Processing instructions

Petra Weiß, Thies Pfeiffer, Hans-Jürgen Eikmeyer, and Gert Rickheit

Abstract. Instructions play an important role in everyday communication, e.g. in task-oriented dialogs. Based on a (psycho-)linguistic theoretical background, which classifies instructions as requests, we conducted experiments using a cross-modal experimental design in combination with a reaction time paradigm in order to get insights in human instruction processing. We concentrated on the interpretation of basic single sentence instructions. Here, we especially examined the effects of the specificity of verbs, object names, and prepositions in interaction with factors of the visual object context regarding an adequate reference resolution. We were able to show that linguistic semantic and syntactic factors as well as visual context information context influence the interpretation of instructions. Especially the context information proves to be very important. Above and beyond the relevance for basic research, these results are also important for the design of human-computer interfaces capable of understanding natural language. Thus, following the experimental-simulative approach, we also pursued the processing of instructions from the perspective of computer science. Here, a natural language processing interface created for a virtual reality environment served as basis for the simulation of the empirical findings. The comparison of human vs. virtual system performance using a local performance measure for instruction understanding based on fuzzy constraint satisfaction led to further insights concerning the complexity of instruction processing in humans and artificial systems. Using selected examples, we were able to show that the visual context has a comparable influence on the performance of both systems, whereas this approach is limited when it comes to explaining some effects due to variations of the linguistic structure. In order to get deeper insights into the timing and interaction of the sub-processes relevant for instruction understanding and to model these effects in the computer simulation, more specific data on human performance are necessary, e.g. by using eye-tracking techniques. In the long run, such an approach will result in the development of a more natural and cognitively adequate human-computer interface.

1. Introduction

Instructions play an important role in everyday communication, especially in the context of education or at work where they are often embedded in task-

oriented dialogs. Research on instruction processing is, among other things, also particularly relevant for the design of human-computer interfaces capable of understanding natural language. The development of such an interface can be the objective only of an interdisciplinary approach undertaken as a joined effort of (psycho-)linguistics and computer science.

1.1. Instructions in the research line of the CRC

Our research follows the experimental-simulative approach. Based on the theoretical background in psycholinguistics, we conduct experiments in order to collect empirical evidence on the performance of human instruction processing. At the same time we approach our research questions constructively from the perspective of computer science and human-computer interface design. The natural language processing interface created for a virtual reality environment is the basis for the simulation of our empirical findings. Using virtual reality techniques allows us to employ a broad range of interaction between human and machine while still being able to concentrate on the higher levels of communication and not being overwhelmed by sensory and motor control problems. Comparing the performance of both systems, the human vs. the machine constructor, leads to further insights into the complexity of the problem of instruction processing and the processes involved in human instruction understanding, which will finally lead to a more natural human-computer interface.

Communication is about contexts. In our setting we placed the scenario of the Collaborative Research Center (CRC) 360, where a human instructor directs an artificial robot constructor, in an immersive virtual environment (cf. Fig. 1). In a collaborative construction task the human instructor guides the system in building a toy airplane from a (virtual) wooden toy kit consisting of a set of generic parts, such as bolts, cubes, or bars. Thus the roles of the interlocutors are not equal, as the instructor is assumed to know how to build the desired object, and the constructor is expected to realize the instructor's directions. The system is represented visually by "Max" (Kopp et al. 2003), a human-sized virtual agent. He provides the human instructor with a conversational partner to attend to instead of addressing the void. This is important in order to establish a more natural communicative situation.

In the research presented in this chapter, we are interested in the role of the constructor and the way she interprets the verbal instructions given by the instructor. In doing so, we concentrate on basic single sentence instructions such as *Connect the red bolt with the cube*, and do not permit full

games with several turns. This allows us to focus on the effects of verb- and object-specificity, prepositions, and the influence of the visual context.

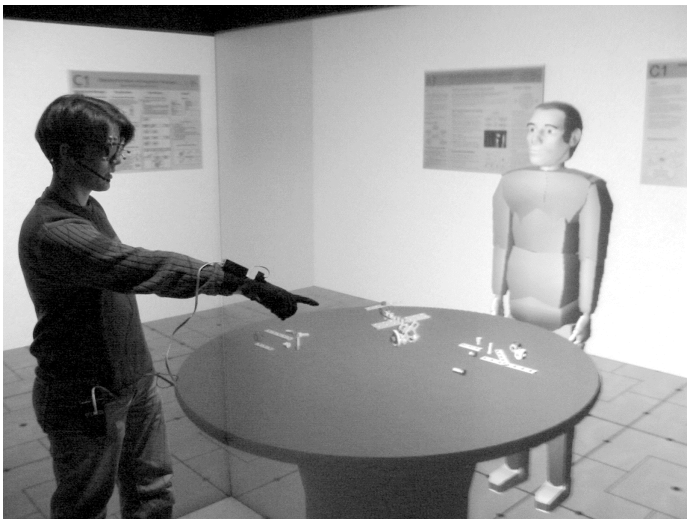


Figure 1. In the virtual reality setting, the user instructs the system, represented by the virtual agent "Max", in building an airplane from toy building blocks. The human-computer interface supports both speech and gestures, as in the example: *Nimm die ↘ rote Schraube! – Take the ↘ red bolt!* (↘ indicates the stroke of the pointing gesture).

1.2. The structure of this chapter

In the first part of this chapter we relate instructions to their theoretical background in linguistics, especially speech act theory, as well as in psycholinguistics and identify important components of instructions. In the second part we present experiments undertaken to investigate how humans perform when processing instructions under different linguistic and contextual conditions. In part three we present the human-computer interface for a situational understanding of instructions. This presentation will mainly focus on reference resolution where the conceptual information conveyed by the instruction and interpreted by a speech-processing system is used to identify the intended objects in the virtual environment. In part four we will develop a local performance measure for instruction understanding in the simulation. This measurement will then allow us to relate the simulative approach to the

results of the psycholinguistic experiments. Using selected examples we will show that the visual context has a comparable influence on the performance of both systems, human and machine. However, this approach is limited when it comes to explaining some effects evoked by variations of the linguistic structure. We will conclude with a discussion and outline our plans for further research.

2. Instructions in linguistic theory

2.1. Instructions as requests

Instructions can be subsumed into the class of utterances called requests (Carroll and Timm 2003; Hindelang 1978). Requests are speech acts with which a speaker wants to prompt his or her partner to do something or to behave in a special way intended by the speaker (Graf and Schweizer 2003; Herrmann 1983: 112–151, 2003), e.g. to take a further step in the assembly of the toy airplane. Based on speech act theory (Austin 1962; Searle 1969, 1976) requests in turn can be assigned to the class of directives (e.g. to command, to request, to permit, to advise, etc.). The basic assumption in speech act theory is that language use is not only information transfer but a special kind of acting with language. In speech act theory, the actions performed with language are categorized according to their communicative function, the illocution (Rolf 1997). This means that utterances are accounted for by their intended or achieved effectiveness.

The realization of the illocutionary act of requesting does not depend on a special grammatical sentence form (Wunderlich 1984). Requests can be formulated using an explicit performative utterance like *I call on you to take the red bolt*, or a declarative sentence with a modal verb such as *You should take the red bolt*, and of course using an instruction with an imperative: *Take the red bolt!* Imperative sentences are prototypical realizations of requests (Wunderlich 1984). Thus, it is not possible to identify an utterance as a request solely from its linguistic form. As a consequence, in speech act theory conditions were formulated that have to be fulfilled in order to identify an utterance as a request (cf. Herrmann and Grabowski 1994: 163; Rolf 1997):

(i) The action will be conducted in the future.
(ii) The speaker wants the partner to conduct the action.
(iii) The speaker believes that the partner is able to conduct the action.
(iv) The speaker believes that the partner is willing to conduct the action.

(v) The speaker presumes that the partner will not conduct the action anyway.

The identification of an utterance as a request depends on the interpretation of a complex combination of linguistic and non-linguistic situational factors and of para-verbal (e.g. smiling) and non-verbal components (e.g. pointing) accompanying the verbal utterances (Grabowski-Gellert and Winterhoff-Spurk 1988). So far, in psycholinguistics there exists no comprehensive or even exhaustive theoretically well-funded systematization of possible variants of requesting. Primarily, two dimensions emerge for the classification of requests: politeness and directness.

With regard to requests, the concept of politeness is closely connected to the idea of "face-work" or "face-management" (Goffman 1989). In a communicative situation interlocutors want to be respected and accepted ("positive face"), and they are afraid of being degraded or losing their reputation ("negative face"; Blum-Kulka, House, and Kasper 1989). For requesting, this means that a speaker has to prompt his or her interlocutor to conduct the intended action and at the same time has to minimize the threatening of the "face" of both (Meyer 1992). Thus, politeness is a possible form of successful "face-work".

Requests can also be either very direct or indirect. Explicit performative utterances and utterances with an imperative verb are direct forms of requests, whereas formulations as questions or as subtle cues are more indirect forms, e.g. by saying *It's very cold in here*, in order to get the partner to close the window. Furthermore, the directness of requests correlates with the degree of politeness (Brown and Levinson 2004). Usually more direct requests are less polite. But very indirect requests are not per se also very polite (Blum-Kulka 1987; Herrmann 1983). Additional factors like cultural norms and situational factors determine whether a special form of requesting is judged as being polite or not (Graf and Schweizer 2003; Herrmann 2003).

2.2. The classification of requests according to AUFF

Especially the dimension of directness with regard to situational factors led to the psycholinguistic classification of (verbal) requests, AUFF, developed by Herrmann (1983: 112-126; Herrmann and Grabowski 1994: 166-174). The acronym AUFF is derived from the German word "*Auff*orderung" [request]. In AUFF five variants of requests are distinguished (cf. Graf and Schweizer 2003; Herrmann 2003):

- Imperative and performative requests (I): The partner is directly committed to do something (e.g. *Take the red bolt! – I call on you to take the red bolt*).
- Requests referring to the legitimation of the speaker (V): The speaker is authorized to commit the partner to do something (e.g. *You must take the red bolt! – I can demand from you to take the red bolt*).
- Requests referring to the secondary goal of the speaker and to conditions concerning the partner (A): The speaker wants the partner to do something and the partner is able and willing to do so (e.g. *Can you take the red bolt?*).
- Requests referring to the primary goal of the speaker and to the conditions concerning the speaker (E): The speaker wants to reach a special target state through the action of the partner (e.g. *I want you to take the red bolt*).
- Requests without referring to the speaker, the partner, deontic conditions (conditions concerning social conventions or norms), or to the action the speaker wants the partner to conduct; "hints" (H): The speaker only refers to conditions of the intended target state (e.g. *A red bolt is missing in the assembly*).

AUFF is a (partial) structure of implications. In this system requests are classified with respect to situational factors like legitimation of the speaker, his primary goals or the intended actions. These factors are interrelated in different and complex ways. In AUFF, the directness of requests is defined by the implicational relations between facts different kinds of requests refer to. Direct and indirect requests can be considered as being polite or not, in dependence on the communicative situation and different verbalizations.

Furthermore, AUFF is not only a descriptive psycholinguistic taxonomy of requests but rather it is conceived as a cognitive scheme represented mentally by an individual. This implies that, given the actual communicative situation, verbalizing of only a few components is sufficient to activate the entire AUFF system – following the principle of "pars pro toto" (Herrmann and Grabowski 1994: 349). As regards directness, the speaker is confronted with a tradeoff between communicative clarity (very direct requests) and the risk of misunderstanding or reactance by the partner. In this respect, requests with medium directness hold a small communicative risk and are used most frequently (Blum-Kulka, House, and Kasper 1989).

Aside from the question of how to classify requests, there is also the problem of which factors determine what kind of request is chosen by a speaker in a specific situation. Following Herrmann and Grabowski (1994;

see also Graf and Schweizer 2003; Herrmann 2003) essentially four factors, as conceived by the speaker, determine his or her choice:

(i) the willingness and
(ii) the ability of the partner to conduct the intended action;
(iii) the speaker's legitimation to request the partner to conduct the action;
(iv) the urgency to reach the primary goal connected with the intended action.

In a couple of experimental and field studies it was possible to identify systematic relations between the four factors mentioned above and the kind of request being produced (Herrmann 2003; Herrmann and Grabowski 1994: 186–205; Hoppe-Graff et al. 1985; for a critical discussion cf. Engelkamp and Mohr 1986).

2.3. Instructions relevant for the CRC scenario

With respect to the communicative situation of the scenario under consideration here, the interlocutors show a clear role allocation. As both share a common goal, their willingness is expected to be high. The instructor, who knows the plan of the model, has the legitimation to give directions and typically takes the role of the speaker. The partner is the constructor and her task is to follow the speaker's instructions by conducting the intended actions. The communicative situation can therefore be considered a standard situation in which normally simple and direct requests are produced (Herrmann and Grabowski 1994). Under this assumption we will concentrate in the following on simple and direct verbal instructions (basic single sentences). Furthermore, we restricted our experiments to instructions related to actions requiring the connection of parts, as these are most frequently used in the corpus of the CRC (cf. Brandt-Pook 1999).

2.4. Linguistic components of instructions for construction processes

Verbal instructions like *Schraube die rote Schraube in den grünen Würfel* (*Screw the red bolt in the green cube*) consist of several linguistic components which have to be processed in order to identify the action to be performed.

2.4.1. Semantic components

At first, the hearer has to interpret the construction verb (*schrauben*, i.e. *to screw*). To interpret a verb in an instruction means to get to know what to do. But the verb on its own does not convey sufficient information to understand an instruction. The constructor also has to know which objects to use in order to carry out the intended action. Therefore she has to interpret the object names (*Schraube*, i.e. *bolt*; *Würfel*, i.e. *cube*). Interpreting the names of the objects does not only mean to understand the literal meaning of the words registered in a kind of mental lexicon, but also to identify the correct objects for conducting the intended action. With action-related instructions this can only be achieved by taking into account the communicative situation. In addition to the linguistic information, especially the visual object context can be consulted for reference resolution and for the identification of the required actions.

Particularly for instructions, verbs and object names cannot be interpreted in isolation. Only the correct interpretation of the combination of object referents and verb admits the processing of the instruction in the intended way. The specificity of verbs and object names plays an important role. A verb like *to screw* specifies a highly specific action, screwing, which in turn imposes further requirements for the objects to be chosen. Whereas an unspecific verb like *to connect* is less restrictive. The same holds for the object naming. Almost all the objects in the context can be referred to by *part*, whereas the name *bolt* matches only with a few objects.

Aside from verbs and object names, prepositions can be important for the adequate interpretation of an instruction. In the example mentioned above, the preposition *in* implicates the direction of the action: The bolt has to be screwed *into* the cube and not, vice versa, the cube *into* the bolt (the "baufix" cubes have six mounting holes, some with a thread and some without). This is different when combining *to screw* with the preposition *an* (*on*). While the established connection is the same, both directions of action are possible: The bolt can be screwed on the cube and the cube can be screwed on the bolt respectively. Hence, the preposition *in* is more specific than the preposition *on*. Furthermore, this aspect is connected with the allocation of the roles of the objects. In combination with *in*, the bolt is the target object which will be chosen, moved, and connected to the reference object, which, in our case, is the cube. In combination with *on*, the role of target or reference object can be assigned to both objects likewise.

2.4.2. Syntactic factors

One important syntactic factor affecting especially the time course of the interpretation process is the variation of the syntactic position of the components mentioned with respect to the concrete formulation of an instruction. In the first instance the position of the verb of action in an instruction is important. There might be a great difference concerning the interpretation process if the verb of action (*schrauben*, i.e. *screw*) is in front position (e.g. *Schraube in den grünen Würfel die rote Schraube* – literally, *Screw in the green cube the red bolt*) or in final position (e.g. *In den grünen Würfel die rote Schraube schrauben – In the green cube the red bolt (is to be) screwed*). In the first case, it is easy to know right from the beginning that the intended action is to screw; in the second case, the instruction has to be processed completely in order to know which action has to be taken.

These considerations can also be applied to the naming of objects: It is possible to mention the target object (*die rote Schraube*, i.e. *the red bolt*) first and then the reference object (e.g. *Schraube die rote Schraube in den grünen Würfel – Screw the red bolt in the green cube*) or vice versa (e.g. *Schraube in den grünen Würfel die rote Schraube – Screw in the green cube the red bolt*). This variation may affect especially the availability of the information about the object referents. Of course, these aspects also interact with the naming of the objects and with context factors.

3. Psycholinguistic experiments on the processing of instructions

In this section we report on a series of experiments in which we investigated the influence of the linguistic components explicated above on the interpretation of simple and direct instructions to conduct assembly actions (connection of parts) within the scope of the scenario of the CRC.

In Experiment 1, we addressed lexical-semantic factors like the specificity of verbs and of object naming in interaction with factors of the visual object context. In Experiment 2, we examined the influence of a syntactic factor, the position of the verbs in the instructions, in combination with the specificity of the verbs and a variation of the visual context. In Experiment 3, we also varied the order of target and reference object (sequence of arguments) in the instructions with regard to the direction of the intended action mediated by the specificity of the prepositions.

Before reporting these experiments in greater detail, we give a brief description of the general method applied in the experiments.

3.1. General procedure in the experiments

In all experiments we used a cross-modal presentation technique in combination with the reaction time paradigm in order to test the influence of the linguistic and the contextual factors on the processing of simple and direct oral instructions.

Participants were presented pictures with arrangements of objects on a computer screen. In a first step, the participants could see the potential target object in combination with contextual objects in the upper half of the screen. Then they were presented an instruction acoustically, and at the same time another object appeared at the bottom of the screen. We will call this object the reference object because the target object has to be moved and fitted to this object depending on the interpretation of the instruction, especially of the construction verb (see Fig. 2; for a terminological discussion cf. Weiß 2005: 31–33).

Participants had to choose one of the objects presented in the upper row as the appropriate target object by pushing the appropriate button on the keyboard or by a mouse click as fast and correctly as possible. Thus the participants had to conduct an action-related decision task. The selection of the target object indicates how they interpret an instruction. Additionally, reaction times were taken as a measure for the processing complexity of the instructions under consideration.

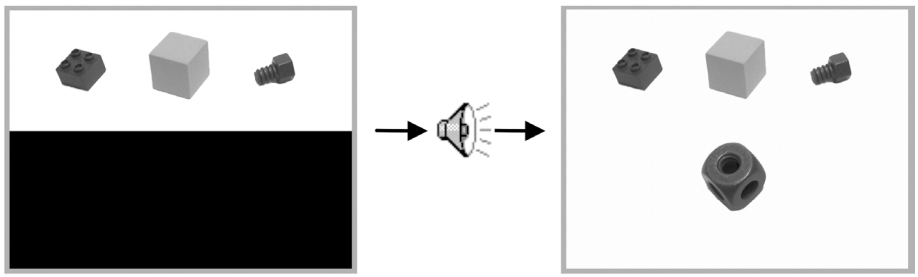


Figure 2. Example of the presentation of an experimental item: First the potential target object (TO) and context objects (CO) are presented (TO top right: red bolt, CO top left and mid: blue LEGO brick, yellow block), then the reference object (RO) (below: green cube) appears, and at the same time an instruction referring to the arrangement of objects is presented acoustically. A sample instruction might be *Schraube die rote Schraube in den grünen Würfel* (*Screw the red bolt in the green cube*).

3.2. Experiments 1 and 2: Influence of verbs, object naming and context

In Experiment 1, we examined the influence of the specificity of verbs and object naming in visually ambiguous vs. unambiguous object contexts. The relevance of the information carried by the construction verb and the contextual information was examined in greater detail in Experiment 2 by systematically varying the position of the verbs in the instructions (Weiß, Hildebrandt, Eikmeyer, and Rickheit 1999; Weiß, Hildebrandt, and Rickheit 1999).

Based on studies that showed sentence processing – in particular, the processing of oral instructions – to take place in an incremental and interactive way (Altmann and Kamide 1999; Tanenhaus et al. 1995), we assumed that, particularly in the case of unspecific linguistic information, the visual object context helps to interpret the instructions.

In these experiments the instructions were always formulated in the following way: The reference object (e.g. *green cube*) was the first object mentioned, and the target object (*red bolt*) the second one, e.g. *Schraube in den grünen Würfel die rote Schraube* (*Screw in the green cube the red bolt*). When considered in isolation, a formulation like *Screw the red bolt in the green cube* might sound more natural. But in the context of the assembly of the toy airplane, there usually is an existing (old) part or aggregate to which a new component should be added. Therefore the constructor has to choose one of the possible objects for a target object and move it to the already determined and fixed reference object. This choice might be easier if the structure of the instruction follows the given-new contract (Clark and Haviland 1977; Hörnig, Oberauer, and Weidenfeld 2002). Also, from the preceding visual presentation of the potential target objects, these objects should already be activated prior to the visual presentation of the reference object and the acoustic presentation of the instruction. Accordingly, by mentioning the reference object simultaneously with its visual presentation, the attention of the participants should not be focused on a particular target object but on the reference object.

3.2.1. Experiment 1: Method, factors, and design

In Experiment 1, several factors were investigated in combination in an orthogonal design. The factor "verb specificity" was varied within cases at two levels (specific vs. unspecific). The factor "specificity of target object naming" was varied between cases at two levels (specific vs. unspecific). Fur-

thermore, two variables concerning the visual "object context", i.e. color and function of the target object relative to two context objects, were varied within cases at two levels (ambiguous vs. unambiguous) each.

- *Verb specificity:* The classification of verbs depending on their level of specificity is based on a transfer of the semantic relation of hyponymy from the classification of nouns (e.g. *a sparrow is a bird*) to the classification of verbs (Miller 1998; Miller and Fellbaum 1991), in which case the relation is termed troponymy (Fellbaum 1998). Troponymy means that specific verbs like *verschrauben* (*to screw*) bear more information (i.e., have higher entropy) than less specific verbs like *verbinden* (*to connect*). Thus, with specific verbs, the possible actions mediated are more constrained than with unspecific verbs. With unspecific verbs, there is a larger amount of possible actions and objects with which these actions can be carried out (Miller 1991: 228–230). The resulting hierarchy of verbs differs from the hierarchy of nouns in that, additionally, the quality of the relation has to be supplied (e.g., "screwing is a special way of connecting"). Taken together, the hierarchy of verbs is shallower than that of nouns, with the number of hierarchy levels normally not exceeding four (Miller 1991: 230). Not in every case is there exactly one superordinate verb for a group of semantically related verbs. As a consequence, it is comparatively difficult to classify verbs based on the relation of troponymy. By using a questionnaire (cf. Weiß, Hildebrandt, and Rickheit 1999), we were able to construct eight pairs of construction verbs differing in their degree of specificity and to combine them with possible objects of action.
- *Specificity of target object naming:* This variation was obtained by using *Teil* (*part*) for an unspecific naming of the target object and a term at the basic level (Rosch 1978) in the specific case, e.g. *Schraube* (*bolt*).
- *Object context:* The referential (un-)ambiguity of color and function of the potential target object was varied in relation to the color and function of two further context objects. In the case of an unambiguous color, only the target object had the color mentioned in the instruction (red, blue, green, or yellow; see Fig. 1); in the case of ambiguous color, all three potential target objects had the same color (for an example, see Fig. 12 below). In the case of an unambiguous function, the intended action could only be carried out with the target object; in the case of functional ambiguity, each of the three objects could serve as the target object (e.g. three bolts; see Fig. 13 below). In a third experimental condition, we combined the referential ambiguity of color and function by creating a set consist-

ing of the target object (e.g. a red bolt), one context object matching the target object in color (e.g., a red cube), and one matching it in function (e.g., a yellow bolt).

In Experiment 1, the unspecific or specific verbs in the instructions were always presented in a sentence final position, as *verbinden* (unspecific) or *verschrauben* (specific) in *Mit dem grünen Teil sollst du die rote Schraube verbinden/verschrauben* (*With the green part the red bolt is to be connected* (unspecific verb) or *In the green part the red bolt is to be screwed* (specific verb).

With this kind of formulation we aimed at making the participants process the information about the object referents first and then the explicit information about the intended action mediated especially by the verb. We assumed that, particularly in combination with ambiguous object arrangements, the information conveyed by (specific) verbs would be of special importance in interpreting the instruction and in selecting the target object.

3.2.2. Experiment 1: Results

On the whole, instructions with specific verbs were processed more quickly than instructions with unspecific verbs. This result holds both in combination with an unambiguous visual object context as in combination with an ambiguous one (Fig. 3).

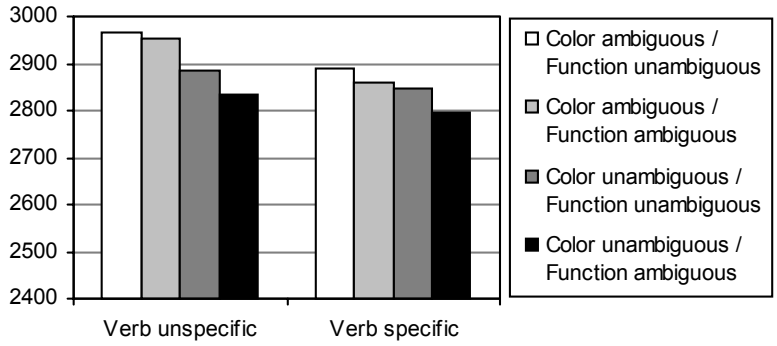


Figure 3. Experiment 1 – Average reaction times (ms) for the choice of the target object for instructions with unspecific and specific verbs in dependence on the visual object context.

Concerning the specificity of object naming of the target object we obtained a contrary result. Here instructions with a specific naming of the target object were processed more slowly than instructions with an unspecific naming (Fig. 4).

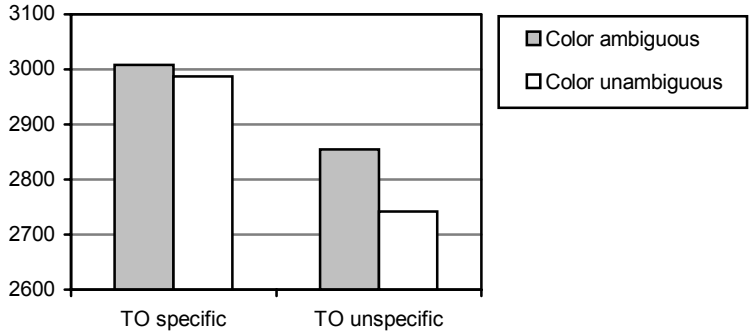


Figure 4. Experiment 1 – Average reaction times (ms) for the choice of the target object: Interaction of specificity of naming of the target object (TO) and (un-)ambiguity of the color of the target object.

With regard to the variation of the contextual factors the following results appeared: Under the condition of referential unambiguous color the instructions were processed more quickly than under the condition with ambiguous color (Fig. 3 and 4). In contrast, instructions related to an object arrangement with a functionally unambiguous target object were processed more slowly than instructions related to an object arrangement with a functionally ambiguous target object (Fig. 3).

Furthermore, the influence of the (un-)ambiguity of the color of the object context interacts with the specificity of the object naming. Especially in the case of an unspecific naming of the target object instructions referring to unambiguous contexts in terms of color are processed more quickly than instructions referring to contexts with ambiguous color. This also means that, especially in the condition with unambiguous color, instructions with unspecific naming of the target objects are processed more quickly than instructions with a specific naming (Fig. 4).

3.2.3. Experiment 1: Discussion

Specific verbs facilitate the interpretation of instructions. But contrary to our expectations, there is no interaction between the specificity of the verbs and

the factors of the visual object context. With specific as well as with unspecific verbs the influence of the variation of the object context is the same (see Fig. 3). This means that the linguistic information mediated by verbs is crucial for the interpretation of the instructions. On the other hand, instructions with specific naming of the target objects are processed more slowly than instructions with unspecific naming. This effect may be due to the fact that in the current situation a specific object naming is redundant and a kind of overspecified object naming (Mangold 1987; Weiß and Barattelli 2003). Especially in the case of unambiguous color of the target object a specific naming of the target object has high entropy which leads to a rich mental representation that results in a complex reference resolution and longer processing times (Fig. 4). This may be not necessary because the correct target object could be selected correctly by its color alone. This corresponds to the main effect that instructions referring to objects in contexts with unambiguous color are processed more quickly than instructions in contexts with ambiguous color (Fig. 4).

The effect of the (un-)ambiguity of the function of the object context is a different one. Here a functionally unambiguous context leads to longer processing times than a functionally ambiguous context (Fig. 3). This rather unexpected result may be due to the fact that the information about the function of the objects is not as directly accessible as the information about their color, which is mediated linguistically (by explicit mentioning) and visually and which in general is central for reference resolution (cf. Weiß and Mangold 1997).

In the following experiment our aim was to find out more about the influence of the verb and context information by a systematic variation of the position of the verbs in the instructions.

3.2.4. Experiment 2: Method, factors, and design

As an additional factor, in Experiment 2 we also varied the position of the verbs in the instructions, aside from verb specificity and (un-)ambiguity of the color and function of the target object in relation to the visual object context. The verbs were presented at either front, mid, or final position. The specificity of the verbs and the color and function of the target object were varied at two levels along the lines of Experiment 1. All factors were varied within cases (see Tab. 1 for the experimental design and for examples of the instructions for assembly). There was no variation of the specificity of object naming.

We expected a replication concerning the effects of the specificity of the verbs and the visual object context. With respect to the factor verb position we expected an interaction with the visual object context: Especially in combination with ambiguous object arrangements, instructions with the verb in final position should be processed more slowly because the utterance has to be processed completely and the decision deferred about the correct target object until the processing of the verb. On the other hand, instructions with the verb in front position should be processed more quickly because right from the beginning on – particularly with specific verbs – it is clear what kind of action has to be conducted.

Table 1. Experiment 2: Design and examples of instructions

	Verb specificity	Verb position	Color context ambiguous / unambiguous
Function context ambig. / unambig.	specific	front	*Verschraube mit dem grünen Teil das rote Teil* *Screw in the green part the red part*
		mid	*Mit dem grünen Teil verschraube das rote Teil* *In the green part screw the red part*
		final	*Mit dem grünen Teil das rote Teil verschrauben* *In the green part the red part is to be screwed*
	un-specific	front	*Verbinde mit dem grünen Teil das rote Teil* *Connect with the green part the red part*
		mid	*Mit dem grünen Teil verbinde das rote Teil* *With the green part connect the red part*
		final	*Mit dem grünen Teil das rote Teil verbinden* *With the green part the red part is to be connected*

3.2.5. Experiment 2: Results

In Experiment 2, the results concerning the effects of the specificity of the verbs were replicated: Instructions containing specific verbs were processed faster than instructions with unspecific verbs (Fig. 5). This effect was independent of the position of the verbs (Fig. 6). Also, the effect of the color of the target object was replicated: Instructions referring to situations with unambiguous target object color were processed faster than instructions referring to a situation with ambiguous target object color (Fig. 5 and 7).

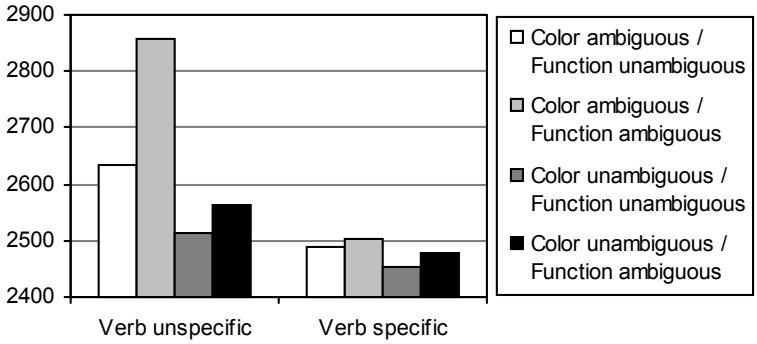


Figure 5. Experiment 2 – Average reaction times (ms) for the choice of the target object for instructions with unspecific and specific verbs in dependence on the visual object context.

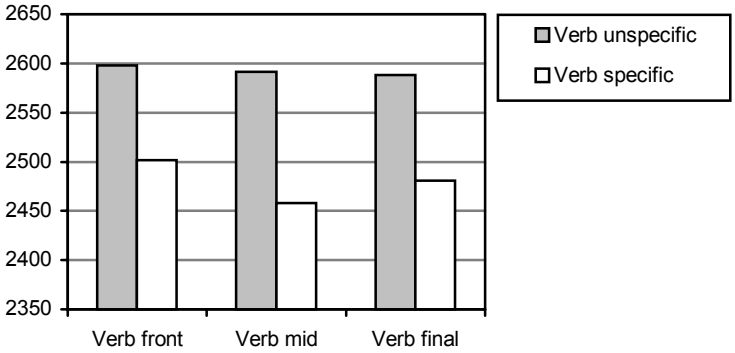


Figure 6. Experiment 2 – Average reaction times (ms) for the choice of the target object for instructions with unspecific and specific verbs in dependence on the verb position.

There was no main effect of the position of the verbs (Fig. 6), but there was an interaction with the ambiguity of the target object color (Fig. 7). As expected, in the case of unambiguous color, instructions with the verb in final position were processed fastest, whereas instructions with the verb in front position were processed slowest; instructions with verbs in the middle took an intermediate time. In the case of ambiguous target object color, the latency for instructions with verbs in front and mid position showed the same course, but contrary to the condition with unambiguous color, there was an increase for instructions with verbs in final position (Fig. 7).

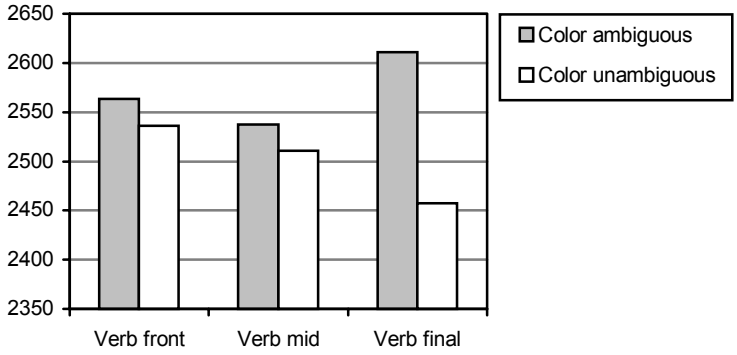


Figure 7. Experiment 2 – Average reaction times (ms) for the choice of the target object: Interaction of color of the target object and verb position.

The effect that the instructions are processed more quickly in combination with a referential ambiguous functional object context could only be replicated for instructions with the verb in final position (Fig. 8). This form of the instructions corresponds to the instructions used in Experiment 1 with respect to the position of the verb. In contrast, instructions with the verb in front or mid position related to a functionally unambiguous context were processed more quickly than instructions related to a functionally ambiguous context (Fig. 8).

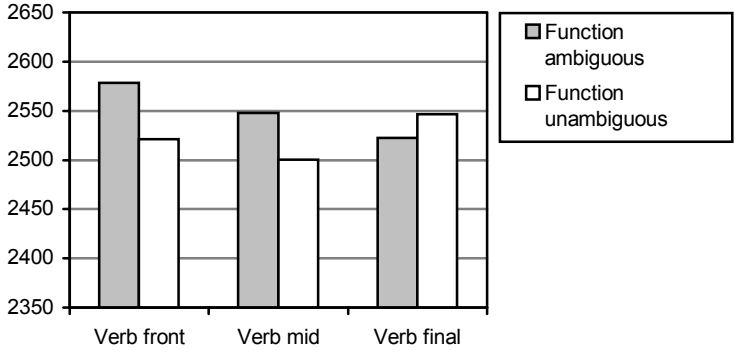


Figure 8. Experiment 2 – Average reaction times (ms) for the choice of the target object: Interaction of function of the target object and verb position.

Though there was no interaction between the factors verb specificity and verb position, we also conducted analyses separated by the verb specificity. It could be shown that the interaction between verb position and color of the

target object could be put down especially to instructions with specific verbs, whereas the main effect of the color of the target object appears particularly in combination with unspecific verbs (Fig. 5).

3.2.6. Experiment 2: Discussion

With Experiment 2, we could replicate the results of Experiment 1 concerning the main effects of verb specificity and of the influence of the target object color on the processing of the instructions under consideration. Again, there is no statistically relevant interaction between verb specificity and the context variables. But by inspecting Figure 5, it becomes apparent that there are clear differences in the influence of the contextual factors on the processing of specific and unspecific verbs. In combination with unspecific verbs, particularly instructions that refer to configurations with ambiguous color and function of the target object lead to longer latencies.

This discrepancy to the result of Experiment 1 might be due to the variation of the position of the verbs in the instructions. This variation leads to differences in the temporal availability of the linguistic information conveyed by the verb on the one hand and linguistic information referring to the context on the other. When the verb is in final position, the information referring to the color of the target object is available early in the interpretation process because it is mentioned prior to the linguistic information about the action as conveyed by the verb. Thus, when the color of the target object is unambiguous, it is possible to utilize this information immediately on processing the color adjective and seeing the object arrangement, so as to directly choose the correct target object. In contrast, in cases with ambiguous color, it is necessary to wait until the verb is interpreted in order to know which action is required and to decide on which object should be chosen as target object – in particular because of the fact that the objects were named unspecifically as *part*. This interpretation is further substantiated by the difference in the reaction times concerning the interaction between color and verb position in the presence of specific and unspecific verbs.

The rather unexpected result concerning the influence of the function of the objects of Experiment 1 – instructions referring to functionally ambiguous object arrangements are processed more quickly than instructions referring to functionally unambiguous arrangements – also becomes clearer when looking at the differences resulting from the variation of the verb position. This effect could only be replicated with verbs in final position. In the case of functional unambiguity of the target object in combination with the verb

in final position, it is necessary to build up more than one possible functional context of action because there are two or three different types of objects. When the verb appears in final position, the verb has to be processed first in order to determine which action has to be performed and with which of the objects this action is possible. In the case of functional ambiguity, however, only one functional context of action has to be built up. This requires less cognitive effort and leads to shorter processing times.

3.2.7. Discussion of Experiments 1 and 2

The results from these experiments show that the processing of instructions does not only depend on linguistic information but also on visual information about the object arrangements under consideration. Especially the linguistic-semantic information mediated by the verb of action as well as the information provided by the visual context (in particular the color of the objects) contributes in a significant way to the processing of the instructions. Furthermore, the syntactic position of the different information units plays an important role. These findings correspond to approaches which suppose that sentence processing or language processing in general can be regarded as an incremental and integrative process (Crain and Steedman 1985; Spivey-Knowlton et al. 1998; Trueswell and Tanenhaus 1994). We interpret these results as evidence that instructions or more generally speaking utterances are processed in a constituent based incremental way (Hildebrandt et al. 1999; Weiß, Kessler et al. 1999).

The findings of Experiment 1 concerning the specificity of the naming of the target object (unspecific naming is processed more quickly than specific naming) were accounted for at first by the fact that in the experimental setting with only three potential objects the specific name of the target object may be redundant. Thus, it might make the understanding of the instructions and the referential interpretation more difficult than an unspecific naming (cf. Weiß and Barattelli 2003).

In two follow-up studies we examined the influence of the specificity of objects naming in more detail within cases. In one experiment we varied the specificity of the naming of the target object and of the reference object. In another experiment we examined whether the number of objects in the visual context (7 vs. 2) takes an effect on the relevance of the specificity of target object naming. We expected that a specific naming of the target object facilitates the processing of the instructions, especially in the case of more than three potential objects of action.

Generally, the result of Experiment 1 concerning the specificity of object naming could not be replicated. In contrast, the reaction times tend to go in the direction expected originally: Instructions with specific naming of the target object were processed faster than those with unspecific naming. But this result only occurs in interaction with the specificity of the naming of the reference object. This effect may be attributed to the linguistic surface of the instructions: The specific naming of the reference object mentioned first in the instructions might lead to a specific naming default. When this specific naming is followed by an unspecific target naming (as in *Screw in the green cube the red part*), this default has to be revised. Such a revision is not necessary with an unspecific naming of the reference object; here, an unspecific naming and a specific naming of the following target object can be processed alike. The fact that the expected effect of the specificity of the object naming did also not occur in the experiment that varied the number of context objects might be taken to indicate that the object arrangements chosen so far are too simple and straightforward to yield clear results concerning the specificity of the object naming. Additionally, the specificity of the target object might not be very helpful because in these experiments there was no variation of the ambiguity of the function or color of the object context. So again, a specific naming in these contexts is a kind of overspecification.

3.3. Experiment 3: Influence of prepositions and sequence of arguments

In Experiment 3 (Weiß 2001), we examined how the specificity of the preposition influences the processing of instructions. We were especially interested in any effects regarding the direction of the intended action, indicated by the assignment of the roles of target and reference object.

3.3.1. Experiment 3: Method, factors, and design

Participants in Experiment 3 viewed pictures with four objects on a computer monitor (e.g. red bolt, yellow bolt, red cube, yellow cube; for an example see Fig. 14 below). At the same time, an oral instruction was presented acoustically. Participants had to choose one of the objects as the correct target object by pressing a key on the keyboard. The reaction times for their decisions were measured.

Three factors were varied within cases: the specificity of the preposition, the sequence of arguments, and the position of the verb. The specificity of

the verbs was not manipulated in this experiment; however, most of the verbs that we used can be classified as specific.

- *Specificity of preposition:* At the level of the verb-argument structure, the variation in the specificity of the preposition refers to whether or not it is possible to unambiguously assign an argument like a prepositional phrase (Britt 1994). For the combination of the verb *to screw* with the preposition *in* and a visual context comprising, for example, a cube with a hole and a bolt, this assignment is specific, since only the prepositional phrase *in the cube* is possible. In contrast, the combination of *to screw* with the preposition *on* in the same context is less specific, because two corresponding prepositional phrases (*on the cube* / *on the bolt*) are possible (cf. Olsen 1996). – We expected instructions with specific prepositions to take more processing time than instructions with less specific prepositions because in the former case, there is only one possible assignment of the arguments. In the latter case, however, participants have to choose exactly one object as correct target object among several candidates – and such a decision process presumably takes time.
- *Sequence of arguments:* As a second factor, we varied the sequence of the naming of the objects and thus, the sequence of the arguments. In the experiments reported so far, the reference object (RO) was always mentioned first and the target object (TO) second, as in *Screw in the blue part* (RO) *the red part* (TO). In the present experiment, we varied the sequence of the arguments and also presented instructions like *Screw the red part* (TO) *in the blue part* (RO). – Based on observations on the processing of instructions for the establishment of spatial relations between objects (Harris 1975; Huttenlocher and Strauss 1968), we expected that instructions in which the potential target object is mentioned first are processed faster than instructions in which it is mentioned last (cf. the "advantage of first mention"; Gernsbacher 1991).
- *Verb position:* The third experimental factor again was the position of the verb of action in the instructions, with front, mid and final position as the factor levels. We expected a modifying influence on the processing of the instructions. This assumption was based on the results obtained so far and on the fact that the variation of the position of the verb also leads to a variation in the availability of the information about the action to be performed.

The orthogonal combination of these factors yields the experimental design which, together with examples of the instructions, is shown in Table 2.

Table 2. Experiment 3: Design and examples for instructions with the verb *schrauben* (*to screw*), comparing the prepositions *in* (*in*) and *an* (*on*)

Preposition specificity	Verb position	Argument sequence	
specific	front	TO…RO	*Schraube ein rotes Teil in ein blaues Teil* *Screw a red part in a blue part*
		RO…TO	*Schraube in ein blaues Teil ein rotes Teil* *Screw in a blue part a red part*
	mid	TO…RO	*Ein rotes Teil schraube in ein blaues Teil* *A red part screw in a blue part*
		RO…TO	*In ein blaues Teil schraube ein rotes Teil* *In a blue part screw a red part*
	final	TO…RO	*Ein rotes Teil in ein blaues Teil schrauben* *A red part in a blue part is to be screwed*
		RO…TO	*In ein blaues Teil ein rotes Teil schrauben* *In a blue part a red part is to be screwed*
un-specific	front	TO…RO	*Schraube ein rotes Teil an ein blaues Teil* *Screw a red part on a blue part*
		RO…TO	*Schraube an ein blaues Teil ein rotes Teil* *Screw on a blue part a red part*
	mid	TO…RO	*Ein rotes Teil schraube an ein blaues Teil* *A red part screw on a blue part*
		RO…TO	*An ein blaues Teil schraube ein rotes Teil* *On a blue part screw a red part*
	final	TO…RO	*Ein rotes Teil an ein blaues Teil schrauben* *A red part on a blue part is to be screwed*
		RO…TO	*An ein blaues Teil ein rotes Teil schrauben* *On a blue part a red part is to be screwed*

3.3.2. Experiment 3: Results

As expected, the reaction times were significantly longer in the case of a specific preposition than in the case of an unspecific preposition (Fig. 9). The difference in the sequence of the arguments did not take a significant main effect, but the average reaction times were on the whole longer with the sequence TO–RO than with the sequence RO–TO. This means that, contrary to our assumption, the processing of the instructions took longer when the target object was mentioned first than when it was mentioned second (Fig. 10).

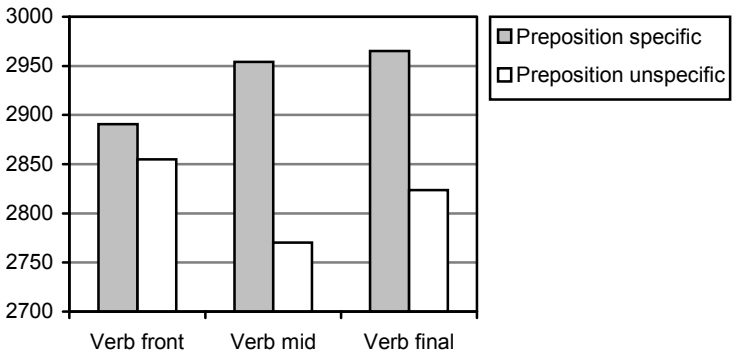


Figure 9. Experiment 3 – Average reaction times (ms) for the choice of the target object: Interaction of verb position and specificity of preposition.

Moreover, the verb position had no significant effect on its own. This factor interacted both with the specificity of the preposition and with the sequence of the arguments. Because of the quality of this interaction, it was possible to also interpret the main effect of the specificity of the preposition in its own right (Fig. 9). The interaction between verb position and sequence of arguments (Fig. 10) required a more differentiated consideration of the conditions, which yielded that only in the condition with the verb in final position there was a statistically significant difference in the reaction times with the sequence TO–RO showing a distinct increase of reaction times compared to the sequence RO–TO.

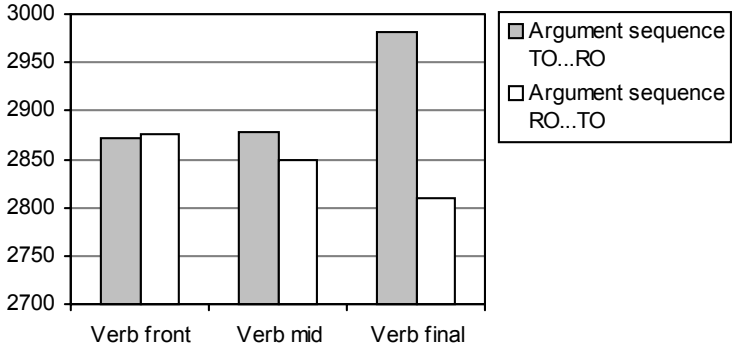


Figure 10. Experiment 3 – Average reaction times (ms) for the choice of the target object: Interaction of verb position and sequence of arguments.

3.3.3. Experiment 3: Discussion

As expected, reaction times for instructions with specific prepositions are longer than for instructions with unspecific prepositions. The processing of the instructions is more costly when participants have to choose exactly one object as the correct target object than in the case with two possible target objects. In a similar way, Chambers et al. (2002) were able to show an influence of prepositions on the processing of instructions with an eye-tracking study. Their instructions varied in the selectivity of the prepositions (specific vs. unspecific). In their setting, the information mediated by the preposition restricted the possible interpretation of the following noun phrase in a prospective way (as measured by the sequence of eye fixations). As in our experiment, the specificity of the preposition leads to a pre-selection of the resolution of the object references.

In our experiment, the sequence of the linguistic components relevant for the processing of the instructions again plays an important role. The effect of the specificity of the preposition appears only in the conditions with the verb in mid or final position (Fig. 9). When the action to be conducted is clearly specified by the verb right from the beginning of the utterance, the information contributed by the subsequent components may already have been established.

With respect to the sequence of the arguments we obtained the unexpected result that reaction times were not faster in the condition TO–RO (in which the target object was mentioned first and the reference object second). But again, the position of the action verb had a modifying influence: Only in the verb final condition, reaction times for TO–RO were significantly longer than for RO–TO (Fig. 10). Evidently, the earlier the verb information (which is relevant for acting) is available, the less influential are the other factors. However, in the condition RO–TO, the reaction time pattern is reversed (Fig. 10). Such a sequence of arguments goes along with an unusual formulation of instructions, and in order to process such instructions it is necessary to jump back and forth between the relevant information units (see Tab. 2).

On the whole, this result corresponds with findings in favor of the idea that it is easier to relate a (new) target object to a reference object already given (Oberauer and Wilhelm 2000; cf. also Hörnig, Oberauer, and Weidenfeld 2002). Furthermore, in the condition RO–TO, the target object is always the object mentioned second. As the experimental task was to choose this target object, it was possible to react immediately after processing this object reference. Thus, this result is in line with the idea of an effect of recency of mentioning (cf. Gernsbacher 1991).

3.4. General discussion of the experimental results

With our experiments we were able to show that linguistic-semantic factors such as the specificity of verbs, objects, and prepositions as well as syntactic factors such as the position of the linguistic components influence the interpretation of the kind of instructions under consideration here. Furthermore, the information conveyed by the visual object context also contributes significantly to the understanding of the instructions. And in some cases even information only on the context leads to a correct choice of the object of action and hence, to an adequate reference resolution.

Comparable results have been obtained for example by Spivey-Knowlton et al. (1998). Their examination of the processing of oral instructions showed an immediate influence of the visual context of objects as well as an important influence of the context of action and the experimental task the participants had to complete. As in our experiments, the authors used the conduction of an action as an indicator for the interpretation of the instructions. Such a procedure differs highly from the traditional ways of examining sentence processing. Here often only the linguistic reception of sentences is analyzed without or with only reduced (linguistic) contexts (e.g. Ferreira and Clifton 1986). Particularly for the processing of more complex instructions, but also for the processing of (syntactically) incomplete or underspecified and elliptical instructions – which typically occur in task-oriented communication –, we expect contextual information to become even more relevant, possibly vital for an adequate interpretation of an utterance.

4. Processing instructions in virtual reality

Having presented insights into the human side of instruction processing, we now want to switch sides and take on the machine's perspective. We present work on understanding instructions in a virtual reality construction task scenario, concentrating on the relevance of verb and object specificity and the temporal availability of information in natural language instructions. This is done under the perspective of reference resolution, i.e. the process of identifying the objects the instructions refer to. We will contrast some of the empirical results on humans with the prospects resulting from a computational approach and discuss how these results can be used to improve the naturalness of the speech understanding system.

In the following, we will concentrate on the description of the framework used for speech and gesture understanding. In doing so, we will emphasize

the reference resolution process in which the effects of verb and object specificity are simulated. Then we will draw a comparison between the empirical findings and the technical approach.

4.1. Speech and gesture understanding

The central module of our system for the understanding of multimodal instructions and direct manipulative actions in virtual reality is a tATN (Latoschik 2003). This is basically an ATN (Woods 1970) specialized for synchronizing multimodal input. As an ATN, it operates on a set of states and defines conditions for state transitions. The actual state represents the context of the utterance processed so far. Possible conditions classify words, gesture content, or test the context of the application. If a condition matches, the associated state becomes the actual state. The most prominent part of the context is the set of visual objects which is represented in the world model. Whenever information about visual objects is processed, the tATN queries a module called "reference resolution engine" (RRE) in order to verify the validity of the complex object descriptions specified so far and find the matching objects in the world model (Pfeiffer and Latoschik 2004). The set of possible interpretations of a complex object description delivered by the RRE is incrementally restricted by adding new constraints in the course of the processing of the utterance by the tATN. If the parsing process finally has been successful, the tATN initiates the execution of the instruction using the prominent entries in the result set.

4.2. Reference resolution

The task of the RRE is to interpret complex demonstrations according to the current world model represented in heterogeneous knowledge bases for symbolic information such as type, color, or function, and for geometrical information. This is done using a fuzzy-logic based constraint satisfaction approach. The tATN communicates with the RRE using a query language interface. After computing the query, the RRE returns a set of possible solutions, assigning entities in the world model to the specified variables, classified according to their relevance.

To parse the instruction *Nimm die rote Schraube!* (*Take the red bolt!*), the tATN would finally end up with a query as shown in Figure 11. It searches for a single entity matching the noun phrase *die rote Schraube* (*the red bolt*).

A query consists of variable definitions, e.g., `(inst ?x OBJECT)`, and a set of constraints `(has-color, is-a, has-type)`. The maintenance of temporal relations by the tATN is necessarily continued in the RRE. This is reflected by an additional parameter in the constraints associating each with a certain time during which the constraint is expected to hold. This may be not so important for the constraints over color or type used in the example, as they refer to static properties, but it will be for those constraints that refer to topological relations and arrangements of objects.


```
(inst ?x OBJECT)
   (has-color ?x RED time-1)
(inst ?x-type TYPE)
   (is-a ?x-type BOLT)
(has-type ?x ?x-type time-2)
```

Figure 11. The figure shows the constraint representation generated when processing the instruction *Take the red bolt!* The upper part depicts the constraint graph view on the textual constraints shown below.

Time is an important factor, as in the dynamic scenes of an immersive virtual environment most of the constraints can only be computed on demand. Especially, geometric constraints conveyed verbally, e.g. *Nimm die Schraube rechts vom Block!* (*Take the bolt to the right of the block!*), are computationally demanding. Even single constraints are highly ambiguous, and the fuzziness keeps adding up when several constraints are spanning over a set of variables. The RRE uses various techniques to overcome this problem: query refinement, hierarchical ordering, and incremental processing.

– *Query refinement:* The tATN only formulates queries made explicit in the utterance. In order to improve performance, the RRE refines these queries by adding constraints that define expectations or assumptions. The search space of potential reference objects, for example, is restricted to those of the toy kit by assuming `(inst ?x OBJECT)` for variables introduced by speech. In addition, the set of relevant objects is restricted to those that are located between the two interlocutors. Other constraints help in differentiating alternative solutions in the case of underspecification. So, for example, objects close to a participant (within reach of the hands) are pre-

ferred, or when connecting two objects, the pairing with minimal distance is preferred.

– *Hierarchical ordering of constraints:* Some constraints (like those that concern symbolic knowledge) are comparatively fast to compute; however, some others (like constraints on the arrangement of context objects), are computationally expensive. Some constraints are highly selective, singling out a small group of objects; some are fuzzy or too general for the context, as in the case of overspecification. In order to speed up computation, the RRE arranges the constraints in a hierarchical order, preferring faster, more selective constraints over more expensive, general ones.
– *Parameterization of the search process:* Occasionally, entities of an utterance, e.g. elements of a verbal expression, directly guide the further course of the search process. A frequently cited example is the handling of definite or indefinite articles. Experiments have shown that noun phrases with an indefinite article are processed faster than those with a definite article (Eikmeyer, Schade and Kupietz 1995). The behavior of the RRE can be changed accordingly. In the case of a definite article, the RRE can make an exhaustive search to ensure that the very best matching object is returned. When handling an indefinite article, the RRE can be requested to search for the first match rated over a specified threshold. In the worst case, this can take as long as in the case of a definite article, but on average it will be faster. – The example of the parameterization of the search process already shows that these features of the RRE do not only improve performance in terms of speed or resources; they can also be used to improve performance in terms of cognitive adequacy. Although the structures and processes used by the RRE are different from those of humans, constraints in time and capacity apply to both systems, human and machine, and the principles of coping with them might be similar.

5. Comparing instruction processing in humans and machines

Both systems, the human and the machine, have to deal with the same problem – the processing of assembly instructions in a specific context –; however, the technical premises on which the systems build are entirely different. Yet, in the end, by and large the same actions are taken on the side of human and artificial constructors. Ideally those are the actions, the instructor had had in mind – and this is, of course, the purpose the machine had been designed for in the first place. As both constructors are getting the same input and produce comparable actions, we are asking:

– How can we compare both approaches to instruction processing?
– How can results of such a comparison be interpreted?

Our idea is that we will get a deeper understanding of human instruction processing by looking at the problem from a different, a machine perspective.

5.1. Performance measurement

Before conducting an experiment and compare the two systems, we have to find an appropriate performance measure. Fortunately, for data on the human behavior we can resort to the results of the psycholinguistic experiments presented earlier in this chapter. The latencies recorded in these experiments provide a valid measure of the efforts required in the processing of instructions under varying conditions of linguistic structure and visual contexts.

Testing the performance of the machine by measuring plain processing times would be simple. However, the machine is much faster than the human, operating in the range of only a few milliseconds. This, together with the fact that the performance of the machine depends highly on the implementation and the hardware, renders this approach invalid for us.

Instead, as the reference resolution engine is based on a constraint-satisfaction technique, our measurement of its performance will use the number of constraint evaluations necessary when interpreting an instruction within a given context. This is a reliable and valid measure in that it is independent of the quality and efficiency of both the implementation and the hardware. However, it still depends on the way the knowledge of the world is modeled within the system.

In the following we will pick out representative examples of the items used in the experiments, look inside the reference resolution engine, and provide a detailed view of the constraint evaluations. For this, we will, on the one hand, make transparent the constraints created during the syntactic and semantic parsing of the instruction done by the tATN. On the other hand, we will show the progress of the RRE by stating the remaining variable assignments valid for a given context after the new constraints have been evaluated. In each step of the understanding process, the number of constraint evaluations depends on the number of variable assignments remaining after the preceding step and the number of constraints added in the current step. For our purposes each processing step can be marked by a single word or expression being actively processed.

5.2. Context effects

We start by investigating the effects of context on the performance of understanding simple noun phrases. This is a good starting point for introducing our notation.

5.2.1. The color

Experiment 1 shows, that instructions referring to color within an object context, in which the intended (target) object is in this regard identifiable unambiguously, are processed faster than in an ambiguous context. This holds at least for instructions with the verb in final position. As an example, the noun phrase *die gelbe Schraube* (*the yellow bolt*) is considered in an unambiguous and an ambiguous context (Fig. 12).

1. *die* (*the*): On encountering the article, the tATN requests the RRE to instantiate a new variable `?x` with the basic type `OBJECT` (see first cell in the "Query" column). The RRE creates the new variable and already tries to find possible assignments according to the current context. The results of this process are shown in the "Assignments" column. In this notation each assignment is represented by a tuple with a number of values corresponding to the number of variables – in our case this is a single value. In both contexts, there are initially four possible assignments for the new variable `?x` (the gender information in the German article is ignored).

2. *gelbe* (*yellow*): Verbal reference to a color is captured by the constraint **`has-color`**. The RRE evaluates this constraint for each assignment, resulting in a total of four constraint evaluations for each context. In the case of an unambiguous color context, the constraint only holds for one assignment. In the case of an ambiguous color context, three assignments pass the evaluation. Thus, verbal reference to the color was more restrictive in the unambiguous context than in the ambiguous one, as had been expected.

3. *Schraube* (*bolt*): As the type or the function of an object is of a different quality than its appearance, it is represented with an additional variable `?x-type`. This also reflects the heterogeneous design of our system, as different knowledge bases are involved for representing the visual information or the information regarding functions or types. The newly cre-

ated variable is then interlocked with the existing variable `?x` by the binary constraint **`has-type`**. Now the selectivity of the reference to color shows its consequences for the number of constraint evaluations. As for the unambiguous case the possible assignments have already been narrowed down to a single tuple, the **`has-type`** constraint has only to be evaluated once. In the ambiguous context, the constraint evaluation count is three.

<table>
<tr><td colspan="4">Interpretation of NPs in contexts with unambiguous or ambiguous colors</td></tr>
<tr><td>Context:</td><td>left: green-ball
blue-knob
yellow-bolt
bottom: red-cube

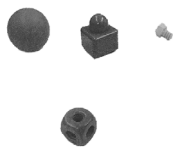
right: yellow-cube
yellow-brick
yellow-bolt
bottom: red-cube</td><td>  </td><td>  </td></tr>
<tr><td>Noun Phrase:</td><td colspan="3">die gelbe Schraube (the yellow bolt)</td></tr>
<tr><td>Surface</td><td>Query</td><td colspan="2">Assignments (?x), later (?x, ?x-type)</td></tr>
<tr><td>die
(the)</td><td><code>(inst ?x OBJECT)</code></td><td colspan="2">(green-ball)
(blue-knob)
(yellow-bolt)
(red-cube)
(yellow-cube)
(yellow-brick)
(yellow-bolt)
(red-cube)</td></tr>
<tr><td>gelbe
(yellow)</td><td><code>(has-color
?x YELLOW)</code></td><td>(yellow-bolt)</td><td>(yellow-cube)
(yellow-brick)
(yellow-bolt)</td></tr>
<tr><td>Schraube
(bolt)</td><td><code>(inst ?x-type TYPE)
(is-a ?x-type BOLT)
(has-type
?x ?x-type)</code></td><td>(yellow-bolt, BOLT)</td><td>(yellow-bolt, BOLT)</td></tr>
</table>

Figure 12. The upper part of the figure shows the context and the noun phrase. The lower part shows results of the speech understanding process: The "Surface" column shows the fragment of speech currently being processed, the "Query" column shows the built query, and the "Assignments" column shows the possible assignments (each in parentheses) as returned by the RRE. In the case of the context with an ambiguous color, more constraint evaluations have to be processed with the last query.

Comparing the results of the constraint evaluation, we may sum up that in the ambiguous context at least two more constraint evaluations have to be computed. This reflects the fact that referencing color is more discriminative in contexts with an unambiguous constellation of colors. In this respect, the processing in the RRE conforms to the observations of the experiment.

5.2.2. The functional context

The second contextual factor varied in the experiments was function. The results of Experiment 2 show that the influence of this factor depends on the position of the verb. When the verb is in front position, processing an instruction in a functionally ambiguous context takes longer than in an unambiguous context. In contrast, with the verb in final position, the difference between the reaction times is only very small but with a slight indication which makes it seem to be the other way round, namely processing the instruction in a functional ambiguous context being slightly faster than in the unambiguous context, which has also been shown in Experiment 1.

In Figure 13, two sentences, with the verb in front position and in final position respectively, are considered within two contexts. The figure shows that the contextual factor "function" causes a large difference when the verb is in front position, thus replicating the experimental findings by positing less constraint evaluations in the functionally unambiguous context than in the ambiguous one. In contrast, an investigation of instructions with the verb in final position yields no difference, at least with the reduced representation we use for purposes of demonstration (Fig. 13).

However, the constraint `(`**`connectable`** `?target ?reference)` is internally translated to:

- `(`**`inst`** `?target-type TYPE)`
- `(`**`has-type`** `?target ?target-type)`
- `(`**`inst`** `?reference-type TYPE)`
- `(`**`has-type`** `?reference ?reference-type)`
- `(`**`connectable`** `?target-type ?reference-type).`

This is done because the information of connectivity is part of the conceptual knowledge about types and functions. When instantiating the variables `?target-type` and `?reference-type` in the context with unambiguous functions, four different values (BALL, KNOB, BOLT, and CUBE) are initially assigned, in the ambiguous context there are only two (BOLT and

<table>
<tr><td colspan="4">Interpretation in unambiguous vs. ambiguous functional contexts</td></tr>
<tr><td>Context:</td><td>left: green-ball
blue-knob
yellow-bolt
bottom: red-cube

right: green-bolt
blue-bolt
yellow-bolt
bottom: red-cube</td><td>  </td><td>  </td></tr>
<tr><td>Sentence:</td><td colspan="3">Mit dem roten Teil das gelbe Teil verbinden!
(With the red part the yellow part is to be connected!)</td></tr>
<tr><td>Surface</td><td>Query</td><td>Assignments (?target, ?reference)</td><td></td></tr>
<tr><td>Mit dem roten Teil</td><td>(has-color
?reference
RED)</td><td>(green-ball, red-cube)
(blue-knob, red-cube)
(yellow-bolt, red-cube)</td><td>(green-bolt, red-cube)
(blue-bolt, red-cube)
(yellow-bolt, red-cube)</td></tr>
<tr><td>das gelbe Teil</td><td>(has-color
?target
YELLOW</td><td>(yellow-bolt, red-cube)</td><td>(yellow-bolt, red-cube)</td></tr>
<tr><td>verbin-den</td><td>(connectable
?target
?reference)</td><td>(yellow-bolt, red-cube)</td><td>(yellow-bolt, red-cube)</td></tr>
<tr><td>Sentence:</td><td colspan="3">Verbinde mit dem roten Teil das gelbe Teil!
(Connect with the red part the yellow part!)</td></tr>
<tr><td>Surface</td><td>Query</td><td>Assignments (?target, ?reference)</td><td></td></tr>
<tr><td>Verbinde</td><td>(connectable
?target
?reference)</td><td>(red-cube, yellow-bolt)
(yellow-bolt, red-cube)</td><td>(red-cube, green-bolt)
(red-cube, blue-bolt)
(red-cube, yellow-bolt)
(green-bolt, red-cube)
(blue-bolt, red-cube)
(yellow-bolt, red-cube)</td></tr>
<tr><td>mit dem roten Teil</td><td>(has-color
?reference
RED)</td><td>(yellow-bolt, red-cube)</td><td>(green-bolt, red-cube)
(blue-bolt, red-cube)
(yellow-bolt, red-cube)</td></tr>
<tr><td>das gelbe Teil</td><td>(has-color
?target
YELLOW</td><td>(yellow-bolt, red-cube)</td><td>(yellow-bolt, red-cube)</td></tr>
</table>

Figure 13. Two complete instructions with the verb in front respectively final position are processed. In the latter case, an ambiguous functional context leads to the processing of the most constraint evaluations.

CUBE). As the possible assignments for `?target` and `?reference` are already restricted to a single tuple, the set of assignments for both type variables are restricted to one as soon as the corresponding **`has-type`** constraints are processed. Thus, when finally processing the computationally demanding **`connectable`** constraint, the same number of constraint evaluations has to be processed in both cases. This leaves the small overhead of two evaluations of the **`has-type`** constraint (which evaluates very fast) for the unambiguous context.

5.3. Effects based on differences in the linguistic material

5.3.1. Specificity of object naming

When a variable for an object is defined by `(`**`inst`**` ?x OBJECT)`, the initial set of possible assignments for `?x` is the set of currently visible objects. This restriction reflects the fact that the instructions in our scenario are all about manipulating objects that are currently available. In the case of a specific object naming, as in *die Schraube* (*the bolt*), the variable for the visual object is tied to a newly created variable for the type: `(`**`inst`**` ?x-type TYPE) (`**`has-type`**` ?x ?x-type) (`**`is-a`**` ?x-type BOLT)`. This is different with an unspecific object naming, as in *das Teil* (*the part*). Here, the noun does not add any further type information, so the variable for the visual object is not linked with a new type variable. Therefore, fewer constraints have to be evaluated when unspecific object names are to be processed – a difference which closely corresponds to the findings from the first experiment.

However, this may only hold for sentences in which the specific variable is fully specified at the time the noun is processed, for example by some preceding reference to its unambiguous color or function. Under such circumstances, adding a specific object naming would lead to an overspecification and impose an additional processing overhead. This is not the case when dealing with underspecifications. Here, the restrictive power of a specific object naming could substantially reduce the number of possible assignments. In that case, we have a tradeoff between the additional constraint evaluations needed to select the visual objects of a given type and the constraint evaluations needed in the subsequent processing steps, which may now operate on a smaller set of remaining possible assignments. Overall, one would expect an advantage for a specific naming, unless there already is an overspecification.

5.3.2. Specificity of prepositions

So far, the observations of the RRE neatly match the data from the experiments. This also holds for the result that instructions making use of a specific preposition, such as *in*, are processed more slowly than those with an unspecific one. However, the dependence of this effect on verb position and the order of the arguments (RO–TO vs. TO–RO) is not replicated, as we will show now.

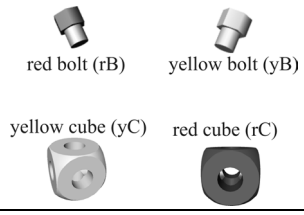
<table>
<tr><td colspan="5">Specific vs. unspecific prepositions with verb in front position</td></tr>
<tr><td>Context:</td><td colspan="4"> 
</td></tr>
<tr><td>Sentence:</td><td colspan="4">Füge ein rotes Teil in/an ein gelbes Teil!
(Put a red part in/on a yellow part!)</td></tr>
<tr><td>Surface</td><td colspan="2">Query</td><td colspan="2">Assignments (?target, ?reference)</td></tr>
<tr><td>Füge</td><td colspan="2">(connectable
?target ?reference)</td><td colspan="2">(rB, yC) (rB, rC) (yB, yC) (yB, rC)
(yC, rB) (yC, yB) (rC, rB) (rC, yB)</td></tr>
<tr><td>ein rotes Teil</td><td colspan="2">(has-color
?target RED)</td><td colspan="2">(rB, rC) (rB, yC) (rC, rB) (rC, yB)</td></tr>
<tr><td></td><td>in</td><td>an</td><td>in</td><td>an</td></tr>
<tr><td>in/an</td><td>(has-port
?target 'MALE)
(has-port
?reference 'FEMALE)</td><td>--</td><td>(rB, rC)
(rB, yC)</td><td>(rB, rC) (rB, yC)
(rC, rB) (rC, yB)</td></tr>
<tr><td>ein gelbes Teil</td><td colspan="2">(has-color
?reference YELLOW)</td><td>(rB, rC)</td><td>(rB, yC) (rC, yB)</td></tr>
</table>

Figure 14. The preposition *in* is more specific than *an* (*on*) adding two additional constraints. This leads to a successful reference with a single solution; not so for the preposition *an*, as it adds no further constraints. The result is ambiguous and either a pragmatic arbitrary choice or a clarifying question has to follow.

Before we present proper examples, we have to explain the mapping of the prepositions to the constraints. For an unspecific preposition this is easy as no further constraints need to be added. But in the case of a specific preposition, the following constraints are added: `(`**`has-port`** `?target 'MALE)`

(**has-port** ?reference 'FEMALE). Ports are used in the knowledge bases to mark areas where objects can be connected (Jung 2003). There are several possible types of ports; in our case we are faced with screws ports. For these ports, two different subtypes exist, 'male' and 'female': A "baufix" bolt typically has one male port and a cube six female ports. The intended direction of the connection is then reflected by specifying the ports needed to accomplish the connection suggested by the preposition.

Figure 14 gives an example for an instruction with the verb in front position and both a specific and an unspecific preposition. After processing the instruction with the specific preposition *in*, the set of possible assignments is narrowed down to a single value. The preposition helps to select the intended order of the objects (bolt into cube). In the unspecific case, the preposition *an* (*on*) does not add any further constraints. The instruction is underspecified and two different assignments remain. In Experiment 3 the subject then had to choose one of the assignments arbitrarily. This underspecification goes along with a smaller number of constraint evaluations. This holds for all the variants of the instructions. Constraint values for different instruction variants are shown in Table 3.

Table 3. Constraint evaluations for specific vs. unspecific prepositions

Condition		Preposition	Constraint Evaluations (CE)				Total
TO–RO	verb front		Verb	NP	Prep	NP	
		- *in*	12	8	4 + 2	2	28
		- *an*	12	8	0	4	24
	verb mid		NP	Verb	Prep	NP	
		- *in*	12	6	4 + 2	2	26
		- *an*	12	6	0	4	22
	verb final		NP	Prep	NP	Verb	
		- *in*	12	6 + 3	2	1	24
		- *an*	12	0	6	4	22
RO–TO	verb front		Verb	Prep	NP	NP	
		- *in*	12	8 + 4	4	2	30
		- *an*	12	0	8	4	24
	verb mid		Prep	NP	Verb	NP	
		- *in*	12 + 6	4	2	2	26
		- *an*	0	12	6	4	22
	verb final		Prep	NP	NP	Verb	
		- *in*	12 + 6	4	2	1	25
		- *an*	0	12	6	4	22

5.3.3. Sequence of arguments

Below, we shall discuss examples that differ from the examples given above with respect to verb position and the sequence of arguments (see Figures 15 and 16).

<table>
<tr><td colspan="5">Specific vs. unspecific prepositions with verb in final position and order TO–RO</td></tr>
<tr><td>Context:</td><td colspan="4">For a picture of the context, see Figure 14</td></tr>
<tr><td>Sentence:</td><td colspan="4">Ein rotes Teil in/an ein gelbes Teil fügen!
(A red part in/on a yellow part is to be put!)</td></tr>
<tr><td>Surface</td><td colspan="2">Query</td><td colspan="2">Assignments (?target, ?reference)</td></tr>
<tr><td>Ein rotes Teil</td><td colspan="2">(has-color
?target RED)</td><td colspan="2">(rB, yB) (rB, yC) (rB, rC) (rC, rB)
(rC, yB) (rC, yC)</td></tr>
<tr><td></td><td>in</td><td>an</td><td>in</td><td>an</td></tr>
<tr><td>in/an</td><td>(has-port
?target 'MALE)
(has-port
?reference 'FEMALE)</td><td>--</td><td>(rB, rC)
(rB, yC)</td><td>(rB, yB) (rB, yC)
(rB, rC) (rC, rB)
(rC, yB) (rC, yC)</td></tr>
<tr><td>ein gelbes Teil</td><td colspan="2">(has-color
?reference YELLOW)</td><td>(rB, yC)</td><td>(rB, yB) (rB, yC)
(rC, yB) (rC, yC)</td></tr>
<tr><td>fügen</td><td colspan="2">(connectable
?target ?reference)</td><td>(rB, yC)</td><td>(rB, yC) (rC, yB)</td></tr>
</table>

Figure 15. In the condition "verb in final position and specific preposition *in*", the correct assignment for ?target can be established as soon as the preposition is fully processed. In contrast, the instruction with the unspecific preposition is underspecified and two alternative assignments for ?target remain. These two assignments emerge when processing the first NP after 12 constraint evaluations. Thus, although the specific preposition allows a full specification, the assignment for ?target is only established after the preposition is processed.

The specific preposition increases the number of constraint evaluations in all cases, singling out one specific assignment. The unspecific preposition leads to an underspecification with two alternative assignments while evaluating fewer constraints. This replicates the results from Experiment 3 that instructions with unspecif ic prepositions are processed faster than those with specific ones. However, regarding the interaction with the position of the verb, the data from the RRE suggest that the computational effort for processing an instruction decreases the later the verb is positioned. In the experiments this only holds for the RO–TO argument order. Also, differences between

specific and unspecific prepositions decrease the later the verb is positioned in the instruction. This runs contrary to the experimental results in which the instructions with the verb in mid or final position show a significant difference, whereas with a front position both variants yield quite similar reaction times.

<table>
<tr><td colspan="5">Specific vs. unspecific prepositions with verb in final position and order RO–RO</td></tr>
<tr><td>Context:</td><td colspan="4">For a picture of the context, see Figure 14</td></tr>
<tr><td>Sentence:</td><td colspan="4">In/an ein gelbes Teil ein rotes Teil fügen!
(In/on a yellow part a red part is to be put!)</td></tr>
<tr><td>Surface</td><td colspan="2">Query</td><td colspan="2">Assignments (?target, ?reference)</td></tr>
<tr><td></td><td>in</td><td>an</td><td>in</td><td>an</td></tr>
<tr><td>In/an</td><td>(has-port
?target 'MALE)
(has-port
?reference 'FEMALE)</td><td>--</td><td>(rB, rC)
(rB, yC)
(yB, rC)
(yB, yC)</td><td>(rB, yB) (rB, yC)
(rB, rC) (rC, rB)
(rC, yB) (rC, yC)
(yB, rB) (yB, yC)
(yB, rC) (yC, rB)
(yC, yB) (yC, rC)</td></tr>
<tr><td>ein gelbes Teil</td><td colspan="2">(has-color
?reference YELLOW)</td><td>(rB, yC)
(yB, yC)</td><td>(rB, yB) (rB, yC)
(rC, yB) (rC, yC)
(yB, yC) (yC, yB)</td></tr>
<tr><td>ein rotes Teil</td><td colspan="2">(has-color
?target RED)</td><td>(rB, yC)</td><td>(rB, yB) (rB, yC)
(rC, yB) (rC, yC)</td></tr>
<tr><td>fügen</td><td colspan="2">(connectable
?target ?reference)</td><td>(rB, yC)</td><td>(rB, yC) (rC, yB)</td></tr>
</table>

Figure 16. The constraint evaluations for instructions with the alternative order of arguments differ only slightly from those shown in Figure 15.

The tendency for the argument order RO–TO to lead to faster reaction times than the order TO–RO is also not replicated by the study of the RRE (see Fig. 15 and 16, also Tab. 3). While both variants do not differ in the number of constraint evaluations when processing instructions with unspecific prepositions, the average number of constraint evaluations for instructions with specific prepositions is slightly higher for RO–TO.

5.4. Discussion

We compared the processing of instructions in humans with a computer science approach based on fuzzy constraint satisfaction. For this, we used the

number of constraint evaluations as a measurement comparable to the reaction times collected in the psycholinguistic experiments described above. The applicability of this approach was then demonstrated on representative examples taken from the experiments.

We started our investigations with a focus on the effects of the contextual influences of color and function. Then we shifted our attention to local semantic differences, investigating the effects of specificity in naming. Finally, we had a look at the effects of the syntactic order of both arguments and verbs. In the chapter in hand, we intentionally skipped a comparison regarding the effect of verb specificity, because the presentation of the necessary knowledge structures involved would have gone beyond the scope of this chapter.

We consequently drew a line from factors regarding external visual context to factors of a structural linguistic kind. In our setting, the structure of the visual context defines the complexity of the reference problem. While it is true that the structure of the linguistic material reflects this complexity, it is mainly influenced by the interlocutors' knowledge of language use, conceptual world knowledge, and internal processes operating on that knowledge.

As both systems are able to solve the reference problem, their performance shows comparable effects with respect to changes in the complexity of the problem domain, i.e. the visual context. In contrast, the results concerning the effects of changes in the linguistic material show that the interpretation of instructions by the machine does not scale up well enough to match the human's performance. Though its capabilities already meet the pragmatic requirements of a human-computer interface, the performance is not cognitively adequate.

We are quite aware of the fact that these results might be artifacts of the measurement of the performance of the machine's processing capabilities by counting constraint evaluations. This is a linear measurement which is implicitly based on the assumption that the constraints are evaluated in a sequential fashion. The reaction time of a system always is an abstraction from the way of processing, be it sequential or parallel. Counting single evaluations also assumes that the time needed to evaluate a constraint is a constant. It has already been shown for the `connectable` constraint that its complexity matters. Some constraints pertain to easily accessible properties, such as color, which can be thought of as computing in linear time. Other constraints depend on the power of the set of contextual objects. Examples (not addressed in the present study) are constraints over relative attributes, such as size or position. The categorization of properties, e.g., when processing a

specific naming of type or function, may also depend on the complexity and the structuring of the world knowledge. A more precise measurement would incorporate all these factors.

6. Conclusion

This chapter presented a closer look at instruction processing in the context of a construction task domain. Instructions can be assigned to the concept of requests in speech act theory. They can also be categorized psycholinguistically according to the system AUFF. Linguistic aspects relevant for processing the special kind of instructions in our scenario are the specificity of verbs, object naming, and prepositions and the sequence of components in the linguistic surface structure. The results of our experiments also show that the interpretation of instructions is not exclusively determined by linguistic information but also by non-verbal information pertaining to the visual context. These findings are in line with recent approaches to sentence processing, which assume that linguistic and non-linguistic information is processed in an interactive and incremental way (Ferstl and Flores d'Arcais 1999). In order to resolve object references and to get to know which action has to be conducted, any adequate information available in the actual communicative situation is pulled up immediately.

In the section on a multimodal human-computer interface for a virtual constructor, we took a computer scientist's perspective and gave an example of a speech understanding system for virtual reality environments. As the syntactic structure of the instructions under consideration is relatively simple, we concentrated on the semantic-pragmatic processing. This is realized in the reference resolution engine, which is responsible for the identification of the objects of action. The solution presented combines constraint-satisfaction algorithms with fuzzy logic in order to approach the problem of vagueness in natural speech.

Comparing the performances of both systems, human and machine, we have found that the performance depends partly on the structure of the problem domain and partly on the structure of the conceptual knowledge and the processes working thereon. While our measurement grasps the content and the order in which information is available to the processing system, it cannot disambiguate effects of parallel processing or capture the complexity of the processes needed to de-reference each chunk of information. It also neglects side effects and interactions of sub-processes, which could explain, for

instance, the effects observed in the interaction of a specific naming of the reference object and the target object.

In both the psycholinguistic experiments and the computer simulation it becomes obvious that the information gathered using reaction time measurement does not yield enough information to get a deeper insight into the timing and interaction of the sub-processes relevant for the comprehension of instructions. In order to model these effects in the computer simulation, more data on human performance are necessary. Further experiments, making use of eye movement tracking or electroencephalography, could provide the necessary empirical basis for an improvement of the human-computer interface and help in making it more cognitively adequate.

The work presented here concentrated on the processing of basic single sentences by the constructor. Work is on the way to extend the studies to the investigation of more complex instructions such as *Nimm die rote Schraube und stecke sie von oben in den grünen Würfel* (*Take the red bolt and put it from above in the green cube*), or instruction sequences such as *Stecke die rote Schraube in den grünen Würfel und die gelbe in den roten* (*Put the red bolt in the green cube and the yellow [one] in the red [one]*) (Weiß, Pfeiffer, and Allmaier 2004).

References

Altmann, G. T. M., and Y. Kamide
1999 Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73: 247–264.

Austin, J. L.
1962 *How to do Things with Words*. Oxford: Clarendon Press.

Blum-Kulka, S.
1987 Indirectness and politeness in requests: Same or different? *Journal of Pragmatics* 11: 145–160.

Blum-Kulka, S., J. House, and G. Kasper (eds.)
1989 *Cross-Cultural Pragmatics: Requests and Apologies*. Norwood: Ablex.

Brandt-Pook, H.
1999 *Eine Sprachverstehenskomponente in einem Konstruktionsszenario*. Dissertation. Bielefeld: Universität Bielefeld.

Britt, M. A.
1994 The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language* 33: 251–283.

Brown, P., and S. C. Levinson
2004 *Politeness: Some Universals in Language Usage*. Cambridge, UK: Cambridge University Press.
Carroll, M., and C. Timm
2003 Erzählen, Berichten, Instruieren. In *Sprachproduktion*, T. Herrmann and J. Grabowski (eds.), 687–712. Göttingen: Hogrefe.
Chambers, C. G., M. K. Tanenhaus, K. M. Eberhard, H. Filip, and G. N. Carlson
2002 Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language* 47: 30–49.
Clark, H. H., and S. E. Haviland
1977 Comprehension and the given-new contract, In *Discourse Production and Comprehension,* R. O. Freedle (ed.), 1–40. Norwood: Ablex.
Crain, S., and M. Steedman
1985 On not being led up the garden path: The use of context by the psychological syntax processor. In *Natural Language Parsing,* D. R. Dowty, L. Karttunen, and A. M. Zwicky (eds.), 320–358. Cambridge, UK: Cambridge University Press.
Eikmeyer, H.-J., U. Schade, and M. Kupietz
1995 Ein konnektionistisches Modell für die Produktion von Objektbenennungen. *Kognitionswissenschaft* 4: 108–117.
Engelkamp, J., and G. Mohr
1986 Legitimation und Bereitschaft bei der Rezeption von Aufforderungen. *Sprache & Kognition* 2: 127–139.
Fellbaum, C.
1998 A semantic network of English verbs. In *WordNet: An Electronic Lexical Database,* C. Fellbaum (ed.), 69–104. Cambridge, MA: MIT Press.
Ferreira, F., and C. Clifton
1986 The independence of syntactic processing. *Journal of Memory and Language* 25: 348–368.
Ferstl, E., and G. Flores d'Arcais
1999 Das Lesen von Wörtern und Sätzen. In *Sprachrezeption*, A. D. Friederici (ed.), 203–242. Göttingen: Hogrefe.
Gernsbacher, M. A.
1991 Cognitive processes and mechanisms in language comprehension: The structure building framework. In *The Psychology of Learning and Motivation 27*, G. H. Bower (ed.), 217–263. San Diego: Academic Press.
Goffman, E.
1989 *Interaction Ritual – Essays on Face-to-Face-Behavior*. New York: Pantheon Books.
Grabowski-Gellert, J., and P. Winterhoff-Spurk
1988 Your smile is my command: Interaction between verbal and nonverbal components of requesting specific to situational characteristics. *Journal of Language and Social Psychology* 7: 229–242.

Graf, R., and K. Schweizer
2003 Auffordern. In *Psycholinguistik: Ein internationales Handbuch,* G. Rickheit, T. Herrmann, and W. Deutsch (eds.), 432–442. Berlin: de Gruyter.

Harris, L. J.
1975 Spatial direction and grammatical form of instructions affect the solution of spatial problems. *Memory and Cognition* 3: 329–334.

Herrmann, T.
1983 *Speech and Situation. A Psychological Conception of Situated Speaking*. Berlin: Springer.
2003 Auffordern. In *Sprachproduktion*, T. Herrmann and J. Grabowski (eds.), 713–732. Göttingen: Hogrefe.

Herrmann, T., and J. Grabowski
1994 *Sprechen: Psychologie der Sprachproduktion*. Heidelberg: Spektrum.

Hildebrandt, B., H.-J. Eikmeyer, G. Rickheit, and P. Weiß
1999 Inkrementelle Sprachrezeption. In *KogWis99: Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft,* I. Wachsmuth and B. Jung (eds.), 19–24. Sankt Augustin: infix.

Hindelang, G.
1978 *Auffordern: Die Untertypen des Aufforderns und ihre sprachlichen Realisierungsformen*. Göppingen: Kümmerle.

Hörnig, R., K. Oberauer, and A. Weidenfeld
2002 Räumliches Schließen als Sprachverstehen. *Kognitionswissenschaft* 9: 185–192.

Hoppe-Graff, S., T. Herrmann, P. Winterhoff-Spurk, and R. Mangold
1985 Speech and situation: A general model for the process of speech production. In *Language and Social Situations,* J. P. Forgas (ed.), 81–95. Heidelberg: Springer.

Huttenlocher, J., and S. Strauss
1968 Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Verbal Behavior* 7: 300–307.

Jung, B.
2003 Task-level assembly modeling in virtual environments. In *Computational Science and its Applications - ICCSA 2003, Proceedings,* V. Kumar, M. L. Gavrilova, C. J. K. Tan, and P. L'Ecuyer (eds.), 721–730). Springer.

Kopp, S., B. Jung, N. Leßmann, and I. Wachsmuth
2003 Max – A multimodal assistant in virtual reality construction. *KI – Künstliche Intelligenz* 4: 11–17.

Latoschik, M. E.
2003 Designing transition networks for multimodal VR-interactions using a markup language. In *Proceedings of the IEEE Fourth International Conference on Multimodal Interfaces, ICMI 2002,* 411–416. Pittsburgh, USA: IEEE Computer Society.

Mangold, R.
1987 Schweigen kann Gold sein – über förderliche, aber auch nachteilige Effekte der Überspezifizierung. *Sprache und Kognition* 4: 165–176.
Meyer, J. R.
1992 Fluency in the production of requests: Effects of degree of imposition, schematicity and instruction set. *Journal of Language and Social Psychology* 11: 233–251.
Miller, G. A.
1991 *The Science of Words*. New York: Scientific American.
1998 Nouns in WordNet. In *WordNet: An Electronic Lexical Database,* C. Fellbaum (ed.), 23–46. Cambridge, MA: MIT Press.
Miller, G. A., and C. Fellbaum
1991 Semantic networks of English. *Cognition* 41: 197–229.
Oberauer, K., and O. Wilhelm
2000 Effects of directionality in deductive reasoning: I. The comprehension of single relational premises. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 1702–1712.
Olsen, S.
1996 Pleonastische Direktionale. In *Wenn die Semantik arbeitet: Klaus Baumgärtner zum 65. Geburtstag* G. Harras and M. Bierwisch (eds.), 303–329. Tübingen: Niemeyer.
Pfeiffer, T., and M. E. Latoschik
2004 Resolving object references in multimodal dialogues for immersive virtual environments. In *Proceedings of the IEEE Virtual Reality 2004* Y. Ikei, M. Göbel, and J. Chen (eds.), 35–42. Chicago: IEEE Computer Society.
Rolf, E.
1997 *Illokutionäre Kräfte: Grundbegriffe der Illokutionslogik*. Opladen: Westdeutscher Verlag.
Rosch, E.
1978 Principles of categorization. In *Cognition and Categorization,* E. Rosch and B. B. Lloyd (eds.), 27–48. Hillsdale: Erlbaum.
Searle, J. R.
1969 *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press.
1976 A classification of illocutionary acts. *Language in Society* 5: 1–23.
Spivey-Knowlton, M. J., M. K. Tanenhaus, K. M. Eberhard, and J. C. Sedivy
1998 Integration of visuospatial and linguistic information: Language comprehension in real time and real space. In *Representation and Processing of Spatial Expressions,* P. Olivier and K.-P. Gapp (eds.), 201–214. Mahwah: Erlbaum.
Tanenhaus, M. K., M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy
1995 Integration of visual and linguistic information in spoken language comprehension. *Science* 268: 1632–1634.

Trueswell, J. C., and M. K. Tanenhaus
1994 Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In *Perspectives on Sentence Processing* C. Clifton, L. Frazier, and K. Rayner (eds.), 155–179. Hillsdale, NJ: Erlbaum.

Weiß, P.
2001 "Schraub' in" oder "schraub' an"? Präpositionen bei der Verarbeitung von Handlungsanweisungen. In *Sprache, Sinn und Situation,* L. Sichelschmidt and H. Strohner (eds.), 75–89. Wiesbaden: Deutscher Universitäts-Verlag.
2005 *Raumrelationen und Objekt-Regionen: Psycholinguistische Überlegungen zur Bildung lokalisationsspezifischer Teilräume*. Wiesbaden: Deutscher Universitäts-Verlag.

Weiß, P., and S. Barattelli
2003 Das Benennen von Objekten. In *Sprachproduktion,* T. Herrmann and J. Grabowski (eds.), 587–621. Göttingen: Hogrefe.

Weiß, P., B. Hildebrandt, H.-J. Eikmeyer, and G. Rickheit
1999 Verb-, Objekt- und Kontextinformation bei der Rezeption von Handlungsanweisungen. In *KogWis99: Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft,* I. Wachsmuth & B. Jung (eds.), 238–243. Sankt Augustin: infix.

Weiß, P., B. Hildebrandt, and G. Rickheit
1999 Empirische Untersuchungen zur Rezeption von Handlungsanweisungen: der Einfluß semantischer und kontextueller Faktoren. *Sprache und Kognition*, 18: 39–52.

Weiß, P., K. Kessler, B. Hildebrandt, and H.-J. Eikmeyer
1999 Konzeptualisierung in inkrementell-integrativer Sprachverarbeitung. *Kognitionswissenschaft* 8: 108–114.

Weiß, P., and R. Mangold
1997 Bunt gemeint, doch farblos gesagt: Wann wird die Farbe eines Objektes nicht benannt? *Sprache und Kognition* 16: 31–47.

Weiß, P., T. Pfeiffer, and K. Allmaier
2004 Blickbewegungsmessungen bei der Verarbeitung elliptischer Konstruktionsanweisungen in situierter Kommunikation. In *44. Kongress der Deutschen Gesellschaft für Psychologie,* T. Rammsayer, S. Grabianowski, and S. Troche (eds.), 307. Lengerich: Pabst.

Woods, W.
1970 Transition network grammars for natural language analysis. *Communications of the ACM* 13: 591–606.

Wunderlich, D.
1984 Was sind Aufforderungssätze? In *Pragmatik in der Grammatik: Jahrbuch des Instituts für deutsche Sprache 1983,* G. Stickel (ed.), 92–117). Düsseldorf: Schwann.