., Weller, A., & Villejoubert, G. (2011). Distributed Cognition and Insight Problem Solving. *Proceedings of Experimental Psychology Society Meeting 13-15 April 2011, Oxford, UK*, 273-278.

Vallée-Tourangeau, F. & Wrightman, M. (2010). Interactive skills and individual differences in a word production task. *AI & Society, 25*,433-439.

Vorwerg, C. & Tenbrink, T. (2007). Discourse Factors Influencing Spatial Descriptions in English and German. In Barkowsky, T., Knauff, M., Ligozat, G., & Monetto, D. (Eds.). *Spatial Cognition IV: Reasoning, Action, Interaction*. (pp.470-488). Berlin, Germany: Springer.

Vygotsky, L. (1986). *Thought and Language. Translated by Alex Kozulin*. Cambridge, MA: MIT Press.

Wachsmuth, I. & Jung, B. (1996). Dynamic Conceptualization in a Mechanical-Object Assembly Environment. *Artificial Intelligence Review, 10*, 345-368.

Wedman, J., Wedman, J., & Folger, T. (1996). Analysis of analogical problem-solving processes via think-aloud protocols. *Journal of Research and Development in Education, 30*, 51-62.

Weinrich, H. (2005). *Textgrammatik der deutschen Sprache. 3.revidierte Auflage*. Hildesheim, Germany: Olms Verlag.

Welke, K. (1965). *Untersuchungen im System der Modalverben in der deutschen Sprache der Gegenwart. Ein Beitrag zur Erforschung funktionaler und syntaktischer Beziehungen.* Berlin, Germany: Akademie Verlag.

Whorf, B.L. (1956). *Language, thought and reality*. Cambridge, MA: MIT Press.

Wiener, J. & Tenbrink, T. (2008). Traveling salesman problem: The human case. *KI 1/08 Themenheft KI und Kognition*. (pp.18-22). arenDTaP Verlag.

Wilson, T.D. (1994). The Proper Protocol: Validity and Completeness of Reports. *Psychological Science, 5*, 249-252.

Witteck, A. & Tomaello, M. (2005). Young children's sensitivity to listener knowledge and perceptual context in choosing referring expressions. *Applied Psycholinguistics, 26*, 541-558.

Wunderlich, D. (1981). Modalverben im Diskurs und im System. In Rosengren, I. (Eds.). *Sprachen und Pragmatik. Lunder Symposium 1980*. (pp. 11-53). Lund, Sweden: CWK Gllerup.

Zacks, J.M., Tversky, B., & Iyer, G. (2001). Perceiving, Remembering, and Communicating Structure in Events. *Journal of Experimental Psychology: General, 130*, 29-58.

Zhang, J. & Norman, D.A. (1994). Representations in Distributed Cognitive Tasks. *Cognitive Science, 18*, 87-122.

Eidesstattliche Versicherung

Ich versichere hiermit, dass ich meine Dissertation "Linguistic Representation of Problem Solving Processes in Unaided Object Assembly" selbständig und ohne fremde Hilfe angefertigt, und dass ich alle von den Autoren wörtlich übernommenen Stellen wie auch die sich an die Gedankengänge anderer Autoren eng anlehnenden Ausführungen meiner Arbeit besonders gekennzeichnet und die Quellen zitiert habe.

Hannover, den 31. Juli 2013

— Linn Gralla —