How to make a wheelchair understand spoken commands

.

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

unterlegt von Daniel Couto Vale

Universität Bremen

Bremen, den 31. Juli 2018

Contents

Chapter 1

Introduction

The community of people who suffer from various degrees of gait impairment accounted for 867,029 individuals among the 80.5 million inhabitants of Germany at the end of 2011 (Statistisches Bundesamt, 2013), about 1.1% of the entire population. All of them are likely to be or become wheelchair users and we can assume they are a potential market for speech-controlled intelligent wheelchairs for two reasons, one legal and the other economical.

According to the German constitution, no human can be underprivileged due to his or her disability (Deutsche Verfassung, Art 3.3) and this constitutional law is currently interpreted as an equality of opportunity supported by corresponding compensatory assistance for individuals' disabilities and/or congenital, social, and physical disadvantages. To implement this, laws have been passed to regulate illness, accident, and assistance insurances coverage to include access to wheelchairs and to oblige all inhabitants of Germany and tourists in Germany to be insured (Krankenversicherung, Unfallversicherung, Pflegeversicherung). Hence, every inhabitant and tourist must be insured and, if so, is guaranteed access to either a wheelchair or appropriate assistance to compensate for their gait impairment both in the cases where they were born with or have acquired this impairment through illness, accident, or ageing on German soil.

As for the market, a major wheelchair distribution channel in Germany is likely to favour speech-controlled wheelchairs. This channel is mediated by assisted living institutes funded by insurances and does not include ownership by the end user. In apartment buildings belonging to such institutes, where care taking is provided to residents, such a technological addition might reduce overall costs for two reasons. First, the costs of additional technology can be amortised over the years among various residents who will live in the same apartment utilising the same wheelchair successively, lowering the cost per year of usage. Second, this technology might reduce the need for paid human labour in mobility-related assistance, that is likely to be more expensive per year than the amortised additional cost of speech-controlled autonomous navigation. Therefore, from the assistance business perspective, including speech control functionality in wheelchairs in Germany can be justified based on the collective financial savings that this technology would bring to assisted living institutes with the automation of human labour in mobility-related assistance.

However, understanding users' utterances such as *ich möchte mir die Hände*

waschen (*I want to wash my hands*) as commands for the wheelchair to take the user somewhere, is a non-trivial task. Unsurprisingly, no command understanding modules for wheelchairs currently exist that can reliably understand which service users are demanding from the wheelchair. The reason is simple: the process of recognising the user's intent regarding his or her utterance is highly dependent on the contexts of situation and discourse, a well-known problem in situated dialogues due to their agentive nature and dependencies on physical context (Ross and Bateman, 2009).

In this thesis, I propose an automation of the understanding of spoken commands for wheelchairs and I shall test it in an assisted living institute apartment. In particular, I shall focus on answering two issues:

1) How to recognise the user's intent regarding his or her utterance relying on the situation the interactants are in and the unfolding discourse.

2) How to create a language-based taxonomy of simple things[1], locations and processes that can be integrated into a rule-based understanding module.

In the following section, I list the challenging linguistic phenomena for automatic understanding that this research will contend with, following which I present the intended contributions to collective knowledge regarding human languages, subsequently describing how the scientific thesis shall be proposed and verified, finally concluding with an outline explaining the contents of each chapter.

1.1 Linguistic Phenomena

To understand spoken commands, a wheelchair needs to contend with various linguistic phenomena, which can be divided into five groups according to their characteristics: speech-and-text, reference, configurations, logical nexuses, and dialogue moves. I describe and exemplify these linguistic phenomena below individually.

1.1.1 Related to speech and text

Automatic speech recognition is the task of automatically representing speech as a stream of characters (Nuance VoCon 3200). Usually, speech is first represented as a stream of phonetic letters from IPA or another proprietary phonetic alphabet. Then it is segmented and replaced by standard spelling in national alphabets, syllabaries, and/or ideogram sets. Most speech recognisers depend on linguistic models for accurate recognition. Since German is a language with a relatively good correspondence between phonetic and standard spelling, speech recognition tends to be accurate for language models based on standard spelling and conversion rules from standard to phonetic spelling (Couto-Vale and Mast, 2012). Accuracy tends to be better for models anchored on rhythmic structures than for those anchored on grammatical constituency alone (Couto-Vale and Mast, 2012).

However, when a wheelchair recognises a user's speech, it recognises the exact sequence of sounds uttered, which may include disfluencies (Halliday, 1987) and repairs (Schegloff et al., 1977; Schegloff, 1979, 1992b, 2000b, 2002b). In other

[1] Simple things include not only regular objects we find in an apartment but also people, furniture, rooms, apartments, buildings and any other thing we can refer to during interaction.

words, **transcribed speech** is not the same as **spoken text**. For instance, let's consider the transcribed speech in Example 1. Two slashes indicate the initial or final boundary of a **speech curve**, a melodic group of phonemes, and one slash indicates a boundary between **speech feet**, a rhythmic group of phonemes.

(1) // ich möchte zum / zur Küche / fahren //
// I want / to go to the / to the kitchen //

When a user produces the sounds in Example 1, he or she intends to replace the word *zum* in the end of the first foot by the word *zur* in the beginning of the second. The corresponding spoken text is the wording in Example 2. Double pipes indicate the initial or final boundary of a **clause** and a single pipe indicates the boundary between **clause constituents**.

(2) || ich | möchte | zur Küche | fahren ||
|| I | want | to go | to the kitchen ||

For a wheelchair to understand which text was spoken by the user, it cannot simply assume that transcribed speech is spoken text. If it were to do this, linguistic analysis would need to account for "ungrammatical texts" such as the transcribed speech in Example 1, adding more complexity to an already difficult task.

Disfluencies are a simple case of repair phenomena. For long-distance repairs, a wheelchair cannot assume that any segment of spoken text is in its final version until the user stops repairing what he or she has spoken thus far. For instance, the transcribed speech in Example 3 should correspond to the spoken text in Example 4.

(3) // ich möchte / zur Tür / fahren // zur Wohnungstür //
// I want / to go / to the door // to the apartment door //

(4) || ich | möchte | zur Wohnungstür | fahren ||
|| I | want | to go | to the apartment door ||

If the wheelchair fails to overcome disfluencies and integrate repairs, it will need to adopt graceful failure strategies such as asking the user to repeat utterances that any human would have understood. In this thesis, I propose a module that consumes transcribed speech and produces the corresponding grammatical text (see Chapter 8).

1.1.2 Related to reference

In computational linguistics, there is a large body of study in reference resolution (Abbott, 2010) and, with the advent of home automation and autonomous devices, we see a growing interest in the area of situated reference resolution (Mast et al., 2012; Mast and Wolter, 2013b,a; Mast et al., 2014a,b). The major difference between reference to absent things and reference to objects or people known to be around us is the way referential grounding works. Resolving a reference to an absent thing depends on two factors: whether the referent is constructed by the uttered words themselves such as some food one wants to

prepare or whether the referent is known to exist. In the latter case, if identifying the referent is relevant for moving on with the dialogue, the referrer needs to provide enough discursive context to establish a frame of reference — where the object is to be found — and enough restrictions to single out a thing in that frame. In contrast, resolving a reference to a present thing depends on the reference resolver's ability to use the context of situation and referential restrictions to establish the adequate frame of reference for singling out the referent (Zender et al., 2009). Due to this difference, many simplifications that work for dialogue systems supported by knowledge bases such as street, stock, movie, and song databases do not work for situated dialogue as we shall see next. In the following, I describe the main referential phenomena that fall into this category.

When a wheelchair user refers to something, the wheelchair must reliably identify what the user is referring to independently of how they are typically referred to in German. For instance, parts of humans and parts of rooms are typically referred to in different ways in German as in Examples 5 and 6, so wheelchairs need to anticipate wordings with different structures for different kinds of things.

(5) ich möchte meinen Mund ausspülen
I want to my mouth clean
I want to clean my mouth

(6) ich möchte zum Küchentisch fahren
I want to to the kitchen table go
I want to go to the kitchen table

If the wheelchair takes the table in Example 6 to be a table with specific structural features that make it fitting for a kitchen and does not understand it as the only table that is part of the only kitchen in the situation, it will fail to understand that a dining table permanently moved from the kitchen to the living room stops being *the kitchen table* and becomes *the living room table*.

The wheelchair must also ascertain what the user is referring to independently of how the user chooses to refer to it. For instance, a user may refer to his or her own hands in different ways in German as in Examples 7 and 8. The wheelchair should understand that the user is referring to his or her own hands in both cases.

(7) ich möchte meine Hände waschen
I want to my hands wash
I want to wash my hands

(8) ich möchte mir die Hände waschen
I want to me the hands wash
I want to wash my hands

If the wheelchair assumes that *mir* (*me*) in Example 8 is the person who the user wants to wash all present hands for instead of the person whose hands are to be washed, it will not be able to determine whose hands the user is referring to when two or more people are present and will fail to understand the user wants it to bring him or her to the wash basin.

Moreover, the wheelchair must recognise what the user is referring to independently of whether the same words carry different meanings in different utterances from a formal perspective. For instance, when the user refers to a book as in Example 9, the wheelchair should not attempt to identify that book as part of the user as it did to the user's hands in Example 8. Nor should it assume any other relation between the user and the book.

(9) ich möchte mir das Buch holen
I want to me the book get
I want to get the book

Furthermore, after the user refers once to an object such as a book (*das Buch*), a key (*der Schlüssel*), or a bowl (*die Schale*), depending on what it is, he or she may refer to it again by *es/ihm* (*das*), *er/ihn/ihm* (*der*), or *sie/ihr* (*die*). Therefore, for Examples 10 and 11 the wheelchair needs to identify what the user is referring to based on not only what has been referred to thus far but also what those objects are.

(10) siehst du das Buch? ich möchte es auf den Tisch legen.
see you that book? I want to it on the table put.
do you see that book? I want to put it on the table.

(11) siehst du den Schlüssel? ich möchte ihn auf den Tisch legen.
see you that key? I want to it on the table put.
do you see that key? I want to put it on the table.

Moreover, one may refer to a simple thing such as a sofa as an instance of a certain category of things while another person may refer to it as an instance of a different category. As illustrated in Examples 12 and 13, it is only by remembering the way simple things have been referred to thus far that the wheelchair can recognise that different users are referring to the same thing by *es*, *er*, or *sie*.

(12) bring mich zum Sofa. weißt du, wo es ist?
take me to the sofa. know you, where it is?
take me to the sofa. do you know where it is?

(13) bring mich zur Couch. weißt du, wo sie ist?
take me to the couch. know you, where it is?
take me to the couch. do you know where it is?

Finally, users often refer to non-unique instances of categories of things in the apartment such as a table (*der Tisch*). In that case, users refer to them sometimes as an instance of a more specific category as in Example 14, sometimes as part of a room as in Example 15, and sometimes as something located in a room or functional area as in Example 16.

(14) bring mich zum Esstisch
take me to the dining table
take me to the kitchen table

(15) bring mich zum Küchentisch
take me to the kitchen table
take me to the kitchen table

(16) bring mich zum Tisch in der Küche
take me to the table in the kitchen
take me to the table in the kitchen

However, users may also rely on their location as well as the wheelchair's location to refer to something that is not unique in the apartment, but is unique in a particular room or functional area. The room or area may be where the user and the wheelchair are currently as in Example 17 or where the user informed he or she wants to be in the future as in Example 18. Failing to identify the table the user is referring to in any of these cases would result in the wheelchair's taking the user to the wrong destination.

(17) bring mich zum Tisch
take me to the table
take me to the table

(18) bring mich in die Küche zum Tisch
take me to the kitchen to the table
take me to the kitchen to the table

Yet users may also refer to something that is unique solely amongst the things that they need to use during the activity they are performing. For instance, if someone knocks on the apartment door and the user begins the activity of opening the door for the visitor, the user may refer to the apartment door not only as the door to the apartment as in Examples 19 and 20 but also as a simple door as in Example 21. They may refer to the door as a simple door independently of where they are currently and independently of how many doors there are in the apartment. Given the sheer number of doors around interactants (cabinet doors, fridge doors, room doors, etc.), failing to identify the relevant door for the activity being performed results in the user's being taken to an apparently random position in the apartment.

(19) bring mich zur Wohnungstür
take me to the apartment door
take me to the apartment door

(20) bring mich zur Haustür
take me to the apartment door
take me to the apartment door

(21) bring mich zur Tür
take me to the door
take me to the door

In short, how simple things are referred to depends on the ways people typically refer to instances of those categories of things in a language, how users choose to refer to them, whether they have been referred to thus far, and whether

they are unique in the apartment, the interaction room, the room last referred to, or amongst the resources required for an activity. In order to understand the user's commands and perform the demanded services properly, the wheelchair needs to identify the simple things being referred to in the situation reliably, handling all these referential phenomena properly.

1.1.3 Related to configurations

Once the wheelchair recognises what the user is referring to, it needs to understand whether he or she is indicating a simple thing's state or a change in its state (event). Recognising represented states in the situation is essential to verify whether the mentioned things are in those states currently and to ensure they arrive at those states in the future. For instance, Example 22 indicates an attribute of the user, namely the user's location, and Example 23 indicates a change in his or her state, that is, a change in his or her location.

(22) ich bin in der Küche
I am in the kitchen
I am in the kitchen

(23) ich bin in die Küche gekommen
I have to the kitchen come
I came to the kitchen

In terms of events, understanding who shall perform *labour* is essential for the wheelchair to cooperate with the user in an orderly fashion. A user may represent an action performed by him or her alone as in Examples 24 and 25. In that case, the wheelchair needs to understand that the user is the one performing the action.

(24) ich habe meine Hände gewaschen
I have my hands washed
I washed my hands

(25) ich habe die Tür geöffnet
I have the door opened
I opened the door

If the wheelchair understands these user actions as actions performed *per labour*, it will not wrongly assume it is supposed to wash the user's hands or open the door for him or her when it hears Examples 26 and 27. The user is the one supposed to wash his or her own hands and open the door, not the wheelchair.

(26) ich möchte meine Hände washen
I want my hands to wash
I want to wash my hands

(27) ich möchte die Tür öffnen
I want the door to open
I want to open the door

The user may also represent services performed by the wheelchair as in Examples 28 and 29. In such cases, the wheelchair should understand that it is the one performing a demanded service after undertaking a command, that is, it should know that someone else had specifically delegated this task to it. Different from the user in his or her actions, it did not act in 'free will'.

(28) der Rollstuhl hat sich aufgeladen
the wheelchair has itself recharged
the wheelchair recharged itself

(29) der Rollstuhl hat mich in die Küche gefahren
the wheelchair has me to the kitchen taken
the wheelchair took me to the kitchen

To distinguish services from actions, one can use the following grammatical operations. All services can be represented *in locution* as in Examples 32 and 33 whereas actions cannot.

(30) ich habe den Rollstuhl darum gebeten, sich aufzuladen
I have the wheelchair . asked, itself to recharge
I told the wheelchair to recharge itself

(31) ich habe den Rollstuhl darum gebeten, mich in die Küche zu fahren
I have the wheelchair . asked, me to the kitchen to take
I told the wheelchair to take me to the kitchen

All services can also be represented as performed *per labour* and demanded *per locution* as in Examples 32 and 33. In all such cases, the person demanding the service is the wheelchair user and the one performing the service is the wheelchair.

(32) ich habe den Rollstuhl sich aufladen lassen
I have the wheelchair itself recharge made
I had the wheelchair recharge itself

(33) ich habe mich vom Rollstuhl in die Küche fahren lassen
I have me by the wheelchair to the kitchen take made
I had the wheelchair take me to the kitchen

In contrast, a user may alternatively represent an action by him or herself that is performed *per locution*, that is, an action that is performed solely by uttering something. For instance, the user may represent the action of recharging the wheelchair by telling it to recharge itself as in Example 34 and the action of going to the kitchen by telling the wheelchair to take him or her there as in Example 35. In such cases, the wheelchair needs to understand that this action consists of delegating a task to someone else and is only completed when the service provider performs the delegated task.

(34) ich habe den Rollstuhl aufgeladen
I have the wheelchair recharged
I recharged the wheelchair

(35) ich bin in die Küche gefahren
I have to the kitchen gone
I went to the kitchen

Unless the wheelchair understands the user is delegating a task to it in these actions, it will not be able to understand utterances such as Examples 36 and 37 as commands for it to perform the delegated task.

(36) ich möchte dich aufladen
I want you to recharge
I want to recharge you

(37) ich möchte in die Küche fahren
I want to the kitchen to go
I want to go to the kitchen

In addition, a user may indicate an action by a team comprising him or herself and the wheelchair. For instance, the team may be referred to as a group as in Example 38 or as enumerated members as in Examples 39 and 40. In all cases, the group is performing an action, not a service, because they did not receive a command from anyone outside the group. However, once we focus on the members of the group, it comprises a service client and a service provider and the group's action is a service provided by the wheelchair to the user.

(38) wir sind in die Küche gefahren
we have to the kitchen gone
we went to the kitchen

(39) der Rollstuhl ist mit mir in die Küche gefahren
the wheelchair have with me to the kitchen gone
the wheelchair went with me to the kitchen

(40) ich bin mit dem Rollstuhl in die Küche gefahren
I have with the wheelchair to the kitchen gone
I went with the wheelchair to the kitchen

The wheelchair cannot assume that these utterances represent services or actions per locution because these utterances may be expanded as in Examples 41 and 42. The one who performs the action of opening the door by labour is the wheelchair user in both cases. However, the one who performs the service of taking the user to the door is the wheelchair in both cases. If the wheelchair assumes that referring to the group is merely an indirect reference to the actual service provider, it would not understand the expanding clause. Hence, the wheelchair needs to understand that the group performed an action by labour and then infer how the group members demanded and performed labour.

(41) wir sind zur Tür gefahren, um die Tür zu öffnen.
we have to the door gone, . the door to open.
we went to the door to open the door.

(42) ich bin mit dem R. zur Tür gefahren, um die Tür zu öffnen.
we have with the w. to the door gone, . the door to open
I went with the wheelchair to the door to open the door.

If the wheelchair cannot infer both demand and distribution of labour, it will not understand who is supposed to do what when the user makes commands such as the one in Example 43.

(43) fahren wir zur Tür, um die Tür zu öffnen.
go we to the door, . the door to open.
Let's go to the door to open it.

Finally, users may also demand a service merely by describing a route or referring to a simple thing in the situation. For instance, a user may demand the service of taking him or herself to the kitchen by representing a route to the kitchen as in Example 44 or by referring to the kitchen as in Example 45. The wheelchair needs to understand that these utterances are not merely representations of a route to the kitchen or references to the kitchen, but commands to take the user to the kitchen. However, not all things mentioned in isolation are related to destinations. In Example 46, for instance, by referring to his or her own legs, the user alerts the wheelchair not to hit them against obstacles in the way to a destination. Therefore, since not all references to a simple thing are commands to take the user to a location relative to it, the wheelchair needs to consider the kinds of things being referred to and who/what they are part of before deciding what service is being implicitly demanded.

(44) zur Küche
to the kitchen
to the kitchen

(45) Küche
kitchen
kitchen

(46) meine Beine!
my legs!
my legs!

In short, the wheelchair needs to understand the roles each person and object plays in the process being described or in potential processes so that it can understand who should perform the represented task per labour and what service is being implicitly demanded. Only so can the wheelchair contend with all the configuration-related linguistic phenomena enumerated above.

1.1.4 Related to logical nexuses

After recognising the roles people and objects play in a represented process, the wheelchair needs to identify which service is being demanded. This is not a trivial task because users do not always represent the service they expect the wheelchair to perform.

By design, any intelligent wheelchair must help its user accomplish what he or she desires to do. Hence, the state that the user wants to achieve is always the end of any sequence of events planned by the wheelchair. End states may vary extensively from "clean hands" and "book in one's hands" to "recharged wheelchair" and "open door". A minor portion of these states are achieved directly by a wheelchair service as in Example 47.

(47) lade dich auf
recharge yourself .
recharge yourself

Most often, the user achieves the end state through an action on his or her own as in Examples 48, 49, and 50.

(48) ich möchte die Tür öffnen
I want to the door open
I want to open the door

(49) ich möchte mir das Buch holen
I want to me the book hold
I want to get the book

(50) ich möchte mir die Hände waschen
I want to me the hands wash
I want to wash my hands

The user being somewhere in the apartment is never an end on its own, so the service of taking the user somewhere does not result in the end state desired by the user. The user wants to be taken to particular locations in an apartment so that he or she can perform an action there[2]. Therefore, the user's final position is always one where the user can perform an action on his or her own. For instance, the locations *at the door*, *at the desk*, and *at the wash basin* in Examples 51, 52, and 53 are not random positions relative to those objects that could be represented with those wordings. They are always positions where the user can perform the respective actions in Examples 48, 49, and 50. Essentially, though the user represents a general location relative to those objects, the user aims to be in any position that fits such a relative location description as long as he or she can perform a non-represented action per labour there.

(51) fahre mich zur Tür
take me to the door
take me to the door

(52) fahre mich zum Schreibtisch
take me to the desk
take me to the desk

(53) fahre mich zum Waschbecken
take me to the wash basin
take me to the wash basin

[2]Or achieve something relevant by moving there such as hiding from someone.

Conversely, by saying that he or she wants to perform an action such as the ones in Examples 48, 49, and 50, the user demands that the wheelchair perform the service of taking him or her to a position where he or she can perform that action. The reasoning behind understanding the desire to perform such actions as demands of the service of taking someone somewhere is that the wheelchair user cannot or should not go to those positions on his or her own. Since the wheelchair needs to help its user perform daily actions by taking the user to positions where he or she can perform those actions, the wheelchair should take the user to an appropriate position for performing the represented action whenever the user must be at a remote position for performing that action.

In particular, the actions in Examples 48, 49 and 50 cannot be accomplished in all positions within an apartment for different reasons. The door to be opened in Example 48 has a constant location relative to the apartment, so it serves as the relatum of the user's destination. The book to be picked up in Example 49 can be anywhere in the apartment and the furniture where it is located currently becomes the relatum of the destination. Finally, the user's hands in Example 50 are part of the user and the location where the user wants to be cannot be represented in relation to them. The user's destination is actually relative to a tool for washing his or her hands, namely the wash basin. This implies that the wheelchair needs to notice whether the user can access the action's goal and reach all the tools he or she needs to perform the action before planning the service of taking the user somewhere.

Moreover, other actions such as reading a book can be performed by the user anywhere in the apartment. Therefore, when users want to read a book or perform another similar action in a particular location, they need to specify that they want to perform an action in that location as in Example 54. The wheelchair should recognise the user's desire to perform this action at that location and plan a service accordingly.

(54) ich möchte das Buch auf dem Sofa lesen
I want to the book on the sofa read

I want to read the book on the sofa

Furthermore, a change in the position of the wheelchair while the user is sitting on the wheelchair causes a change in the user's position. Hence, a user may represent the wheelchair service of going somewhere as in Example 55, but once the user is sitting in the wheelchair, he or she can represent a change in the location of the wheelchair that causes the intended change in his or her location as in Example 56 or a change in both his or her location and the wheelchair's as in Example 57. In Example 55, the wheelchair service of going to the bed is represented. In Examples 56 and 57, the service of taking the user somewhere is not represented by the clause. Example 56 represents the behaviour of the wheelchair that causes the intended change in the user's position and Example 57 represents both that and the intended change in the user's position, not the service.

(55) komm zum Bett
come to the bed

come to the bed

(56) fahr zum Bett
go to the bed
go to the bed

(57) fahren wir zum Bett
go with me to the bed
go with me to the bed

Hence, when the user is sitting on the wheelchair, it is always the final position of the user, not the wheelchair's, that needs to enable him or her to perform a non-specified action, and this is the case no matter whether the user's location or the wheelchair's gets represented by a wording (see Examples 56-59). In all four cases, users demand that the wheelchair take them somewhere; the user's final position is either represented by a wording as in Examples 57, 58, and 59 or is a location relative to the wheelchair whose position is represented by a wording as in 56, 57, and 58.

(58) fahr mit mir zum Bett
go with me to the bed
go with me to the bed

(59) fahr mich zum Bett
take me to the bed
take me to the bed

Moreover, when users intend to perform an action, they may represent both their action and the wheelchair's service. Alternatively, the wheelchair's service may be replaced either by its operation or an action per locution involving it. In particular, users may utter two independent clauses consecutively as in Example 60 or they may create a clause complex as in Example 61. The wheelchair must understand that these two events are related to each other in the sense that the second is the reason or purpose for the first; the first represents directly or indirectly the service being negotiated whereas the second represents the action being enabled.

(60) fahr mich zur Tür. ich möchte die Tür öffnen.
take me to the door. I want to the door open.
take me to the door. I want to open the door.

(61) fahr mich zur Tür, damit ich die Tür öffnen kann.
take me to the door so that I the door open can.
take me to the door so that I can open the door.

Finally, the user may demand two services from the wheelchair with a single command. For instance, while one is still sitting on the bed and the wheelchair is still located at the charging station, one may ask the wheelchair to take them to the door because one wants to open it. Let us assume their utterances resemble Examples 60 and 61. The sequence of events that must occur includes changes in both interactants' locations. First, the wheelchair needs to go to the bed where the user is sitting; second, the user needs to sit down into the wheelchair; and finally, the wheelchair needs to take the user to the door so that he or she

can open it. The wheelchair needs to plan this sequence of events to understand which services are demanded from it. In this situation, the two services by the wheelchair consist of going to the bed and taking the user to the door.

In short, the wheelchair needs to understand the dependency relations between preconditions and events and infer what end state the user desires. Only then can the wheelchair understand how represented services and actions are supposed to be performed and which non-represented services and actions need to be performed.

1.1.5 Related to dialogue acts

After understanding all phenomena related to reference, configurations, and logical nexuses, the wheelchair needs to assess if its understanding of the user's utterance is reasonable given the speaker's identity and the current state of things in the situation. Only then can it recognise the user's dialogue act. For instance, let us assume Example 62 has been uttered.

(62) Rolland. ich möchte zum Waschbecken.
Rolland. I want to go to the wash basin.
Rolland. I want to go to the wash basin.

If the speaker is the wheelchair user, this is a command for the wheelchair to take the user to the wash basin. Whether the wheelchair needs to pick up the user somewhere before taking the user to the wash basin is a matter of planning how to perform that service, not of which service to perform. However, if the speaker is another person and the wheelchair is blocking that person's way, this is a command of another sort, namely a command for the wheelchair to move out of the way. How the wheelchair needs to move out of the way is, again, a matter of planning. Alternatively, if the speaker is not the wheelchair user and the wheelchair is not blocking his or her way, the speaker might be asking for information regarding the location of the bathroom if he or she is still new to the apartment. In this case, the wheelchair should simply tell him or her where the bathroom is.

These different understandings depend on interpersonal relations between interactants such as the speaker being *a person who uses the wheelchair* and their domestic roles such as the speaker being *a person who is visiting the apartment* or *a person who lives in the apartment*. The wheelchair needs to consider these relations and roles to recognise the intended command or question in the user's utterance. These understandings also depend on the current circumstantial attributes of the interactants such as one *being in the other's way to a destination*.

Moreover, a wording uttered by the same person may have different meanings in different situations. For instance, if a user is sitting on the bed and the wheelchair is located at the charging station, the utterance in Example 63 is a demand for the wheelchair to come to the bed where the user is sitting. However, if the user has just called the wheelchair by saying *komm her* (*come here*) and the wheelchair has not only undertaken the command, but also entered the bedroom, in this discursive context the utterance in Example 63 is a further specification of the previous command for the wheelchair to come to the user, not a new command.

(63) zum Bett
to the bed
to the bed

This further specification would only be required if the wheelchair had not assumed it needed to go to the bed to pick up the user. In that case, a response such as *ok, ich komme zum Bett* (*ok, I'll go to the bed*) could be perceived as a confirmation that the wheelchair misunderstood the user's intent with the first utterance that would, in turn, reduce the user's trust in the wheelchair's understanding. Therefore, such unnecessary further service specifications should be met with responses such as *ich weiß* (*I know*).

In short, not only interpersonal relations and interactants' domestic roles should be considered by the wheelchair when deciding what the user's utterance means, but also the current state of things, how recent their attributes are, and negotiated services in the contexts of both discourse and situation. With these situational features, the wheelchair should be able to respond to the user's utterances in a manner that assures him or her of its awareness of the surrounding conditions. The evaluation experiment to be reported in Chapter 14 verified the extent to which these situational and discursive features can lead the dialogue system to a reasonable understanding of commands.

1.2 Research Goals

Besides creating a working automation, this thesis aims at answering the existing scientific question of how to uniquely determine the illocutionary force of an utterance (the user's intent) based on the utterance itself and its situational and discursive contexts employing symbolic processing alone. To achieve this, it is necessary to understand referential descriptions of simple things, location specification, and proposals of actions and services by interactants. Since understanding references, specifications, and proposals is a precondition for understanding utterance situationally, I will pose a secondary goal, namely, that of showing how a particular kind of language-based taxonomy of simple things, locations, and processes can be used to support reference, specification and proposal understanding. These two goals can be formulated in the following way:

1. How to recognise the user's intent relying on what the user meant by the words chosen, the situation the interactants are in, and the ongoing discourse of interaction, making use of only symbolic processing?

2. How to create a language-based taxonomy of simple things, locations and processes that can be integrated into a rule-based understanding module?

The primary goal consists of integrating what the user says with the situation and discourse. This includes:

1.a. Resolving reference to actual things in the situation and actual and potential positions of those things respectively in current and subsequent situations.

1.b. Assigning participant roles to mentioned things in the described processes.

1.c. Building a logical sequence of events from the current state of things in the situation to the end state most likely to be intended by the user given the utterance, situation, and discourse from a need ranking and planning perspective.

1.d. Accounting for personal rights and duties as well as potential misunderstandings by the user of what the wheelchair is doing or did.

The secondary goal comprises modelling language as to improve the reusability of linguistic models and recognition of dialogue acts. For this, I shall use a combination of **Systemic Functional Linguistics** (SFL) and **Combinatory Categorial Grammars** (CCG) on the following grounds.

Particularly in situated dialogue, where a robust approach to pragmatic interpretation is required, SFL claims to provide a holistic account of language where dialogue acts are not 'exported' to a later pragmatic step, but are described as situationally potential lexicogrammatical features realised by wordings (Thibault and Van Leeuwen, 1996). If this claim of reliability is accurate, it would be possible to make use of this theory computationally to understand users' requests to an intelligent device in situated dialogues. Therefore, this is a claim worth testing. However, current systemic functional models of language were proven to be inefficient for parsing and understanding in general (Bateman, 2008).

On its turn, CCG is a formalism of grammar, not a theory of language in use as SFL claims to be (Steedman, 1996). This formalism allows for efficient parsing and it possesses a "transparent" interface between grammatical and semantic structure, indicating that each grammatical constituent corresponds to a semantic constituent. It is, however, agnostic regarding which grammatical structure is mapped to which semantic structure as it does not include a theory of semantics and pragmatics. Since CCG is open with respect to its semantic stratum and it is compatible with SFL treatment of text as representation (Couto-Vale et al., 2014), it is worth testing whether it is also compatible with SFL treatment of text as exchanges of service.

This brings us to the following question: SFL posits no semantic loss, but systemic functional models of language cannot support linguistic analysis. Could CCG supply what is missing? If the answer to this question is positive, utterances will be understood as a specific dialogue act in a situation, they will not be interpreted in an open ended inferential step.

SFL is also a theory of language that claims to explain why text means what it does (Halliday and Matthiessen, 2014, p.3). That is useful for predicting both what people might say in the situation they are in, and how they might compose a wording based on what those words mean. The first claim is worth testing for better restrictive models for speech recognition and the second claim is worth testing for language-based ontology construction.

Additionally, SFL's mainstream tradition also employs a rank scale (Halliday and Matthiessen, 2014, p.5), mapping semantic elements to groups and phrases and semantic configurations to clauses, which is a practice that reduces misalignment between grammatical and semantic composition.

Combining CCG flexibility with theoretically motivated semantic structures, I intend to align grammatical structure ranks in CCG resources with semantic structure ranks in both theory and implementation for making semantic structures be the same for parsing and generation and useful for both managing dialogue and tracking the states of things in the situation. In turn, this will reduce the need for conversion and, most importantly, prevent any loss of information at the interface between different components. Ultimately, since no information would be lost, this isomorphism would also enable automatic understanding of delicate interpersonal features.

The linguistic resource design subgoals to achieve Goal 2 of integrating a taxonomy into a rule-based understanding module are listed below:

2.a. Assuring that grammatical composition enables semantic composition both in terms of experiential semantics and speech acts.

2.b. Assuring that each grammatical structure rank corresponds to a semantic structure rank, thus making groups and phrases correspond to semantic elements and clauses correspond to semantic configurations of elements.

2.c. Assuring that a multiword expression such as *dreht sich* (*turns*) is treated as a form of a single term associated with a single class of events.

2.d. Assuring that the boundaries of grammatical structures are not limited to the boundaries of written words[3].

In addition, there are also two subgoals for further developing the theory of language in SFL. These theoretical subgoals are:

2.e. Developing the theory of lexis in SFL with multiword expressions.

2.f. Developing the theory of logical inference in SFL.

Finally, with the CCG resource developed for the understanding module, I aim at showing:

2.g. That CCG can be employed to create a rank structure with multiword expressions, in particular, covering reflexive, divisible, and prepositional verbs in German.

2.h. That CCG can be employed to suggest illocutionary forces.

The scope of this thesis is limited to these enumerated goals and the understanding components developed for this research will be evaluated in such terms.

1.3 Research Methodology

The research process can be divided into four steps:

1. Collection of linguistic data

[3]Which is relevant for understanding *der Küchentisch* (*the kitchen table*) as a reference to the same table as the one referred to in *der Tisch in der Küche* (*the table in the kitchen*)

2. Development of a taxonomy and a combinatory categorial grammar
3. Implementation of a text producer, components of understanding, and a dialogue system that utilizes them
4. Evaluation of the automation

First, **data collection** was performed with a **Wizard-of-Oz experiment**, wherein a hidden researcher remotely controlled the wheelchair speech and action while a user interacted with it assuming the wheelchair understood him or her. The user's spoken commands were recorded with a video camera and a microphone, transcribed and maintained as a corpus of spoken commands.

Second, a **language-based taxonomy** and a **combinatory categorial grammar** (CCG) were constructed based on the corpus of spoken commands. Special attention was given to multiword expressions such as *er kennt sich damit aus* (meaning *he knows it*), particularly, reflexive, divisible, and prepositional verbs, descriptions of simple things such the kitchen table and dining tables, whose constituing words make drastically different semantic contributions, and simple things such as sofas that can be instances of multiple language-based classes of things, namely *Sofa* (*Sofa*) or *Couch* (*Couch*), each of which is a subclass of a different class of things for anaphoric reference resolution: for instance, *Sofa ist Das* (*a Sofa is a "Das"*) and *Couch ist Die* (*a Couch is a "Die"*).

Thirdly, a **text producer** was implemented for producing grammatical wordings out of recognised speech. Programming modules were created to encapsulate and simplify access to the parser OpenCCG (Bozşahin et al., 2005), the ontology manager OWLAPI (Parsia et al., 2012), and the reasoner HermiT (Glimm et al., 2014). An understanding module that uses those modules was implemented with four integration components for 1) resolving reference, 2) understanding configurations of elements, 3) inferring nexuses between configurations of elements, and 4) noticing intent based on rights and duties, tackling misalignments in grounding, and making moves in dialogue flows. Subsequently, a dialogue system was implemented with the dialogue management framework DAISIE (Ross and Bateman, 2009) that utilises both the text producer and understanding module.

Finally, I will report on the **evaluation experiment** conducted. The **results** of the experiment will be presented in terms of the success rate of the entire understanding module. In the **discussion** of the results, each failure will be traced back to the first component that fails and potential improvements in coverage or architecture will be proposed.

1.4 Listening and understanding components

The listening and understanding modules are divided into the following six components:

1. **Text producer** a component that produces a grammatical text from speech transcriptions
2. **Language Analyser** a component that builds semantic structures for dialogic text segments

3. **Reference Integrator** a component that identifies the represented simple things and locations
4. **Configuration Integrator** a component that specifies the roles those simple things and locations can play in the represented configuration
5. **Nexus Integrator** a component that infers configurations logically related to the one represented that could count as a user-intended end state and plans how interactants can achieve that state together
6. **Move Integrator** a component that fits the user's utterance as a dialogue move taking the ongoing exchanges of services and information state into consideration

The inner working of each component is described in a separate chapter.

1.5 Outline

This thesis is divided into four parts:

1. State of the Art
2. Resource Construction
3. Architecture and Implementation
4. Evaluation and Conclusion

Below, I shall describe what I shall present in each part in general terms.

1.5.1 State of the Art

In the State of the Art, I describe other researchers' findings and their description of the linguistic phenomena they encountered, indicating the gaps in their approaches that I aim to fill and the limitations that I aim to overcome in this work.

Chapter 2 Dialogue in Interaction In this chapter, I present both non-computational and computational linguistic theories that allow us to predict and explain interactant's contributions to dialogue and interaction, focusing on those theories and computational methods that are most relevant for dialogues unfolding during interaction. I conclude the chapter with a list of overlooked areas in computational approaches to determine the speaker's intent in situated dialogues.

Chapter 3 Society and Words In this chapter, I review different models of grammar with their consequent theories of lexis, including computational approaches to multiword expression detection and the incipient theory of lexis in SFL. This includes generative compound words, semantic lexica such as word net and verb net, and domain ontologies and taxonomies. I conclude this chapter by explaining which kinds of multiword expressions need to be covered by a systemic functional theory of lexis for supporting semantic composition, which are the goals 2.a-f of this research.

1.5.2 Resource Construction

The process of building a linguistic resource can be divided into three subprocesses: collecting linguistic data, modelling, and building up a vocabulary. In particular, the model of the apartment and the processes that occur in it required for text processing is not the same as models created by autonomous wheelchair developers for controlling the wheelchair and its motions. It is actually a model of things as they are described, the circumstances those things are described in and the processes they are described as participants of. It is a linguistic ontology of things, circumstances, processes, and configurations thereof as they are represented in utterances, not a domain ontology. Similarly, the vocabulary created for text processing is not merely a list of all words spoken in commands, but a further specification of the linguistic ontology as a taxonomy, associating each semantic class with one or more German terms, and instance of those classes with one or more German names when applicable.

Chapter 4 Data Collection In this chapter, I explain the process of collecting a corpus of commands to an intelligent wheelchair. I describe the Wizard-of-Oz experiment conducted for this purpose and the corpus that was created alongside.

Chapter 5 Ontology Creation In this chapter, I list and describe the classes of simple things, circumstances and processes described in commands to the wheelchair as well as their represented configurations in the commands. I also list the actual things and their actual and potential locations in the apartment, that are instances of the described classes.

Chapter 6 Taxonomy Creation In this chapter, I describe how semantic classes and their instances are mapped to single-word and multi-word expressions found in corpora. Particularly, I will describe how names and terms are realised by single-word and multi-word expressions and how they are associated with actual things and classes of things, circumstances, and processes. Throughout this chapter, I theorise lexis using a systemic and functional approach, filling the theoretical gap therein as far as lexis is concerned.

1.5.3 Architecture and Implementation

After describing how the linguistic resource was constructed, I will describe the actual dialogue system that utilises the created taxonomy and six of its modules. The modules selected for description are those that will be considered in the evaluation of the linguistic resource and proposed integration process.

Chapter 7 System Architecture In this chapter, I provide an overview of the dialogue system that utilises the created taxonomy, describing the blackboard supporting the inter-module communication, the components that take turns in writing on the blackboard, and how they decide whose turn it is to write.

Chapter 8 Text Producer In this chapter, I describe the component that reads a sequence of recognised speech segments and produces a grammatical wording from it.

Chapter 9 Lexicogrammatical Analyser In this chapter, I describe the component that analyses wordings linguistically. This component incorporates the taxonomy described in Chapter 6 and a CCG described throughout the chapter to produce a semantic structure build up of classes and instances described in Chapter 5.

Chapter 10 Reference Integrator In this chapter, I describe the component that determines which simple things present in the situation, their current and potential locations to which a speaker refers. This component considers who is speaking to whom and the perspective from which one is speaking and the addressee is listening. It also utilises classes and configurational roles described in Chapter 5 to establish reference between symbols realised in utterances and actual or potential elements.

Chapter 11 Configuration Integrator In this chapter, I describe the component that maps linguistic representation to actual or potential states and events. This modules also checks whether represented actors can perform the material actions they are represented doing and whether the represented matter affords the material actions under representation. In addition, material actions are further divided as either actions per labour, services per labour, or actions per locution so that specified rights and duties can be used to verify whether the person represented as a service client can demand the represented service from the person represented as a service provider and whether those represented as service providers have to perform the service for them.

Chapter 12 Nexus Integrator In this chapter, I describe the component that maps a valid configuration of roles onto another that is not represented but can be integrated into discourse as an exchange move. Particular cases are discussed: the interdependency between the location of the wheelchair and the location of the wheelchair user, actions to be performed in a specified location, and actions that can only be performed in a particular location.

Chapter 13 Move Integrator In this chapter, I describe the component that tracks the ongoing exchanges in the current situation as well as the potential moves that can be performed at the current moment. This module is responsible for updating the discourse state by integrating an utterance as a move in an exchange, also considering other non-verbal moves that have been integrated thus far.

1.5.4 Evaluation and Conclusion

In Chapters 5-6, a taxonomy was created based on a corpus of spoken commands and, in Chapters 8-13, I described the dialogue system that utilises this taxonomy and its components for the automatic understanding of spoken commands. Here, I shall evaluate the resulting dialogue system and present my conclusions

regarding achieved goals and remaining issues as far as dialogue systems for intelligent wheelchairs are concerned.

Chapter 14 Evaluation In this chapter, I evaluate the dialogue system on two criteria: how frequently uttered commands are accurately processed in general and how frequently a wheelchair user can produce at least one command that is accurately processed when demanding a service from the wheelchair (more details in Section 14.1.9). This chapter is divided into four sections: description of the experiment, the success criteria, the results, and a discussion of failures focusing on what is still required in which component for the wheelchair to understand infelicitous utterances too.

Chapter 15 Conclusion In this chapter, I present my overall conclusions regarding achieved goals and remaining issues as far as dialogue systems for intelligent wheelchairs are concerned.

Part I

State of the Art

Chapter 2

Dialogue in Interaction

Different linguistic theories differ in terms of what they allow us to predict and the degree to which they allow us to explain interactants' contributions to dialogue. In this chapter, I introduce both non-computational and computational approaches for describing dialogue, focusing on those theories and computational methods that are most relevant for dialogues unfolding during interactions between humans and intelligent wheelchairs. I conclude the chapter with a list of the elements that computational approaches lack for determining what speakers want to achieve in this type of situated dialogues.

2.1 Non-Computational Approaches

In this section, I review non-computational approaches for explaining how interactants achieve what they achieve with utterances. First, I visit studies of utterance as observable physical phenomena. Then I review Austin's theory of utterance, which proposes the notion of illocutionary force as a separate notion from perlocutionary effect. Finally, I move to more specific theories of utterance that aim at either formalising or explaining illocutionary force. On the explanation front, I present Grice's theory of implicature and Schegloff's theory of social interaction. On the formalisation front, I present Searle's theory of speech acts, Austin's functional understanding of modal verbs such as 'can', and Matthiessen and Halliday's theory of speech function.

2.1.1 Speech as text production

In sociolinguistics, researchers observe and analyse spontaneous dialogues considering all perceivable cues that a listener must have considered in order to interpret a speaker's utterance and behaviour as he or she did (Schegloff, 2002b, 2004).

Using this evidence-based approach to dialogue analysis, Schegloff (Schegloff et al., 1977; Schegloff, 1979, 1992b, 2000b, 2002b) showed how interactants repair previous speech segments by rephrasing what they have uttered thus far and how they repair earlier parts of a speech segment in later parts of the same segment. To model this type of repair phenomena, Halliday (1987) proposed considering spoken text as a product of speech, not as its direct transcription. In such an

approach, each speech segment would perform one of three editing actions: they would either append a new segment to a collectively constructed dialogic text, replace a segment of dialogic text by a new one, or delete a segment by turning it obsolete. For the dialogue system of this thesis, a module for producing text based on speech segments was implemented following Halliday's suggestion (see Chapter 8).

2.1.2 Barge-ins

It has also been shown that speech overlap occurs whenever listeners start speaking to claim their turn (barge-in) and that overlap ceases whenever all but one speaker concede the turn (Sacks et al., 1974, 1978; Jefferson, 1984, 1986; Schegloff, 1992b,a, 2000a, 2001, 2002b). Most current commercial dialogue systems support user barge-in. However, since barge-ins did not occur in the linguistic data (see Chapter 4), support for them was not implemented in the present work.

2.1.3 Utterance force and effect

In contrast to its physical counterpart, an utterance as a product of speech is a text segment (not a continuous speech segment) that can be taken as a command, a question, a statement or some other type of dialogue act. The task of determining whether an utterance is to be understood as a command, a question, or a statement has long been known to be difficult. In his lectures on how humans do things with words, Austin (1962) stated the following:

> [It is doubtless that] both grammarians and philosophers have been aware that it is by no means easy to distinguish even questions, commands, and so on from statements by means of the jejune grammatical marks available, such as word order, [inflectional] mood, and the like, though perhaps it has not been usual to dwell on the difficulties which this fact obviously raises: For how do we decide which is which? What are the limits and definitions of each?

Regardless of how difficult this problem is known to be, we need to solve it in order to create intelligent devices such as the autonomous wheelchair that was addressed in this research. The first step towards this goal was taken by Austin himself. Previously, statements had been treated as propositions in Propositional Logic. In the case of personal diaries and fictional stories, treating sentences as propositions is probably effective because the author either writes to him or herself or assumes the readers will accept statements as true within a fictional world. However, when we move from monologues to dialogues, a statement ceases to be a proposition and becomes an attempt to convince others of something. If the dialogue is orderly, a proposition is first established when a statement is acknowledged by the addressees.

Austin reached this realisation by using formulas. For instance, the formula *John told Mary that X is the case* contrasts with *John convinced Mary that X is the case*, the first being a **statement** and the second being a **felicitous statement**. In the same way, the formula *John told Mary to do X* represents John's attempt to make Mary do something and *John made Mary do X* represents John's success in making Mary do something by uttering words, which

means that Mary did what she was told to do. The first formula represents a **command** and the second a **felicitous command**.

Using such pairs of formulas, Austin (1962) described **utterances**[1] as attempts to achieve something in a social interaction. In his account, by uttering words, a person can exert his or her power over others, suggesting or forcing them to behave in intended ways. In this sense, a person would apply "force" onto another while uttering words, a force that he calls **illocutionary force**. The resulting state after the addressees' reaction is the interpersonal effect the speaker achieved by uttering words, that is, the **perlocutionary effect**. Finally, an utterance also has content, that which is represented by the utterance, the **locutionary content**.

According to Austin, it is by describing the situational conditions for utterance felicity that we can predict whether an illocutionary force will result in the intended perlocutionary effect.

1. Turning Austin's argument around, we would arrive at the proposition that it is possible to determine all potential felicitous utterances in a situation if we know all conditions for utterance felicity.

2. Assuming that we have this set of utterances and that a wheelchair user knows the situation in which he or she is placed, it is also possible to determine the set of utterances that intelligent device users are likely to say based on their situation.

That we are able to predict what people say in different situations is a standard assumption in the design of dialogue systems and it can be stated by now that there is accumulating evidence that situation and discourse are enough constraints for us to predict most utterances and understand utterer intents given the fact that dialogue systems exist and work most of the time. What computational approaches lack at the moment is a method to predict utterances and understand their meanings when the system is required to react properly in a variety of different situations or when the situation is constantly changing. In such cases, we need to automate situated understanding of utterances and not only provide a dialogue system with a fixed meaning for each utterance. This has not been done thus far.

2.1.4 Implicatures

Another major task in the process of determining what a speaker wants from the addressees is that of creating a bridge between, in Austin's terms, locutionary content and illocutionary force. Locutionary content, as what is represented by the utterance, is not easy to model, but it is the most straightforward aspect of an utterance in terms of modelling (see Chapter 5). The relation between utterance as representation and what the speaker wants from an addressee is much less straightforward and considerably more dependent on contexts of situation and discourse.

[1] He also calls utterances by the Latin term "locutions" and all his adjectives related to utterance such as "locutionary", "illocutionary", and "perlocutionary" are derived from the latter term.

In an attempt to determine which utterances in a discourse do not represent the negotiated information, service or good, Grice (1975) proposed understanding dialogue as a cooperation where four maxims of communication apply:

1. **The maxim of quantity** whereby one gives only as much information as required
2. **The maxim of quality** whereby one states only what one can know
3. **The maxim of relation** whereby one gives only information that is relevant to the discussion
4. **The maxim of manner** whereby one refers only to things that can be identified and only uses names and terms that can be understood in the situation

Utterances that apparently fail to comply with any of these maxims would leave the addressee asking him or herself why the speaker uttered all those words and would trigger inferential processes until some inferred content complies with the maxims. The represented content is the **explicature** and the implied content is the **implicature**, both of which are situational as we shall see in the remainder of this and the next chapter.

Grice's maxims apply to situated dialogues with wheelchairs in different degrees. The maxims of quality and relation tend to apply to discussions about particular topics where exchange of information is in the center. In our case, intelligent wheelchairs are designed to provide services to their users, not information, and therefore, these maxims apply to a lesser extent to the types of dialogue we need to cover. The maxim of manner is quite central to reference integration (see Chapter 10) and the assumption that this maxim is respected informs how names and terms are understood and how referenced things are identified.

As for the correlation between inferences and noncompliance with maxims, I could not find any examples in the linguistic data of this research where any one of the maxims is violated; nonetheless multiple inferences are required for an adequate exchange of services to take place. For this reason, the split between explicature and implicature is a useful concept for intelligent devices, but the usage of Grice's maxims to predict inferences is not equally as useful.

2.1.5 Routines and subdialogues

When describing social interactions, conversation analysts (Schegloff, 1968; Jefferson, 1972; Reichman, 1981; Schegloff, 1986, 2002a, 2011) showed that bounded dialogues such as phone calls are not necessarily opened and closed solely by greetings and goodbyes. There are other types of utterances typical of openings and closings for each particular register. For instance, a person calling into a radio talk show is likely to be asked to identify him or herself to listeners at the beginning of the dialogue and this constitutes a pattern of dialogue openings for radio talk shows, although giving one's own name and asking for another person's name are not a mere act of opening a dialogue. The caller telling his or her own name is expected when opening a dialogue in a talk show and it is the routine, not the dialogue act, that makes such utterances typical in this phase of the call.

Conversation analysts also showed that pairs of utterances such as a question and the respective answer constitute a unit independently of whether they come together or not. Related utterance pairs can be separated by a number of other utterance pairs in dialogue. This observation has consequences to the modelling of subdialogues of several types. Referential grounding subdialogues (Mast and Wolter, 2013a; Mast et al., 2014a,b) are just one example of places where insertion sequences occur. Conversation analysts also showed that the absence of an expected response to a request functions as an actual response if the addressee is known to have heard and understood the request.

In the linguistic data collected for this research (see Chapter 4), we observed that most dialogues between the wheelchair and its user comprised only two utterances, a command and an undertaking, which means that there was no interaction routine underlying dialogue openings and closings. In addition, linguistic data from the Wizard of Oz experiment presented no insertion sequences as a consequence of the interaction design. As a result, the intelligent wheelchair developed based on these linguistic data was required to understand utterances without relying on insertion sequences, that is, without making clarification and referential grounding questions or questions of any type before undertaking the command. The wheelchair needs to understand all references to simple things in the situation without subdialogues or failure if it is to simulate what humans do. In the evaluation chapter (Chapter 14), I report that the dialogue system developed for this research is indeed capable of reliably resolving references without clarifications.

2.1.6 Speech acts

When describing what speakers achieve with words, Austin focused on the difference between the intent of the speaker when making his or her utterance (illocutionary force) and the actual effect he or she achieved with it, where Grice focused on the difference between what is represented by the utterance and what the speaker implies. Searle (1965, 1975a,b) noted that there is another dimension to the task of understanding an utterance.

Let us consider a grammatical frame to be a fixed sequence of slots that can be filled by words. If we describe clauses in terms of grammatical frames, they can be divided into the following three types of **direct speech acts** based on the sequence of filled slots. Parentheses symbolise the slots. **Declarative**, **interrogative**, and **imperative** are types of grammatical frames, and **statement**, **question**, **command** are dialogue acts.

1. (Rolland) (went) (to the kitchen). **Declarative Statement**
2. (Did) (Rolland) (go) (to the kitchen)? **Interrogative Question**
3. (Who) (went) (to the kitchen)? **Interrogative Question**
4. (Where) (did) (Rolland) (go)? **Interrogative Question**
5. (Go) (to the kitchen)! **Imperative Command**

However, there are some clauses such as the ones below, which present the same grammatical frames as any other interrogative clause, but are primarily commands for all practical purposes, not questions.

6. (Will) (you) (go) (to the kitchen)? **interrogative Command**

7. (Can) (you) (go) (to the kitchen)? **interrogative Command**

According to Searle, the primary meaning of a clause in terms of a dialogue act is its **primary speech act** and the type of clause based on a grammatical frame is its **secondary speech act**. The primary speech act is what the speaker is doing with his or her words whereas the secondary speech act is what can be mapped to sequences of slots.

In the study reported in this thesis, my objective was to determine the dialogue act of an utterance (primary speech act) based on the situation and the discourse. For this purpose, secondary speech acts were required only in the sense that all grammatical frames needed to be recognised. For the short dialogues between humans and wheelchairs, where people do not report what others said, there was no need to classify clauses at the secondary speech act level. Moreover, Searle's typology of primary speech acts (illocutionary acts) was not used because all user utterances in the study belonged to the same type of speech act in his typology.

2.1.7 Modal verbs

Concerning the determination of a dialogue act based on the grammatical frame and the situation, Austin (1956) made a further observation that is useful for situated dialogue systems. He pointed out that the word *can* may have various situational meanings depending on the utterance content and the situation. When applying his notions to the classification of user utterances, we can also observe this variation. Words such as *can* indeed have different meanings for the same content in different situations. For instance, the word *can* in *Can you take me to the shower?* may represent a wheelchair's ability if a potential buyer directs this question to an intelligent wheelchair in a store, but it may also be a grammatical mark of a polite command if a wheelchair user gives this command to his or her own wheelchair in an apartment. In turn, this word also has different meanings for different contents in the same situation: for instance, while *can* realises a polite command in the utterance above, it realises the dialogue act of politely making oneself available for a command in *What can I do for you?*

In this research, auxiliaries such as *can* and *will* were understood to carry different meanings depending on what was being represented and on the situation (see Chapters 9, 12, and 13).

2.1.8 Speech function

Halliday and Matthiessen (2014) proposed a theory of how dialogue acts are realised by grammar that generalises speech act theory. In their description, a **grammatical speech function** is realised by a mood structure comprising a Finite verb and potentially a Subject. This mood structure is a segment of a grammatical frame and it realises three types of speech functions: declarative, interrogative, and imperative. A **first-level semantic speech function** is then realised by a combination of the represented content, the grammatical speech function, and potentially a modal auxiliary such as *can* and *will*. The grammatical speech function corresponds to a direct speech act (secondary speech act)

and the first-level semantic speech function corresponds to an indirect speech act (primary speech act).

1. Go to the kitchen! **Imperative Command**
2. Can you go to the kitchen? **Interrogative Command**

However, if a speaker tells the addressee to tell him or her something, the resulting imperative clause complex (Examples 2 and 3) functions as a single question clause (Example 1). In other words, a command to tell something is a question.

1. What time is it? **Interrogative Question**
2. Tell me what time it is! **Imperative Command$_{\textbf{tell}}$ Question**
3. Can you tell me what time it is? **Interrogative Command$_{\textbf{tell}}$ Question**

According to their description, processes of telling, asking, wanting, knowing, and so on belong to a special category of semiotic processes that are capable of realising a second or upper level speech function in this fashion.

In the linguistic data collected for this research, second and upper level speech functions did not occur. For this reason, this generalisation could not be applied.

2.1.9 Implicatures revisited

All implicatures described by Grice — taken as indirections in utterances and inferences in understanding — are a known issue for the description of dialogue acts as grammatical phenomena such as speech acts and speech functions (Schegloff, 1988). For instance, while it is relatively straightforward to understand Examples 1-4 as commands with a semantic speech function, it is less obvious how to understand Example 5 as a command using speech function theory.

1. [(Take) (me) (to the kitchen)]! **Imperative Command**
2. [(Will) (you) (take) (me) (to the kitchen)]? **Interrogative Command**
3. [(Can) (you) (take) (me) (to the kitchen)]? **Interrogative Command**
4. [(You) (should) (take) (me) (to the kitchen)]. **Declarative Command**
5. [(I) (need to) (go) (to the kitchen)]. **Declarative Statement?**

However, there is a point at which speech function cannot be treated as a dialogue act at any upper semantic level. For instance, in our linguistic data Example 1 is a command for the wheelchair to take the user to the washbasin. The service of taking the user to the washbasin is not represented by the utterance; however, it is implied given that the user can wash his or her hands only at the washbasin and cannot go to the washbasin on his or her own, and that the wheelchair has the duty of taking the user where he or she needs to go.

1. [(I) (need to) (wash) (my hands)]. **Declarative Statement?**

This means that a speech function theory needs to be supplemented by logical reasoning if one wants to support implicatures of the types recognised by Grice. In this thesis, the module for logical reasoning required for this is described in Chapter 12.

2.1.10 Preparations and follow ups

Halliday and Matthiessen's theory of speech functions (2014) includes not only initial utterances, but also responses. Treating dialogue flows as state transducers where each utterance is a move from one dialogue state to another, they described an initial utterance as an initiating move and a response as a responding one. For simple exchanges of services and information involving an initial utterance and a response, such a model suffices.

Schegloff (1980, 1988) noted that there are exchanges of services as well as of information that start with neither an offer nor a demand for services. Using an example plausible for the wheelchair in this research, we would have the following:

`Robot:` What can I do for you?
`Human:` Can you take me to the kitchen?
`Robot:` Ok, I'll take you there.

In the example above, the wheelchair's utterance *What can I do for you?* is not a proper demand of information the response to which is information. It is a demand for a demand for services, that is, it is a demand for a command. It makes the robot available for a command. Schegloff (1980) called a sequence of utterances prior to an actual exchange of services or information a "presequence" because they function as **preparations** for a command, statement, or question. Moreover, Tsui (1994) pointed out that the same occurs after an exchange. Interactants may thank each other or criticise each other's actions. Utterances in postsequences function as **follow ups** as illustrated in the dialogue below.

`Human:` Can you take me to the kitchen?
`Robot:` Ok, I'll take you there. *Robot takes Human to the kitchen.*
`Human:` Thanks!
`Robot:` You're welcome!

In the linguistic data, preparations did not occur, but follow ups did. For this reason, a model of dialogue acts that includes not only pairs of moves was required.

2.2 Computational Approaches

In this section, I review computational approaches for determining what a speaker wants to achieve with his or her utterance. Some approaches consist of classifying utterances in terms of what the speakers are doing with their words based on a corpus. These approaches can be divided into those with domain-specific dialogue acts and those with cross-domain dialogue acts. Other approaches consist of classifying utterances according to what they represent and leave all or most of the decision about what the speaker wants to achieve to a component that determines the listener's obligations. Finally, other approaches do not commit to any manner of understanding utterances and concentrate on the information state, which needs to be updated at each utterance. These include Belief-Desire-Intention (BDI) and plan-based architectures.

2.2.1 Domain-specific dialog acts

One of the earliest dialogue systems that included a data-based taxonomy of dialogue acts for utterance classification was VERBMOBIL (Jekat et al., 1995;

Alexandersson et al., 1997). The taxonomy is shown in Figure 2.1.

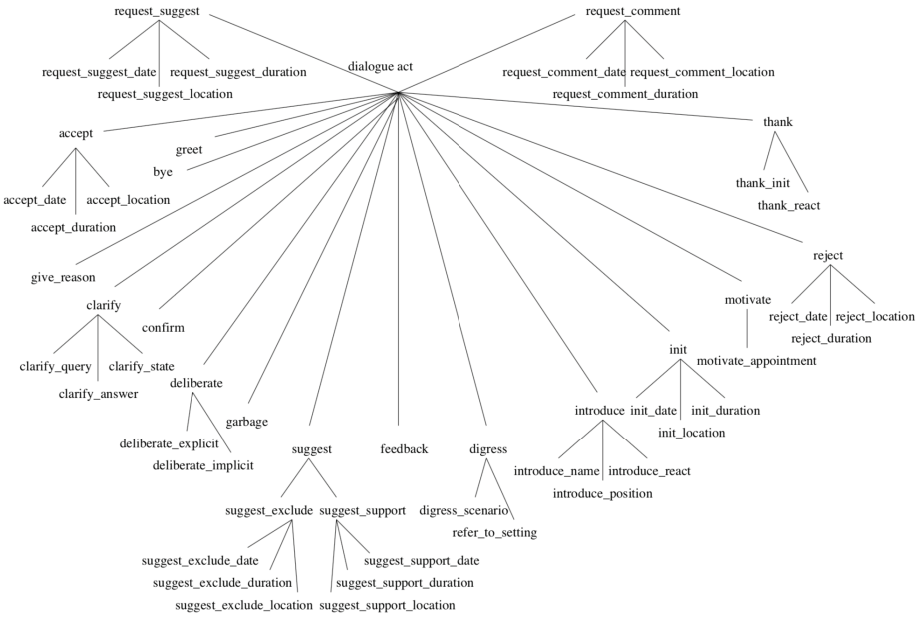

Figure 2.1: Dialogue Act Taxonomy for VERBMOBIL

This dialogue system was designed to play the role of an interpreter for a German speaker who wants to negotiate a meeting with an English speaker. The fact that the German speaker wants to negotiate the date/time, duration, and location of a meeting has already been established when the virtual interpreter is called into the interaction.

The VERBMOBIL taxonomy of dialogue acts is a domain-specific taxonomy with a predominantly two-level classification of dialogue acts: the general class of utterances distinguishes utterances according to whether the speaker is suggesting an attribute for the meeting or favouring, disfavouring, accepting, or rejecting an attribute suggested by other interactants; in contrast, the specific class of utterances distinguishes them not only by the negotiation act that the speaker is performing, but also by the type of meeting attribute that the speaker is negotiating.

This two-level classification approach can be and frequently is applied to utterance classification in dialogue systems and it is effective when interactants are planning an event such as a meeting or ordering a customised product such as a pizza. The classifier is, however, applicable only after it has been established that an event of a given type is to be planned or a customised product of a given type is being ordered. The reason for this is that the attributes to be negotiated, and consequently the corresponding specific classes of utterances, depend on the attributes of the future thing that needs to be negotiated.

Such an approach is inadequate for the negotiation of services in situated dialogues because it depends on the assumption that the situation is not changing and that the activity is solely the negotiation of the attributes of a future thing, which is not what occurs between a wheelchair and its user.

2.2.2 Cross-domain dialogue acts

Domain-specific taxonomies such as VERBMOBIL can be applied only to a single domain. A statistical classifier can be trained to recognise specific classes of utterances for a particular domain whereas the general classes function as an upper model, which is cross domain.

In an attempt to overcome the domain-dependency, many collective efforts were made to propose a cross-domain taxonomy of dialogue acts that can be applied to the task of classifying utterances independently of the domain and the activity in which the interactants are involved. All such efforts failed to provide categories of utterances that can be used for the task of classifying actual utterances based on a general-purpose corpus.

One of the most recent of these efforts was the taxonomy of dialogue acts DAMSL, which stands for Dialogue Act Markup in Several Layers (Allen and Core, 1997; Jurafsky et al., 1997; Core and Allen, 1997; Core, 1998). In this markup, each utterance is classified three times. First, utterances are classified according to whether the speaker did something by uttering words and, if not, why not. For instance, utterances may be "uninterpretable", they may be "abandoned" by the speaker through repairs, and they may be directed to the speaker him/herself, i.e., "self-talk"; only when speakers do something with their words are their utterances classified according to what they did. The second-level utterance classification divides utterances according to whether the speaker is performing a task within the activity in which interactants are involved such as booking a train ticket, whether the speaker is managing the tasks within the activity such as postponing a task to complete another task first, whether the speaker is managing the dialogue flow itself such as greeting, asking for repeats, or saying goodbyes, or whether the speaker is doing something else (other). The task utterances are then classified for the third time according to whether the user is requesting something from the addressees (forward-looking) or responding to a request made by the addressee (backward-looking). Under each of these two categories, a range of dialogue acts is proposed: 1) demanding and offering information, 2) demanding and offering services, 3) opening and closing an exchange (*What can I do for you?*, *That's it!*), 4) thanking, welcoming, reprehending, and apologising to other interactants, 5) reacting to unpredicted events with sounds such as *Ouch!* or 6) reacting to one's own mistakes with sounds such as *Oops!*.

This taxonomy of dialogue acts is quite sound as far as dialogue management is concerned. Any dialogue system needs to have a means of integrating an utterance that is not a statement, question, command, or offer into the dialogue and such a taxonomy is a step in this direction. However, when researchers attempted to let an utterance classifier learn to classify utterances in a cross-domain fashion based on a corpus, they showed that utterances cannot be reliably classified into such classes based on their form alone (Core and Allen, 1997). Amongst the factors that contribute to this result is the fact that the same wording may be used to perform different dialogue acts in different situations and at different points of discourse as illustrated in Section 1.1.5.

This typology was not used in the study because almost all utterances classified for dialogue acts belonged to one single class in DAMSL, namely, that of commands. The classes of dialogue acts to be described here are all subclasses of commands. However, as mentioned above, all deployable dialogue systems

need to be able to handle task-unrelated utterances and DAMSL serves as a overview of dialogue acts most frequently encountered in dialogue systems.

2.2.3 Representation and discourse obligations

Other researchers moved away from determining a dialogue act via formal utterance classification and approached the problem from a different perspective (Traum and Allen, 1994). On the one hand, they proposed that an utterance should be classified as a representation of a state or an event based on a corpus, not as a dialogue act. The assumption behind this decision to bypass dialogue act classification is that process verbs and their complements are more stable across different clauses and situations than the delicate formal features such as subject-finite order and modal verbs, which contribute to the realisation of a dialogue act. Therefore, a statistical utterance classifier has a greater chance of learning the represented content than a dialogue act from a corpus. On the other hand, these researchers proposed the task of classifying events according to who must perform them or who must give information about whether they happened, are happening, or will happen. A dialogue act such as demanding a service is assumed if the event is taken to be a service to be provided by the addressee and it should trigger some inference process if it is taken to be an action to be performed by the speaker.

Such an approach results indeed in a higher success rate in utterance classification for dialogue acts at the expense of restricting the dialogue act classes to information and service exchange and of allowing no fine-grained control by the speaker of how an utterance is to be understood. For instance, if a dialogue system assumes it is supposed to perform a service for the speaker, it will perform this service whenever the service is represented, regardless of whether the speaker wants it to perform this service or only asks whether this service is available or something else. Nonetheless, the increased dependence on situation and discourse makes such an approach useful whenever a dialogue system can interpret representations of an event or a state in one single way, independently of grammatical features related to the dialogue act.

Such an approach cannot be adopted directly for developing situated dialogue systems because of the resulting coarse understanding of utterances, but it presents a method for dealing with utterance classification in the contexts of situation and discourse, which can be taken as an inspiration for situated dialogue systems. The approach adopted in this research was inspired by it.

2.2.4 Belief-Desire-Intention (BDI) agent architecture

In the development of ARTIMIS, Sadek and colleagues (1994; 1996; 1997; 1997; 1999) developed a dialogue system architecture based on interactant's beliefs, desires, and intentions. In this approach, each interactant is modelled by the system from its own perspective. What the system assumes each interactant believes to be the case is represented as a belief, what they want as a desire, and their planned actions as intentions.

With the help of keyword spotters, Sadek et al. classified utterances as representing either the speaker's belief (statement) as in *I am sitting on the bed*, a speaker's desire (statement of will) as in *I need to go to the kitchen*, or a speaker's intention (command) as in *Take me to the kitchen!* With this simple

classification of utterances, the assumption that the speaker wants to delegate a particular task to the addressee, and a theory of possible worlds, they were able to control a dialogue system for booking flight tickets and other form-filling dialogues.

The classification of utterances into three categories, namely the expression of personal belief, desire, or intent, is not grounded on utterance structure. As a result, domain-specific reasoning needs to be implemented in terms of utterance classes. For instance, the utterance *I need to be in New York on Monday by 9 a.m.* needs to be classified as a personal plan to fly to New York arriving before 9 a.m. The fact that the speaker represented his or her location before a time point is not taken into account. It is bypassed by utterance classifiers based on keyword spotters. The novelty in this approach lies in the fact that statements of personal desires such as *I want to do something* and commands such as *Do something!* may be given two different categories, which facilitates both utterance classification with keyword spotters and further processing.

In this research, as explained in Chapter 5, I make use of situational semantics. This means I did not implement a world model nor possible worlds. Moreover, the wheelchair also did not need to infer the user's beliefs, desires, and intentions to understand spoken commands for the collected linguistic data. Keyword spotters were also not used. In other words, a BDI agent architecture was not implemented.

2.2.5 Plan-based agent architecture

A myriad of plan-based architectures have been proposed. This approach to dialogue systems consists of two stages. First, utterances are classified for domain, then subclassified for user intent (dialogue act + representation). The user intent is taken as the planner's goal and the planner finds out whether the utterance is sufficient for reaching the intended goal. If they are not sufficient, disambiguation, referential grounding, clarification, and confirmation subdialogues are automatically triggered before the intended goal can be reached.

This process is reliable whenever each wording can be associated with a single domain and a single intent. However, this research is concerned with cases where the dialogue act, thus user intent, is strongly dependent on the situation, including who the interactants are, their social roles and relations with each other, and their corresponding duties and rights in the situation. A contextless utterance classification for domain and user intent based on wordings alone cannot be used in this case. A situational utterance classification is required.

A plan-based approach to dialogue management could potentially be merged with a move integrator such as that proposed in this thesis (see Chapter 13). This merge would, however, introduce no gain to the architecture regarding situational dialogue act determination because doing this with precision for a given situation requires solving reference to simple things and recognising actual and potential locations for them beforehand. The problem here is that the planner can only be activated for triggering subdialogues after it is given a goal. However, the challenge in situated dialogues is not triggering subdialogues, but to determine the goal that the wheelchair wants to achieve by planning material actions. In other words, when the planner starts planning future actions, all dialogue related processes are already done. Nonetheless, service planning is a task that the wheelchair needs to perform and planning is required for this

purpose. The sole thing that the wheelchair does not need do is to plan its moves in dialogue due to the fact that it just responds to commands and does not try to convince the user of something or get the user to do something.

2.3 My Contribution

In computational linguistics, currently no approach exists for determining a dialogue act based on both the wording and the contexts of situation and discourse.

With the objective of answering the question of how the wheelchair users' intent can be recognised relying on what they mean by their words in each given situation and discourse (Research Goal 1), in this thesis I propose a grounded approach to dialogue act determination. The proposed process of understanding an utterance has three steps. First, incomplete utterances, repairs, and utterances that do not result in a dialogue move are treated by a text producer prior to lexicogrammatical analysis of text (Chapter 8). Second, one or more semantic structures are built for the text produced or edited by the last utterance (Chapter 9). Finally, each semantic structure is integrated progressively into the situation in four stages: first, references to simple things and locations are integrated (Chapter 10); second, participant roles such as who does what to what are checked for feasibility considering object affordances, personal abilities, and personal rights and duties (Chapter 11); third, logical relations between what is represented and services that can be demanded are established (Chapter 12); and, finally, a dialogue move is proposed given the current state of discourse (Chapter 13). The first contribution of this thesis to computational linguistics is evidence that this approach for determining dialogue acts is feasible and reliable at least for controlling a wheelchair.

Chapter 3

Society and Words

Different theories of grammar result in different theories of how words relate to represented phenomena. In this chapter, I review formal and functional approaches to vocabulary description and show how multiword expressions (MWE) makes it difficult to model associations between lexical items and their meanings with current linguistic theories, including Systemic Functional Linguistics (SFL). Next, I review computational approaches for handling MWEs in current use and point out to their theoretical commitments, which are different from the ones I make in this thesis. Once the gap in SFL is filled (Chapters 5 and 6), we will be able to create a language-based taxonomy of simple things, locations, and processes to support the automatic understanding of spoken commands proposed in Chapters 9-13.

3.1 Theories of Lexis

There are two major approaches to describing the vocabulary of a language. The first is the approach adopted in formal linguistics, which includes Generative Grammar (Horrocks, 2013), studies concerning the innateness of the human language faculty in Psycholinguistics (Pinker, 1995), language descriptions supporting Montague's Semantics (Abbott, 2010), and, to some extent, most semantic lexica such as WordNet (Miller, 1995). In these works, lexis has the shape of a dictionary. The second approach comes from early work in anthropology (Malinowski, 1923, 1932, 1935, 1948) and sociolinguistics (Firth, 1936, 1950, 1951b,a, 1957a), in particular, the ones applying ethnographic methods, and later mainstream work in functional linguistics (Halliday, 1966a,b; Hasan, 1987; Halliday and Matthiessen, 1999; Halliday, 2003, 2005; Halliday and Matthiessen, 2014). Here lexis is described as a delicate system of options realising, amongst others, terms in a taxonomy of people, objects and processes. Particular emphasis is given to relations between multiple phenomena and a class of phenomena (*instance-of* and *type-of*), amongst phenomena (*part-of* and *whole-of*) and amongst classes of phenomena (*subclass-of* and *superclass-of*). A taxonomy comprises associations between terms and classes of phenomena, the associations between names and recognisable phenomena not being part of it. As a result, terms keep relations with others such as **hyponymy/hyperonymy** between *cat* and *animal* and **potential meronymy/holonymy** between *paw* and

cat. These term relations do not exist in dictionaries, though hyponymy and potential meronymy can be found in modern semantic lexica such as WordNet (Miller, 1995).

In the following, I provide a broad characterisation of these two approaches to vocabulary description and explain why the meanings of MWEs present challenges to lexis modelling independently of which approach is adopted.

3.1.1 Dictionaries in Formal Linguistics

In all main-stream work in Generative Grammar including both Standard Theory and Extended Standard Theory with X-Bar and Government-and-Binding as well as opposing traditions such as Generalised Phrase Structure Grammar and Lexical Functional Grammar (Horrocks, 2013) and most work in Logical Grammar supporting Montague's semantics (Abbott, 2010), **grammatical words** are taken as the entry point of description. Each word is classified based on solely distributional and morphological criteria. Typical **word classes** include, but are not restricted to *determiners*, *nouns*, *pronouns*, *adjectives*, *prepositions*, *verbs*, and *adverbs*. Phrases are classified bottom up based on the classes of their constituents.

Under morphological analysis, nouns such as *table* and *tables* share the same **stem** *table* and differ formally only for the presence or absence of the **ending** *s*. The noun comprising the stem *table* and the ending *s* is said to be *plural* and, in contrast, the one consisting of only the stem *table* – or the stem *table* and the ending *ø* – is said to be *singular*. Each **set of words** sharing a stem (e.g. *table*) and a word class (e.g. *noun*) counts as a **lemma**.

For a formal computational grammar, the construction of a lexicon consists of two steps. First, lemmas are collected from a text corpus or some other linguistic resource such as a paper dictionary. Second, in its simplest version, the vocabulary is modelled as a monolingual dictionary: each lemma in the resulting lemma set is associated with one or more alternative **concepts** or **composite descriptions**.

With the advent of computational linguistics, dictionaries evolved into semantic lexica such as WordNet (Miller, 1995). In these lexica, lemmas are members of one or more **synonym sets**, associated with a concept or a composite description. Modern semantic lexica keep relations between synonym sets such as hyponymy and potential meronymy.

3.1.2 Taxonomies in Functional Linguistics

In functional linguistics, the entry point for description is different. One starts by collecting a set of different wordings that can be chosen in a given situation. For illustration, let's assume a situation in which a mug tree and three empty liquid containers are present: namely, a wine glass, a water glass, and a tea cup. An English speaker would be able to tell someone to bring him or her one or more objects by uttering the following wordings:

(64) Bring me the wine glass!

(65) Bring me the water glass!

(66) Bring me the tea cup!

(67) Bring me the glasses!

(68) Bring me the cup!

(69) Bring me the mug tree!

The fact that the two present glasses includes both the wine glass and the water glass implies that both *wine glass* and *water glass* are subclasses of *glass*. In addition, the wordings *wine glass* and *water glass* include the word *glass*. These wordings represent glasses of different kinds while the word *glass* in these wordings represents glasses in general. This subclass-of relation between a wording and one of its constituents is taken as evidence for **semantic composition**, that is, evidence that the sign *wine glass* is composed of at least two other signs: *wine* and *glass*.

Assuming the same semantic composition, we would conclude that *tea cup* is a subclass of *cup*, not a synonym, even though these two wordings have the exact same effect in this situation. In contrast, the wording *mug tree* represents a mug tree, which cannot be represented by the word *tree* alone. Neither the fact that both *mug* and *tree* are potential words in English nor the fact that a mug tree has the shape of a leafless pine tree are relevant for discriminating objects in this situation provided that all interactants know what a mug tree is. So the wording *mug tree* can be treated as an atomic unit regarding its association with a class of things in the situation.

For this reason, the expressions *glass*, *cup*, and *mug tree* function as **classifiers** of the referenced objects whereas the words *wine*, *water*, and *tea* function as **subclassifiers**, that is, they specify a subclass of *glasses* or *cups* for discriminating the object[s] the speaker is referring to.

Functioning as classifiers, both words *glass* and *glasses* have the stem *glass*, but only the word *glasses* has the ending *es*. The word *glass* is chosen when referring to a single glass and the word *glasses* is chosen when referring to two or more glasses. In contrast, functioning as subclassifiers of glasses and cups, the words *wine*, *water*, and *tea* are neither singular nor plural. They do not vary according to the number of referenced liquid containers. Sets of words sharing the same experiential function (classifier, subclassifier, etc.) and the same experiential type (glasses, cups, wine containers, water containers, tea containers, etc.) are terms in a taxonomy.

Now let's assume a different situation. A person enters a wine store and there are several wine labels to select from. This prospective buyer can ask a vendor the following questions:

(70) Which wine do you recommend giving to a friend as a birthday gift?

(71) Which wines do you recommend giving to a friend as a birthday gift?

In such a situation, the words *wine* and *wines* represent wine labels from which the speaker might want to buy one or more wine bottles, the physical goods. These words function as classifiers of wine labels (not wine bottles). In turn, names for each wine label such as *Jacob's Creek Reserve* can function either as a reference to a wine label on its own as in *I'd recommend Jacob's Creek Reserve* or as a classifier of wine bottles as in *What about taking two Jacob's Creek Reserves?*

In these illustrations, the term *wine* that subclassifies glasses and the term *wine* that classifies simple things are two different terms, the former has a single form independently of how many glasses are referenced and the latter has two, one for a single wine label and the second for multiple. In a taxonomy, these two terms count as different entries. In contrast, the wording *Jacob's Creek Reserve* can be either a name for referring to a particular wine label or it can be a term for classifying wine bottles. Only the term is part of a taxonomy of simple things. The name belongs to a different lexical set for named entities.

In contrast, a dictionary or semantic lexicon would have a single entry for the lemma *wine* and this lemma would be associated with a single composite description such as *fermented juice of grapes* (taken from WordNet) for both cases. No information would be kept about the experiential contributions that words from this lemma have in different wordings and situations. It would be the task of an interpreter to map concepts or composite descriptions such as *fermented juice of grapes* for *wine* in *bring me the wine glass* to actual classes of simple things and instances thereof in a situation. Moreover, wordings such as *Jacob's Creek Reserve* usually do not have an entry in dictionaries and semantic lexica, even though they may function as classifiers for referents such as wine bottles. Again, it would be the task of an interpreter to map entity names to classes of simple things and their instances in a situation.

In this thesis, I aim at creating a language-based taxonomy of simple things, locations, and processes that can be integrated into a rule-based understanding module operating without an interpretation step. For this purpose, the approach to lexis description required must result in a map between terms and classes of simple things as in taxonomies, not a map between lemmas and concepts or composite descriptions as in dictionaries or semantic lexica.

3.1.3 Multiword Expressions (MWE)

A multiword expression comprises multiple words in the same wording that were inserted and ordered in a particular way to realise a single semantic contribution. For instance, let's consider the following statements about two animals in a picture:

(72) These are the animals.

(73) This is the prey animal.

(74) This is the animal of prey.

(75) This is the prey.

(76) This is the predator.

Both *prey animals* and *animals of prey* are *animals* as expected through semantic composition. *Prey animal* is the same as *prey* and *animal of prey* is the same as *predator*. Following functional composition, *animals*, *preys* and *predators* function as classifiers of living beings, the top class in the domain of biology, whereas both *prey* and *of prey* function as subclassifiers of *animals*.

However, there is more to functional vocabulary description than experiential contributions and experiential types. If alone, the word *prey* functions as a classifier of living beings. It only functions as a subclassifier of *animals* if it

precedes the classifier as in *prey animals* and *prey mice*. Moreover, the expression *of prey* also functions as a subclassifier of *animals* if it follows a classifier as in *animals of prey*, *beasts of prey*, *cats of prey*, and *birds of prey*. Adding to the complexity is the fact that *of prey* is a two-word expression that functions as a single subclassifier. So considering this a single experiential contribution, *of prey* counts as a multiword expression (MWE). This is the case because the experiential meaning is not carried by any of the two words *of* and *prey* individually nor by the two words in combination through a composition rule, it is carried by the presence of these two words in this relative order to each other after the classifier. Humans who fail to recognise semantic composition in such terms misunderstand written and spoken text (Couto-Vale and Heilmann, 2016) and so will machines.

In formal linguistics, a dictionary contains lemmas and each lemma is potentially associated with one or more word senses (or lemma senses). Such a dictionary does not have the necessary structures to store selections of words in a wording, their relative order, and their experiential contributions. As a result, linguists such as Pinker (1995) model wordings such as *prey animals*, *wine glasses*, *water glasses*, *tea cups*, and *mug trees* as "single words comprising multiple words", that is, compound words. No explicit treatment is given to wordings such as *animals of prey*, but a similar treatment can be assumed. Since contributions of semantic composition are not described, computational lexicographers adopting such an approach to create a semantic lexicon would need to list all potential compound words and corresponding lists of lemmas (**listemes**). In turn, they would need to associate each listeme to a different class of phenomena. For a small domain with little composition, listemes might suffice. For a general purpose robot, such listeme-class maps are not scalable for the following reason.

The number of combinations between n classifiers and m subclassifiers is $n \times m$ for two-word compounds and $n \times m_1 \times m_2 ... \times m_i$ for i sets of subclassifiers. For instance, the wording *student semester transport chip card* has 4 subclassifiers and one classifier. Each subclassifier such as *semester* belongs to a set with other items such as *year*, *month*, *week*, *day*, *3-hour*, and so on. The combination of items from the classifier set and each subclassifier set would result in an enourmous set of listemes and a corresponding enourmous set of atomic concepts. The process of creating such a large set of listemes by enumeration, associating each listeme with a different class of phenomena, and then determining subclass-of relations between the resulting classes is by no means efficient, both in terms of human labour and memory allocation. A compositional solution is preferable for a general purpose robot.

In functional linguistics, the mainstream tradition consists of adopting a rank structure, mapping groups and phrases to semantic elements such as simple things and locations, simple clauses to semantic configurations, and clause complexes to sequences of configurations. In a referential description of a simple thing, each constituent such as *animals* and *of prey* makes a different experiential contribution, that is, each contributes to the classification of the referenced thing in a different way. For this reason, constituents that function together as a complex classifier of a simple thing do not need to be treated together as a single compound word. Instead, they can be treated as combined items from different sets of options. Therefore, in such an approach, there is no compound word, no word that is part of another. However, though semantic composition

for referential descriptions is supported by a functional approach, no treatment has been proposed for terms realised by multiword expressions thus far. In this thesis, I aim at filling this theoretical gap, with emphasis in separable, reflexive and prepositional verbs in German (Goal 2.e).

3.2 Computational methods

In computational applications of formal linguistics, there are widely spread misconceptions about the contextual relation between lexis and semantics. The worst type of misconception is the one where researchers are not aware that multiword expressions (MWE) are simple signs and that simple signs can be determined only if we distinguish expressions that count as simple signs from those that count as sign complexes created through composition. Here is an example of how researchers argue in favour of discarding the compositional criterion when describing multiword expressions.

> The "compositional criterion" is a problematic concept in semantics, since it has been shown how difficult it is, in language, to define component parts, rules or functions involved in compositionality (Casadei, 1996) and, above all, that it is impossible to give words an absolute meaning independent from their context (Firth, 1957; Hanks, 2013). Because of this, the problem of subcategorizing the heterogeneous set of MWEs must be based on more reliable and testable criteria. (Squillante, 2014)

These researchers are right in one point. The situatedness of human adult language (Firth, 1957b) indeed hinders attempts to give words an absolute meaning independent from their context of situation and search corpora for multiword expressions in a contextless manner. For this reason, researchers performing corpus studies without keeping track of the situation in which texts are spoken or written indeed cannot describe how signs are composed of other signs and cannot find multiword expressions if we define them as a single sign realised by multiple words. However, the fact that their methods for finding and classifying multiword expressions must be based on "more reliable and testable criteria" is a shortcoming of any methodology that ignores the context of situation, not an issue with the compositional criterion when determining what counts as a multiword expression.

Another serious misconception is the one whereby researchers acknowledge the contextual nature of human language but assume that the meaning of a word varies solely depending on the words that occurred before and after it, that is, depending on the **context of wording** (a.k.a. co-text), ignoring what we can and cannot perceive around us.

Here is an example of paper where this reduction occurs:

> Computational metaphor processing refers to modelling non-literal expressions — e.g., metonymy, personification, idiom, simile, and metaphor — and is useful for improving many NLP tasks such as Machine Translation (MT). [For instance], Google Translate failed in translating *devour* within [the] sentence *she devoured his novels* into Chinese. [...]

> Metaphor identification is highly dependent of its context. Therefore, phrase-level models will miss the chance to identify phrases such as *climb the social ladder* where the phrase *climb ladder* appears literal out of context. In addition, parsing sentences into phrases causes another issue: the classifier predicts a label (metaphorical or literal) for the phrase, e.g., *kill process*, rather than identifying a single metaphorical word, such as *kill*; such identification is important if we want to interpret the metaphor. [...]
>
> Our metaphor identification framework is built upon word embedding, which is based on... (Mao et al., 2018)

Devouring books is indeed a way of *reading* books. It is not the same as physically *chewing* or *eating* them. It is also true that the *social ladder* is none of the *ladders* sold in online shops and that the action of *climbing* a *ladder* is not the same as the action of *climbing* the *social ladder*. Some people might think they are good at one and terrible at the other whereas other people might be skeptic about there being a *social ladder* in the first place. In contrast, no sane person in an industrialised society would doubt the existence of *ladders*. Therefore, from an epistemic perspective, the action of *devouring* books and the action of *climbing* a *social ladder* have different statuses. While *devouring* books can be treated as a directly observable human behaviour, *climbing* a *social ladder* cannot. It can be indirectly observed only if someone's vertical location on the *social ladder* is mapped to some other personal attribute such as that person's yearly brute income or the upper limit for that person's consumption of products and services in a particular calendar year.

For this reason, the mapping between terms and classes of things, locations, and processes around us is indirect only in the latter case. From a synchronic perspective, there are at least two terms *to devour* for human actions, one meaning *to enjoy avidly* in the case of books and the other meaning *to eat greedily/immoderately* in the case of food (definitions from Vocabulary.com). This means, *devouring* in *devouring books* has a single categorial meaning, that of *avidly enjoying books*, which is a directly observable behaviour. The mapping between terms and classes of things, locations, and processes around us is indirect only when the things, locations, and processes described cannot be directly observed around us, which is not the case of the action of *devouring* in *devouring books*.

In this regard, an intelligent wheelchair needs to understand reference to simple things around us not only when users count on the fact that the referent exists (exophoric reference) and when they count on the fact that the referent has been recently mentioned or shown (anaphoric reference), but also independently of whether the referent is directly observable such as people and objects or only indirectly observable such as the referenced distance between physical things in the examples below.

Representation of personal location
Congruent vs Metaphoric

1. Wir sind zu weit weg vom Waschbecken.
   We are too far from wash basin.

2. Unser Abstand vom Waschbecken ist zu groß.
   Our distance from the wash basin is too big.

References to simple things may be anaphoric or exophoric depending on whether the user counts on the contexts of discourse or situation and both anaphoric and exophoric references can be congruent or metaphoric depending on whether the referent can be directly or only indirectly observed. Failing to model such metaphoric references properly results in grammatical models that cannot be used to map grammatical to semantic structures in ranks.

Consequently, the fact that researchers cannot treat wordings such as *the social ladder* as indirectly observable referents using their semantic models is a limitation of their approach for describing semantic systems, not a limitation imposed by grammatical structures per se. Furthermore, the fact that an automatic process classifies *kill process* as a multiword expression is due to a bug in their tokeniser or parser, not something caused by the presence of grammatical structures in text.

In this research, the wheelchair had access to words and expressions associated with classes of things, locations, and processes. The task was not that of learning what counts as a multiword expression in a contextless corpus of utterances. It was that of recognising a known multiword expression in an utterance to understand its meaning in the current situation.

Aiming at achieving this goal, a few model-theoretic treatments of multiword expressions have been proposed in the literature, all of which — to the best of my knowledge — within the scope of formal linguistics and, most of which, with the above described misconceptions, thus they could not be used for the automation of command understanding in this study. In the following, I shall review two of them: namely **moves** for contact/syntactic structures and **catenas** for dependency structures. As mentioned above, no functional approach to multiword expression was available in computational linguistics as of the time of this study and still is not, as far as I know.

3.2.1 Moves in syntactic structures

In Generative Grammar (Horrocks, 2013), a syntactic or contact structure is a structure composed of adjacent constituents. For illustration, using the Stanford Parser (Stanford NLP Group, 2012), the resulting syntactic structure for the wording *Can you come?* is the one represented in Figure 3.1. The word *can* is classified as a **modal verb** (MD), the word *you* as a **personal pronoun** (PRP), and the word *come* as the main **verb** (VB). There are three composite syntactic structures: **noun phrase** (NP), **verb phrase** (VP), and **inverted sentence** (S).

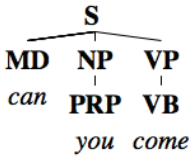

Figure 3.1: Syntactic structure generated by the Stanford Parser

In German, the intransitive verb for *coming* is *herkommen*, a separable two-word expression composed of the word *her* and the word *kommen*. Figure 3.2 shows a syntactic structure for the wording *Kannst du herkommen?* that is analogous to the English counterpart, which was created manually for illustration. The word *her* is classified as a **verb participle** (VB2) and the word *kommen* as a **verb base** (VB1).

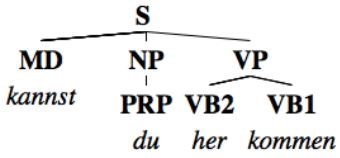

Figure 3.2: Analogous syntactic structure for German

Syntactic structures can accommodate multiword expressions such as *herkommen* well when these expressions are continuous. However, in wordings with similar meaning such as *Kommst du her?* multiword expressions are not continuous. For those wordings, the Standard Generative Theory of Language (Horrocks, 2013, Chapter 2) is not sufficient. In one of the extensions of his theory, namely Government and Binding (Horrocks, 2013, Chapter 3, p.99), Chomsky proposed an operation *move* α where α is any syntactic element that a child learns to move around within boundaries (Bounding Subtheory) to achieve a particular semantic contrast.

For this reason, the wording *Kommst du her?* would have not only a syntactic structure (S-structure), but it would also project the syntactic structure of the wording that was used as input for creating the current wording through element movement (D-structure). Given the fact that the wording *Kannst du herkommen?* is more frequent than *Kommst du her?* in our data, let us assume that the latter was created based on the former by removing the modal verb *kannst* resulting *Du herkommen?*, replacing the verb base ending *en* by *st* to create *Du herkommst?*, and moving the verb base *kommst* to the front to build *Kommst du her?*. The result of this process results in the syntactic structure of Figure 3.3, where the symbol ø is a place holder for the **basic position** of the moved word.

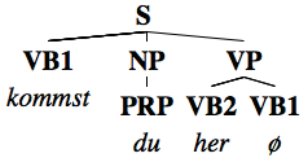

Figure 3.3: Syntactic structure with a word moved from its original position

The most up-to-date proposal to interpret multiword expressions in syntactic structures consists of taking a syntactic element as an input for lexical analysis and determining whether there is a multiword expression comprising of some of the constituents (Osenova and Simov, 2014). This solution presupposes that we

have a syntactic structure with movement links for moved syntactic elements such as the one above.

However, movement links between moved element positions and their basic positions can only be detected for these cases if we detect the multiword expression prior to building up the syntactic structure. For this reason, even if a syntactic structure with movement links can be used for determining the meaning of a multiword expression, it cannot be produced before we recognise multiword expressions. Because of this, though syntactic structures with movement links can store information about discontinuous multiwords, the fact that — to the best of my knowledge — no rule-based parser can create such structures at the moment prevents these structures to be used in dialogue system implementation.

3.2.2 Catena in dependency structures

A dependency structure consists of pair-wise tagged dependency relations between words in a wording. Figure 3.4 shows a dependency structure for the wording *Can you come?* produced by the Stanford Parser (Stanford NLP Group, 2012). In this wording, the word *can* is classified as an auxiliary (aux) of the word *come* whereas the word *you* is classified as a nominal subject of the word *come*. The word *come* is classified as the root of the wording.

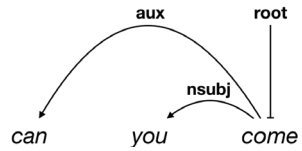

Figure 3.4: Dependency structure generated by the Stanford Parser

Dependency structures are an alternative to syntactic structures or other types of constituency structures. They capture long-distance dependencies such as the dependency between the auxiliary *can* and the word *come* that it helps.

Since they capture long-distance dependencies, they can store information about the relation between words in a multiword expression such as the **separable verb** *herkommen* in German. In Figures 3.5 and 3.6, I propose two analogous dependency structures for the wordings *Kannst du herkommen?* and *Kommst du her?* The word *her* is classified as a particle (part) for the words *kommen* and *kommst*.

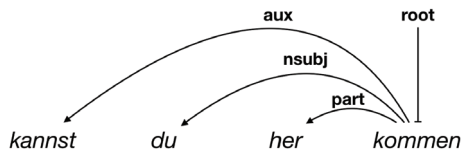

Figure 3.5: Analogous dependency structure for German

A multiword expression such as *herkommen* can be spotted in a dependency structure with a **catena** pattern (O'Grady, 1998) defined in the following terms.

Wording any continuous sequence of grammatical words

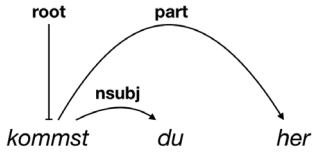

Figure 3.6: Discontinuous verb *kommst ... her* in dependency structure

Catena any chain including a word, zero or more dependents, and zero or more dependents of included dependents (chain of head-to-head relations)

According to these definitions, if we represent the words in *kannst du her kommen*, respectively, by the numbers 1, 2, 3, and 4, the following sets of words would be wordings and catenas:

Wording [1], [2], [3], [4], [1, 2], [2, 3], [3, 4], [1, 2, 3], [2, 3, 4], [1, 2, 3, 4]

Catena [1], [2], [3], [4], [1, 4], [2, 4], [3, 4], [1, 2, 4], [2, 3, 4], [1, 2, 3, 4]

A multiword expression is always a catena, but not necessarily a wording. For the above example, the separable verb *herkommen* is the catena/wording [3, 4]. For the example *kommst du her*, the separable verb is the catena [1, 3], which is not a wording.

Although we can easily spot a multiword expression in a properly built dependency structure, there is no general purpose parser at the moment which can provide reliable dependency structures for multiword expressions such as the separable verb *herkommen*. The reason for this is, again, simple, to provide a dependency structure such as the ones described above, the parser needs to recognise multiword expressions before building the structure.

In this thesis, I aim at providing a symbolic method for recognising and understanding multiword expressions in a wording (Goal 2.c). This method can be used to create dependency structures such as the ones described above or syntactic structures such as the ones described in the previous section. However, generating such structures is not necessary for the present study because my task consists of solely understanding the wording in its context and no module of the dialogue system developed utilises dependency or syntactic structures.

3.3 My planned contribution

In this thesis, I aim at proposing a functional model for one-word and multiword expressions in rank structures (Goals 2.b and 2.c). To comply with the requirements imposed by the goals of this research, this functional model needs to enable semantic composition (Goal 2.a) and be compatible with categorial and systemic functional models of vocabulary and grammar (Goals 2.e, 2.g). With this model, we shall have a single compositional semantic structure that is useful for both parsing and generation.

Part II

Construction of a Linguistic Resource

Chapter 4

Data Collection

Interaction with intelligent wheelchairs can only be considered "intuitive" if most commands people make to wheelchairs are readily undertaken and properly executed. To achieve this, instead of imagining what people would say to an intelligent wheelchair ourselves, it is an established practice in dialogue system development to collect an initial set of actual commands to intelligent devices uttered by prospective device users and take these commands as our initial linguistic data. During implementation, developers make sure that the intelligent devices respond properly to either all commands or a subset of the most frequent ones. In this chapter, I explain the process of collecting an initial corpus of user commands to an intelligent wheelchair prior to dialog system implementation by conducting a Wizard-of-Oz experiment. After reporting on the experiment, I describe the corpus created.

4.1 Wizard-of-Oz Experiment

Prior to this thesis, several experiments were carried out by researchers in I5-DiaSpace for collecting corpora of spatial language. These experiments included human-human interaction for placing objects within a doll house (Tenbrink et al., 2008; Ross et al., 2010), human-human interaction for taking an avatar to a virtual room, human-notebook interaction for the notebook to take the avatar to a virtual room (Tenbrink et al., 2010), and human-wheelchair interaction for path description (Tenbrink and Shi, 2007; Shi and Tenbrink, 2009). These experiments were designed so as to stimulate the largest variation possible in spatial language. The resulting corpora were then used for describing the ways in which space is represented in human languages and how spatial relations are negotiated in dialogue. None of these corpora included the **most typical** displacement commands that wheelchair users indeed enact in a routinely basis. The most trivial and typical ways humans represent space were actually discouraged in favour of a larger variety of expressions and wider German-language coverage. For instance, in the experiment where humans interacted with an autonomous wheelchair, the wheelchair informed the user that the location was unknown whenever the user told the wheelchair to take him/her to a destination – usually relative to a room – without providing further information about the path leading to that place (Tenbrink and Shi, 2007). The focus was placed in

how humans describe routes and not in how they tell wheelchairs to take them to places.

In this thesis, one of the central concerns is to automate the understanding of spoken commands that wheelchair users will typically give to wheelchairs, tackling the most typical linguistic phenomena. For this purpose, we cannot count on linguistic data coming from experiments where the wheelchair declares it cannot understand the simplest commands a user can come up with such as *kannst du in die Küche fahren* (*can you go to the kitchen*), *fährst du in die Küche* (*will you go to the kitchen*), *in die Küche fahren* (*go to the kitchen*), *fahr in die Küche* (*go to the kitchen*), *in die Küche* (*to the kitchen*), and *Küche* (*kitchen*). The reason why we cannot rely on such corpora is trivial. If the wheelchair tells the user it does not know where the kitchen is in an apartment, users lose trust on the wheelchair's knowledge and/or intelligence and they probably will not make most situation-dependent commands typical of situated dialogues. They will most likely fallback to other simple movement commands such as *turn right* and *go forward* as attested in previous studies (Tenbrink and Shi, 2007), and, if told not to do so, they will fallback to partial-route commands with complex referring expressions such as *the door at your right* in *go through the door at your right*.

What would make an interaction intuitive is not how many different kinds of spatial representation the wheelchair can handle well, but how frequently a spatial representation is understood. Therefore, it is important to know the actual variation in spatial representation that wheelchair users are likely to produce and if they will produce spatial representations at all. For collecting such a corpus, a different approach was required, an approach that aimed at collecting a corpus of service exchanges in an interaction with a wheelchair that is **purposeful**, **simple**, **easy** and **intuitive**. In the following sections, I describe these properties of interaction and how such an interaction was achieved in the Wizard-of-Oz experiment.

4.1.1 Properties of interaction

Four properties of interaction were pursued in the experiment: namely, purposefulness, simplicity, ease, and intuitivity. I shall explain them in the following in such a way that they become both intuitively understandable and objectively verifiable.

Purposefulness The purposefulness of an interaction between a human and a machine in an experiment is to be understood as the recognisable correspondence between the experimental interaction and the actual usage of a particular machine in a future situation from the perspective of the participants. This notion is similar but not equivalent to the notion of spontaneity required by studies in Conversation Analysis. Spontaneity in human-machine interaction would be a property of the interaction between a deployed machine and humans who want to use it for their own purposes. This spontaneity cannot be achieved in projects such as the development of a wheelchair that understands spoken commands, especially so in initial stages of development when deployment is unachievable. A human-machine interaction recorded in an experiment is only purposeful if participants commit themselves to making the experiment be successful by acting such as to inform the researchers how they themselves want the machine to

work in future situations. For ensuring purposefulness, researchers can explain an experiment to a participant in such a way that the participant can foresee the situation in which the technology will be used, consider it to improve people's lives, and become aware of the personal and social role they should act out in the experiment. If this is the case, participants will use the machine for a purpose that they made their own during the experiment and they will act out their role as a performing artist does in theatre. They may represent either someone else (imagined or known) using the machine now or themselves using the machine in a future situation as they foresee it. Otherwise, the interaction will lack a personal purpose from the perspective of the participant. Evidence of such a "co-designing" involvement can be seen in the collected data if participants direct utterances to themselves or researchers as in *Naja. Vielleicht geht's ja auch so.* (*Fine. It may just work out this way as well.*) and as in *Schade!* (*What a pity!*). These utterances directed to people other than the intelligent wheelchair can be seen as evidence that the way the wheelchair behaved mismatched how participants think a wheelchair should behave in that situation, but they may also offer evidence that participants are committed to making the project successful. A thorough analysis of participants' utterances directed to themselves and the researcher in the corpus of user commands collected in this experiment was carried out by Vales (2014).

Simplicity Simplicity of interaction consists of fully understanding user requests as they come and responding to users as frequently as possible in an adequate way. Simplicity of interaction from the addressee's perspective relies on a situational understanding of utterances and, in particular, reference. In contrast, a complex interaction consists of "bureaucratic" dialogues in which an incontestable contextless interpretation of a linguistic representation must be agreed upon and in which every doubt is checked in order to avoid mistakes. For instance, if someone knocks on the apartment door and the user asks the wheelchair to take him or her to *the door*, a complex interaction would be one where the wheelchair clarifies whether the user wants to go to the apartment door, the kitchen door, the bedroom door, or another door in the apartment. A simple interaction would be one where the wheelchair just takes the user to the right door based on the contexts of wording, discourse, and situation. In the experiments to be reported in this thesis, simplicity was achieved by making no clarification questions for utterances such as *Küche* (*kitchen*), *zur Küche* (*to the kitchen*), *fahr zur Küche* (*go to the kitchen*), and *fahr mich zur Kūche* (*take me to the kitchen*) and no clarification questions for utterances such as *pass auf!* (*watch out!*), *meine Beine!* (*my legs!*), and *stopp!* (*stop!*). In dialogue flow design, simplicity of interaction consists of systematically selecting the most purposeful interpretation of an utterance as its pragmatic meaning instead of starting a clarification dialogue unless there is a very strong reason not to do so, such as a prediction that irreversible damages might be unintentionally caused.

Ease Ease, on the other hand, is a property of situated interaction from the requester's perspective. It consists of the appropriateness of linguistic resources available to the requester for making a meaningful contribution to a particular interaction. A criterion for foreseeing whether an interaction will be easy is to enumerate the references to entities that must be negotiated for achieving the

personal goals of the interaction in the experiment and then to estimate the appropriateness (Mast and Wolter, 2013b,a; Mast et al., 2014b,a) of referring wordings for each reference in a given situation within the reference frame that the situation predicts (Vorwerg and Tenbrink, 2007). From the speaker's perspective, the intersubjective appropriateness of a referring wording is a conjoint derivate of the nomenclatory and discriminatory powers of a wording in a situation. The nomenclatory power consists of how certain the speaker is that an entity can be represented by its name or as an instance of an entity class by the associated term (see Chapter 6). The discriminatory power consists of how certain the speaker is that referring to that entity by its name or as an instance of a named class discriminates it from other entities in the given situation. From the listener's perspective, nomenclatory power can be understood as how likely the listener is to accept the associations between names and entities or terms and entity classes as standard. From the same perspective, discriminatory power can be understood as how likely the listener is to identify the referenced entity with a perceived or previously mentioned entity in a situation. The existence of appropriate referring wordings to be used in speech in a situated interaction is an indicator that grounding will likely be achieved. In experiment design, ease is achieved by controlling the number of similar things in a situation and the necessity of identifying particulars for the interactants to achieve their personal goals in the interaction[1]. A situation in which an interactant needs to refer to a particular door amongst many similar doors in a long corridor will make interaction difficult. In contrast, a situation in which an interactant needs to refer to the only door present will make interaction easy.

Intuitivity The last property of interaction that I pursued was intuitivity. Intuitivity of a human-machine interaction is proportional to the frequency in which a machine understands user utterances. It is not coverage in the sense of utterance types, but coverage in the sense of utterance instances. While simplicity consists of the way a particular utterance is negotiated and understood, intuitivity consists of the range of variation in utterances that a particular machine understands. If the machine understands utterances independently of which symbols are used and independently of their granularity, then the interaction is intuitive because interactants are not forced to use a limited set of wordings amongst those that are available to them in language. Methodologically, intuitivity of interaction is achieved by extending the coverage of understanding based on a corpus, giving precedence to the most frequent utterance types.

[1] In the field of referring wording resolution, perceivable entities that could be mistakenly identified as the referenced entity are known as **distractors**. Ease of interaction will not always be guaranteed by controlling perceivable distractors. In situated interaction, the same notion of distractors is to be applied not only to perceivable entities but also to other entities. For instance, a meaningful contribution may depend on a mentioned entity being identified as an entity mentioned previously in the dialogue as in *Ich möchte **es** auf dem Sofa lesen.* (*I want to read **it** on the sofa.*) or by a location where the wheelchair has been previously told to take the user as in *ich möchte **zurück zum Tish** fahren.* (*I want to go **back to the table**.*). This means that previously mentioned or shown entities and locations are also potential referents and grounds for representation and will affect how hard or easy the interaction is at any point in time.

4.1.2 Designing interaction

As explained in the previous section, the **purposefulness** of the interaction with the wheelchair is dependent on how well participants can correlate their interaction with the wheelchair in the experiment with an actual usage of the wheelchair in the future or elsewhere. Whatever tasks that experiment participants are assigned by researchers are necessarily tasks that they did not come up with on their own. So for interaction to be purposeful, users must be aware that researchers are designing an autonomous wheelchair and that their part in the design is to demonstrate how people from their linguistic community will truly speak with such a wheelchair.

Open experiments where researchers do not tell participants what to do with the wheelchair do not solve the issue of an interaction being purposeless because participants do not need or want to do anything in a laboratory. Even if they had the option of deciding where they wanted the wheelchair to take them inside a laboratory, they would not go around to accomplish their daily tasks. They would only play with the wheelchair as if it were a toy. In ordinary life, using a wheelchair is not an activity on its own: it is not a labour, it is not a work, and it is not a personal action in Arendt's terms (1958). For an interaction with a wheelchair to have any purpose, the interaction needs to facilitate or enable a human activity. In other words, it needs to facilitate or enable a labour such as washing one's hands, a work such as writing a thesis, or a personal action (self-initiative) such as reading a book on the sofa. A gait-impaired human may find it difficult to or be incapable of going around an apartment to perform such activities and may need a wheelchair for that. Because of this, my assumption during interaction design was that, if participants were put in the position of a gait-impaired person that needs to use a wheelchair to do a sequence of activities, they would be able to look at the current situation from a gait-impaired person's perspective and interact with the wheelchair in a similar way as a wheelchair user would do even if they are not gait-impaired themselves.

In that way, by assigning a series of actions for the participants to perform (instead of a list of displacements), I aimed at increasing the chances that participants understood the social roles that they needed to play in the experiment, that they made the purpose of the imagined person they interpret their own, and that they committed to making the experiment work out for its purpose: namely that of enabling the improvement of an intelligent wheelchair. When watching the recorded interactions, I had the impression that participants were indeed working hard to interpret gait-impaired people. For instance, many of them made a great effort to move from the bed or the sofa onto the wheelchair and vice versa without using their legs. Another evidence of commitment to the experiment was seen in the kinds of suggestions they made during a retrospective protocol. Many of the suggestions were related to the end position (orientation and location) of the wheelchair in relation to the bed, the sofa, the wash basin, and the prep area next to the kitchen sink, which was not perfectly functional. This serves as evidence that they wanted the wheelchair to work properly and were engaged in giving relevant feedback to researchers. In other words, they showed a co-designing behaviour.

Since the goal of the experiment was to collect a corpus, we used a wizard-of-oz experiment design. Dr. Dimitra Anastasiou interpreted the wizard. She could see the current position of the user and the wheelchair captured by strate-

gically positioned ambient cameras and heard the dialogue captured by a strategically positioned microphone. She could control whether the wheelchair should move or stand still, which was the next position the wheelchair should move to in a list of 7 predefined positions, and when the wheelchair should say which of 8 canned responses such as *guten Tag!* (*good afternoon!*), *ja!* (*yes!*), *ok, ich komme sofort* (*ok, I'm coming*), *ok, ich fahre dahin* (*ok, I'll go there*), and *ok, ich fahre dich dahin* (*ok, I'll take you there*). To guarantee that interacting with the wheelchair was kept **simple**, the wizard had no option to make clarification questions and had to accept all commands as they came, at the abstraction level in which they came. If a command such as *bring mich ins Arbeitszimmer* (*take me to the office*) had three possible positions, the wizard had to judge on her own which position was most reasonable for the current situation and pick one. If she was not sure whether an utterance such as *ich möchte ein Buch lesen* (*I'd like to read a book*) was a command or a statement of a wish to perform a personal action, she had to decide whether to undertake the command or acknowledge the statement by saying *ok* (*ok*). Systematically opting for interactional simplicity was only possible because no permanent damage can be caused by misunderstandings in such a situation. With this design, I counted on the fact that participants had always the chance of judging whether the wheelchair understood them correctly after a response. For instance, acknowledgements such as *ok* (*ok*), *dann mach es* (*then go for it*), and *stimmt* (*you're right*) imply that the participant's utterance was understood as a statement. If the participant did not mean to enact a statement and received an acknowledgement as a response to his or her request, he or she can enact a new directive request that is less likely to be mistaken for a statement. Arriving at the wrong destination inside a room is also no tragic happening. A repair command to the desired location can be enacted by the participant as soon as he or she notices that the wheelchair is heading to an undesired location. No permanent damage is expected to be caused by such misunderstandings when moving around inside an apartment so maximum simplicity can be achieved.

Another concern was that interacting with the wheelchair should be **easy** for the participants. The **ease** of interaction depends on the participants being able to rely on their linguistic habits, so having a physical device to interact with is a good start. To make common linguistic habits available, I chose to observe humans while they interacted with a *seeing hearing human-sized* device, namely, a wheelchair equipped with both a camera and a microphone. Alternative options included a non-corporeal all-seeing all-hearing entity such as a virtual agent, which has no size and has no current position in space, thus having no perspective. Another option would be a seeing hearing container such as an apartment, which would have a *global* perspective to the situation. Both the absence of perspective and a global perspective are unusual in our daily interactions with other humans, therefore they are likely to make interaction **harder**.

By equipping the wheelchair with a visible camera behind the user (inaccessible to the wizard), I aimed at guaranteeing that gestures and spatial relations such as *vor das Waschbecken* (*[to] in front of the wash basin*) had the same overlap of valid interpretations such as the intrinsic front of the wash basin, the relative front of it from the wheelchair perspective, and the absolute front of it given the room orientation. All of these interpretations render one and the same position since they do not conflict. In contrast, they might have conflicted if the

wheelchair saw the scene from a camera installed on the back upper left corner of the room. This would have created the necessity of grounding the perspective by saying unusual wordings such as *vor das Waschbecken von meinem linken Hand her*[2] (*to the front of the wash basin from my left hand*). The very absence of such complicated representations in the actual linguistic data suggests that the interaction might have been easy from the participant's perspective (notes in Section 4.6). Moreover, the fact that the wheelchair has about the same size as a human also makes the interaction easy. If participants were to interact with larger objects such as an apartment or a building, small changes in position such as *ein Stückchen weiter ran* (a little bit closer) might not have been represented. The larger the addressee is, the harder it is to refer to relevant objects in the given situation and to talk about functional positioning.

Furthermore, the wheelchair had a personal name – *Rolland* – and was the only wheelchair in the situation. This meant that the wheelchair could be called into dialogue with the vocatives *Rolland* (*Rolland*), *Rollstuhl* (*wheelchair*), and *Rolli* (*wheelchair*) and did not need to be called with general vocatives such as *hallo* (*hey you*) and *Entschuldigung* (*excuse-me*). In addition, the positions that the wheelchair needed to reach in order to enable human activity were representable as positions relative to single objects: i.e. *neben dem Bett* (*next to the bed*), *neben dem Waschbecken* (*next to the wash basin*), *neben dem Vorbereitungsbereich* (*next to the prep area*), *neben dem Schreibtisch* (*next to the desk*), *neben dem Sofa* (*next to the sofa*), and *neben der Wohnungstür* (*next to the apartment door*). All of these classes of things had a single instance in the situation. This is a very different situation from that of going around a university building where most doors are office doors, most corridors are full of office doors on both sides and are connected to other corridors full of office doors on both sides in very similar ways. Therefore, descriptions such as *the third red door on the left side that is right from the whiteboard* were not required from the interactants. In addition, the position that the wheelchair was to arrive at was any one of a set of positions that were representable both as a projection from a physical thing and as a position for performing an action. They were not positions on an empty field, on an open sea, or on a desert where Earth axis and celestial bodies become relevant. In those cases, one needs spatial expressions such as *to reach the city, you'll need to go north for six miles and then go northwest for another 10 miles.* Finally, relative locations were chosen so that locations would not be difficult to describe in relation to objects. All positions (location + orientation) enabled a human activity. They were not random positions that would demand representations such as *half a meter right from one meter in front of the wash basin facing the left wall.* In addition, the description of the enablable human activity did not depend on mentioning particular things such as *the second light bulb from left to right at the left back corner of the bedroom* or *the kitchen door to the corridor that gives access to the bathroom.* Since I made sure that linguistic habits available to participants were sufficient to coordinate actions, I expected that interaction would be easy. This was partially the case. More on that issue shall be discussed in Section 4.6.

Finally, since the interaction partner of experiment participants was controlled by a wizard, the interaction was as **intuitive** as it could ever be.

[2]Such an utterance did not occur.

4.1.3 Selecting routine activities

When deciding which activities were to be included in the experiment, I assumed that humans do not do activities such as *washing one's mouth*, *washing one's hand*, *eating*, and *reading books* in random order. Even if washing one's mouth is an end on itself, a hygienic labour, it may be simultaneously a preparation for eating breakfast. In the same way, cleaning one's hand is not only a hygienic labour after eating something with one's own hands, but also a preparation for holding and reading a paper book after touching greasy food.

In addition, some activities such as to eat something are further qualified depending on the time of the day. Having something to eat after waking up – i.e. having breakfast – is likely to be carried out when one is alone as picking something in the refrigerator and eating it at the prep area whereas having lunch may be carried out as going to a cafeteria or to a restaurant, heating up food in a microwave, or some other variation depending on individual dietary habits. Therefore, specifying that the gait-impaired person – who is supposed to be interpreted – lives alone and that this experiment wants to capture a morning routine of that person not only helps anchoring *having something to eat* on a human habit but also makes this activity something that needs to follow waking up.

Finally, since the scenario needed to correspond to the mourning routine of a single person living in an apartment in an institutional house (see Chapter 1), I included a care taker regularly visiting the apartment and disrupting whatever personally determined activity is taking place. The routine-disruptive and random nature of a visit allows it to be placed anywhere within the participant's routine. Of course, since this was an experiment, we chose the best point to interrupt the routine from a researcher's perspective and interrupted the routine of all participants at the same point. However, from a person executing the routine, the external interruption of someone knocking on the apartment door is never personally planned.

The resulting sequence of human activities was:

1. wake up
2. do a mouth wash
3. eat something
4. wash your hands
5. read a book on the couch
6. open the door whenever someone knocks

A human is *in bed* when one wakes up, *in front of the wash basin* when one does a mouthwash, *in front of the prep area* when one eats something, *in front of the wash basin* when one washes one's hands, *on the couch* when one reads a book, and *next to the door* when one opens it.

However, the target group of this research are gait-impaired humans. They either lie in bed or on the sofa or sit on the bed, on the sofa, or in the wheelchair. They usually do not sit on regular movable chairs because there is little gain in moving onto a regular chair and then back into the wheelchair. When sitting,

they move from bed into the wheelchair and back, they move from the sofa into the wheelchair and back, and they cannot move from bed onto the sofa and back, if the bed and the sofa are not next to each other. In this sense, two gait-impaired human seats (the bed, the sofa, or the wheelchair) need to be next to each other in a way that enables moving from one to the other whenever a human has a subtask of moving from one to the other. Moreover, the position that enables these humans do activities such as being in front of the wash basin in order to do a mouth wash is dependent on the position of the wheelchair they are sitting in. In this sense, the wheelchair has positions of its own that correspond to human positions that enable the performance of human activities by the seated human. Moreover, in addition to positions that enable moving between seats and performing activities, the wheelchair has a position of its own, namely it can stay on the charging station for recharging. In other words, the wheelchair has positions it needs to reach which are not correspondent to the positions of any activity by the human or by the wheelchair as in *zum Bett* (*to the bed*), it has positions that correspond to human positions that enable human activities such as *vor das Waschbecken* (*in front of the wash basin*), and it has positions that enable wheelchair activities such as *auf der Ladestation* (*on the charging station*).

This has two consequences: one is that the wheelchair may need to put itself in more than one position in order for the human to perform a single activity, e.g. going to the bed side for the human to move into it then going to in front of the wash basin for the human finally to do a mouth wash; the second consequence is that some of the positions of the wheelchair are reached not because of human activities but because of activities of the wheelchair on its own such as recharging. This means that, on the one hand, a human activity that is bound to one human position may result in more than one wheelchair displacements and, consequently, more than one displacement commands, and, on the other hand, that the routine must have not only human activities but also wheelchair activities.

The resulting sequence of human and wheelchair activities that reflect those constraints was:

1. wake up
2. do a mouth wash
3. eat something
4. wash your hands
5. read a book on the couch while Rolland recharges
6. open the door whenever someone knocks

Tables 4.1 and 4.2 list all displacements that were expected in the morning routine and Table 4.3 appoints the activities that demanded these displacements. Displacements WD2, WD3, WD4, WD5, and WD9 ended with the wheelchair in a position in relation to a necessary position of the user for performing the activity or a subtask while sitting in the wheelchair. Displacements WD1, WD6, and WD8 ended with the wheelchair in a position that enabled the user to move into and out of the wheelchair, respectively displacements HD1,

HD2, and HD3. Displacement WD7 ended with the wheelchair in a position where the wheelchair was to perform an activity, namely, that of recharging. In other words, all displacements ended in a position that was useful for something, so being in that position was not an end by itself.

Displacement	Source Location	Target Location	Human Location
WD1	charging station	bed	on bed
WD2	bed	wash basin	on wheelchair
WD3	wash basin	prep area	on wheelchair
WD4	prep area	wash basin	on wheelchair
WD5	wash basin	desk	on wheelchair
WD6	desk	sofa	on wheelchair
WD7	sofa	charging station	on sofa
WD8	charging station	sofa	on sofa
WD9	sofa	entrance door	on wheelchair

Table 4.1: Wheelchair Displacements

Displacement	Source Location	Target Location	Wheelchair Location
HD1	bed	wheelchair	next to bed
HD2	wheelchair	sofa	next to sofa
HD3	sofa	wheelchair	next to sofa

Table 4.2: Human Displacements

Activity	Displacement
2	WD1 HD1 WD2
3	WD3
4	WD4
5.a	WD5 WD6 HD2
5.b	WD7
6	WD8 HD3 WD9

Table 4.3: Displacements for Activities

4.2 Preparing instructions

By proposing the previous sequence of activities, I aimed not only at convincing participants that the experiment corresponds to a likely future routine of someone in Germany who interacts with such a robot when they become available,

but also at avoiding words that participants could reuse during the experiment. If I were to instruct experiment participants to wash their hands in the wash basin by referring to the wash basin, I would be increasing the likelihood that they would talk about that place as in *zum Waschbecken* (*to the wash basin*) and not as in *ins Badezimmer* (*to the bathroom*). In fact, for most of the instruction it was possible to avoid reusable references to things and locations with a single exception: it was not possible to prepare instructions without saying that the book reading should take place either *auf dem Sofa* (*on the sofa*), *auf der Couch* (*on the couch*), or *im Wohnzimmer* (*in the living room*). Likely because I chose the wording *auf dem Sofa* (*on the sofa*) to be part of the instruction, most participants made displacement commands for the wheelchair to go *zum Sofa* (*to the sofa*). One participant said *zur Couch* (*to the couch*) and revealed in the end that he chose a different word on purpose because he was testing whether the wheelchair could understand everything. No participant told the wheelchair to take them *ins Wohnzimmer* (*to the living room*) without further specification. For all other locations relative to pieces of furniture and utensils, which were not mentioned in the instructions, participants represented displacements both [in]to rooms and [up] to pieces of furniture and utensils. This is confirming evidence that, even if a researcher wants to make the interaction easy, mentioning things and locations as instances of named classes during instruction defeats the purpose of experiments aiming at capturing typical linguistic variation.

4.3 Participant selection

When selecting participants for an experiment with intelligent devices, one of the concerns is to control the selection of participants based on linguistic competence. A typical method is to ask for the participant's mother's tongue (or simply **mother tongue**), to ask for the participant's first language (or simply **first language**) or to invite **native speakers** only. All of these are ways of classifying people according to **linguistic nativity**, a notion according to which humans are born in linguistic communities in the same way as they are born in territories. In other words, if a person is born in an 'American English'-speaking community, they are born 'American English' speakers in the same way as being born in 'American' territory makes one be 'American'. This view implies that human newborns have a genetic endowment to learn the language of their linguistic communities and being born in a linguistic community is decisive in developing their linguistic competence. Such a nativity of language competence is put forth by a strand of linguistics that assumes that a linguistic community is somewhat or completely homogeneous and that being a competent 'language user' is to speak exactly like others.

As a side-effect of such a world view, some speakers of a language are classified as competent by birth – entitled of being creative and of speaking in untypical ways – and others as incompetent by birth and in need of mimicking natives. Moreover, such a world view is also overshadowed by a strand of bad linguistics that interprets language change as degeneration: i.e. word forms and lexical items may become 'unused' or 'misused' whereas new ways of meaning are taken to be 'borrowings', 'neologisms', and 'slangs'. Studies in such a line seem to be attached to a pursuit of some pure essence of a language, which is conflated with a pursuit of the essence of a homogeneous linguistic community,

in turn conflated with a pursuit of the essence of an isolated nation (reproduction group), in turn conflated with a pursuit of the essence of a particular human 'race'. This is not to say that all such studies conflated all such traits, but the rationale that supports the nativity factor is definitely one that attempts to partition humanity into disjoint linguistic communities without overlaps.

This whole idea that linguistic communities are disjoint and homogeneous could not be farther from what can be observed. The membership to linguistic communities is not given by birth. And the territory where one lives does not coincide with the birth territory. For instance, all EU citizen can *enter* and *stay* in any EU country and city (right to come and go, right to stay), live in any city of those countries (residence right), and have children with and marry any human on Earth (now human rights). This means that there is in practice no race, no isolated nation, and no homogeneous linguistic communities for children to be born into.

In addition, technological tools such as an intelligent wheelchair are often offered in multiple languages and their success depends not on how well these tools discriminate a 'purer' variant of a language but rather on how well they can cope with all variants of human language in a particular territory. In other words, the success of such a tool depends on it understanding all user utterances that take place when users interact with them. In this sense, it is possible to talk about a single 'human language' and its variations and, in this context, 'language switch' (or 'code switch') becomes a mere switch from a named set of meaning-making resources to another named set of meaning-making resources, all of which belong to an overarching human language.

Assuming this view, humans classify sets of meaning-making resources that they use for interacting with others (folk taxonomy) by giving names to those sets such as *Deutsch*, *Schwäbisch*, *Schweitzerdeutsch*, *Englisch*, *Türkisch* and so on. And they do this depending on their awareness of the heterogeneity of their linguistic community. Sets of linguistic resources that are associated with courses in the educational system and that are official languages of some country tend to be called **languages** while those that are not tend to be called **dialects**.

Taking this into account, I devised a method for selecting participants that is not based on linguistic nativity. It consists of two steps: an invitation and an online questionnaire. The invitation was written in the language of the experiment, which is readable only by those who can read the language, printed out and distributed on campus. The invitation informs we aim at collecting what people would actually say to a wheelchair that understands spoken commands in German (a named language in the folk taxonomy). With that formulation, we left to German speakers to choose whether they should be taken into account for what people actually say in what they perceive to be the German-speaking community. Finally, because I do not have access to how they have assessed themselves, I formulated a questionnaire that asks for the names of the languages and dialects that participants have used with their parents, with other family members, partners and people they lived with, within the school system and on the streets, friend circles, and city life (we did not include work experience because the subjects were university students living in Germany). With that, I collected some ethnographic classification of personal experience with human-human interaction, something I could go back to and rely on for, potentially, rewording or discarding some collected utterances that would reduce the quality of speech recognition or semantic accuracy of the wheelchair. The benefit of

doing this is twofold: 1) design decisions are made consciously and 2) the impact of those decisions on usability for those who need to use the product is known in advance.

For instance, this questionnaire was indeed useful for justifying the rewording of some utterances of one participant who said *zum Waschbechen* (sic) instead of *zum Waschbecken* (*wash basin*) and *komm hier* (sic) instead of *komm her* (*come here*). This participant had interacted in German only on the streets and at the university, possibly only in the last few years before the experiment took place. All other participants, even though they had different backgrounds, some of them never talking in German with their parents, seemed to share a lexicogrammar, i.e. to follow similar criteria for selecting lexical items and grammatical structures when telling the wheelchair what to do. No participant had to be entirely excluded based on a nativity factor.

4.4 Experiment run

From the student's perspective, each experiment run consisted of 10 phases. In the following I shall explain what happened in each one of them.

4.4.1 Invitation Flyer

Dr. Dimitra Anastasiou and I created invitation flyers in German that informed we were searching for German speakers for an experiment with an intelligent wheelchair. We also informed that the experiment would take approximately 1 hour and that participants would receive 8 euros and a small snack. Instead of 8 euros, participants could alternatively receive 2 credits if they needed experiment credits for their courses[3]. Finally, it informed that the registration for the experiment would take place at the web address *http://appengine.com/i5-diaspace*.

4.4.2 Website Registration

The website entered into more detail and informed that we wanted to collect what people would actually say to an intelligent wheelchair in German when they used it to go around an apartment. The website informed visitors about the address of the laboratory, the options of being paid in euros or quitting bachelor credits, and a form. The form had a field for name, a field for gender, another for age, multiple open fields for the language and dialects they spoke with their parents, with other members of their families including their partners and partners' families and people they lived with, at school and with people on the streets. Finally, it had a field for contact (telephone or e-mail) and a text area for notes. The registration website was online for two weeks.

4.4.3 Telephone Call

Dr. Dimitra Anastasiou called the registered students and scheduled an experiment session with each one of them. No extra information about the experiment

[3]The English Language Department at the University of Bremen makes it obligatory for students to take 3 credits in experiment participation so that students get to know what an experiment in linguistics looks like.

was given. Some students wanted to know if they needed the credits and Dr. Anastasiou informed them they had to check with their own department and that it varied depending on when they started the course.

4.4.4 Terms of Agreement

Our laboratory is located on the second floor of Cartesium, a university building in Campus. When participants arrived there, Dr. Dimitra Anastasiou greeted the students and asked them to follow her to the second floor. She introduced me to the participants as the experiment instructor. Most of the participants were students and most of the students attended my lectures in computational linguistics. This means they were familiar with me.

Right after they arrived, they were asked to sign a term of agreement stating that they *were not forced* (*wurde nicht gezwungen*) to participate in the experiment and that they *allowed* (*erlaube*) the recordings to be used in the context of scientific research. They were informed that the recordings would not be used out of that context. This information was written in the first person on a paper that they had to sign in the end as in *I acknowledge that I was not forced to participate in this experiment and that....* In addition to that, I explained what was written in the paper as in *when you sign this paper, you acknowledge that you were not forced to participate in this experiment and that....* Finally, I explained to each one of them that this was a formality that the university makes us do to guarantee that no one is ever forced to participate in experiments and to make sure that the recordings can be used for the research afterwards.

4.4.5 Apartment Tour

Following this, the participant was presented the Bremen Ambient-Assisted Living Laboratory (BAALL). This laboratory was built in the form of a fully functional apartment and was conceived as a research environment for usability tests of commercial and prototype appliances. It is equipped with monitoring cameras that enable researchers to see what is happening in the laboratory remotely and also to record an experiment run for future processing.

For the apartment tour, Dr. Dimitra Anastasiou and I wrote a presentation script with the following constraints. We wanted to know what states, actions, services, locations and things participants would represent and how they would represent them. For this reason, we avoided referring to those phenomena in ways that could be re-used in the experiment. For instance, if a person would need to tell the wheelchair to go to the wash basin, we opted to refer to a location in relation to that object not as *on the wash basin* during the apartment tour, but as *here (point to a location on the wash basin)*. In this way, participants did not have access to the way researchers classified these phenomena nor the class names that researchers might have 'taught' to the wheelchair. Example 4.4 shows the script for the apartment tour.

The script for presenting BAALL was designed in order to avoid using the lexical items that participants would need during the experiment so as to avoid priming the lexical item choice. I rehearsed the script until I could repeat it *ipsis litteris* convincingly as if I were not repeating a script. The informality in the language is intentional since the formality level I used with students I knew by both first name and last name was *Du*. Due to social conventions,

Dies ist unser Labor. Dies ist auch eine Forschungswohnung. Die wurde gebaut, um Technologien fr Menschen mit motorischen Behinderungen zu testen. Sie ist wie eine normale Wohnung aufgeteilt. Hier ist ein Ort, wo man schlafen kann (berhrt das Bett). Hier sind Bcher (berhrt das Regal). Ich lege ein Buch hierhin fr das Experiment (legt das Buch *Merlin* auf den Tisch). Ich erklre dir gleich, was deine Aufgabe ist. Komm mal mit! Hier ist eine Mundsplung fr das Experiment (berhrt die Flasche Mundspllung) und hierhin stelle ich einen Becher fr dich (legt den Becher auf das Waschbecken). Es ist auch fr das Experiment. Hier gibt es Wasser (ffnet den Wasserhahn). Dieser Raum wurde so gebaut, dass man mit einem Rollstuhl reinkommen kann. Siehst du? (deutet mit beiden Armen) Es gibt eine Tr da (zeigt auf die Tr), aber die lassen wir auf. Hier ist eine Dusche und hier ist eine Toilette (zeigt auf die Dusche und Toilette). Komm mal mit! Hier gibt es Essen! (berhrt den Vorbereitungsbereich) Die sind lecker! (schaut auf Gummibrchen) Das ist auch fr das Experiment. Du kannst sie essen, wenn du willst. Aber noch nicht. Das ist der Khlschrank (berhrt den Khlschrank, macht ihn auf). Was ist hier? Uh! Das ist neu (berhrt eine Bierflasche, macht den Khlschrank zu)! Hier gibt es noch einen Tisch (berhrt den Tisch). Und hier werden wir ein Fernseher spter aufhngen (mahlt ein Viereck mit dem Finger in der Luft vor einer Wand nach). Die Wohnung ist noch nicht komplett, es fehlen noch Sachen.	This is our lab. This is also a research apartment. It was build, to test appliances for people with motor deficit. It is organised like a normal apartment. Here is a place where we can sleep (touches the bed). Here there're some books (touches the shelves). I'm putting one here for the experiment (puts the book *Merlin* on the desk). I'll tell you what your task is in a moment. Come with me! Here there's some mouth wash for the experiment (touches the mouth wash bottle) and here I'll leave a plastic cup for you (puts a plastic cup on the wash basin). It's also for the experiment. Here there's water (tries out the tap). This room was built in such a way that we can enter with a wheelchair. See? (makes a two arm swing gesture) There's a door there, but we'll leave it open (points at it). Here there's a shower and here there's a toilet (points at them). Come with me! Here there's some food (touches the prep area)! Pooh, yummy (looks at gummy bears)! This is also for the experiment. You can have some if you want. But not yet. This is the fridge (touches the fridge, opens it). Let's see what is inside! Wow! This is new (touches a beer glass, closes the fridge)! Here there's another table (touches the table). And here we'll place a tv set in the future (point-draws a rectangular region on the wall). The apartment is still not complete. It still needs some extra things.

Table 4.4: Script for the apartment tour

I would not be able to switch the formality to *Sie* during the apartment tour for the students I knew personally and the small age difference allowed me still to perform the instruction with *Du* for unknown students. I strategically used causal clothes to make informal modality more expected. As I am a foreigner and have a strong accent, I assume that the possibly felt inadequacy of *Du* formality level and expressions such as *Komm mal mit!* (*Come with me!*) are likely to be attributed to the accent. The script can be seen in Example 4.4.

4.4.6 Wheelchair Demonstration

After showing the apartment, the participant and I arrived at the living room where the wheelchair was waiting for us. To present the wheelchair, Dr. Dimitra Anastasiou and I wrote a second script. The demonstration should give the impression to participants that the wheelchair understood us properly with spoken commands and pointing gestures. The script is presented in Example 4.5. Both the wheelchair Rolland and I as a researcher had turns in the dialogue. When pointing to the region that Rolland was supposed to reach, I made a relaxed arm shape and a relaxed hand shape point-drawing a circle on the floor, an approximate area. Both the fact that I point-draw a circle on the floor and the fact that the gesture was not precise were intentional. With this, we made sure that the user did not immitate my pointing gesture when interacting with Rolland because there is no situation in which such pointing gestures were useful. Touching things for indicating them (often used the instruction scripts) was also not performable by a person sitting on the wheelchair. All other pointing gestures (such as pointing to the participant's feet and to the shower) were made with relaxed hand and in a very imprecise way, again intentionally so that these gestures were not taken as a muster for gestures in the experiment. Dr. Anastasiou reported on the gesture corpus (Anastasiou, 2011; Anastasiou et al., 2012), what goes beyond the scope of this research.

As for the commands *Rolland, kannst du dahin fahren?* (*Rolland, can you go there?*), *Danke, Rolland!* (*Thank you, Rolland!*), *Kannst du zurck zu deiner Ladestation fahren?* (*Rolland, can you go back to your charging station?*), I opted for a relatively strong politeness level for *Du*-references, expecting that this would show the wheelchair can understand commands of the same kinds that humans understand and not only *fahr dahin* (*go there*). Since one of the goals of this thesis was to collect a corpus with variation in the interpersonal component of language, this seemed reasonable. Another important point was to use a lexical item for charging station. Since we did not have a charging station for Rolland to dock on, we had to tell participants that Rolland thought the carpet was a charging station, otherwise participants would not know that they could mention the carpet as an instance of charging stations. Parts of the script such as *er ist sehr intelligent* (*it/he is very intelligent*) were meant to be ironic since they followed a reference to an entity that the wheelchair "believes" to be a charging station, but which one can see is not. It was spoken with a laughing voice, ironic smile, and a negative head shake. The intention of such a comment was both to acknowledge that the wheelchair does not behave exactly like a human (robotic voice, utterances that are clearer than what one would expect from a human, delayed response, poor awareness of obstacles such as feet, programmatically overwritten perception etc) but also to state that the wheelchair can do at least what we want it to do for current purposes.

Finally, the two displacements of the wheelchair during the demonstration were also functional but in a way different from the ones that would take place in the experiment. Whereas wheelchair displacements during the experiment were executed so as to enable the user or the wheelchair to perform actions (they were embedded as a subtask of an activity), those during the demonstration did not need to enable any action, their purpose was to demonstrate that the wheelchair understands spoken commands. In those cases, the wheelchair staged the movements. Both what is represented in such commands linguisti-

cally *kannst du dahin fahren* (*can you go there*) and the positions reached are different from those that the experiment demands from participants. In this way, I assumed such a command would not prime participant strongly. Indeed, we have evidence it did not. More on this shall be presented in the results.

As well as previously, every single script line and gesture was rehearsed until I could play my role naturally without relying on a paper.

Und das ist Rolland (berhrt ihn). Dieser Rollstuhl ermglicht behinderte Menschen, sich in der Wohnung zu bewegen. Er kann uns sowohl hren als auch sehen; man spricht mit ihm durch diesen Mikrofon (haltet Kopfhrer mit Mikrofon). Wenn ich das trage... (setzt die Kopfhrer auf) gut, jetzt kann er mich hren. Und lass ich die Kamera einschalten (schaltet die Kamera ein) er... (wartet, bis die Kamerarotlicht blinkt) kann uns durch diese Kamera sehen. Gut, sie ist an. Er kann Gestik verstehen. Guck mal! (wendet sich an Rolland) Rolland, kannst du dahin fahren? (zeigt einen Bereich auf dem Boden) – Ja, ich fahre dahin! – (wendet sich an den/die Teilnehmer/in) Er kann uns gut verstehen. Achtung! (zeigt auf die Fe des/der Teilnehmers/in) Ein Schritt zurck! (wartet, bis Rolland zum gezeigten Ort kommt) (wendet sich an Rolland) Danke, Rolland! Kannst du zurck zu deiner Ladenstation fahren? – Gerne, ich fahre dahin! – (wendet sich an den/die Teilnehmer/in) Er denkt, seine Ladestation ist hier (zeigt auf einen Teppich). Er ist sehr intelligent!	And this is Rolland (touches it/him). This wheelchair enables disabled people to move inside the apartment. It/he can both hear and see us. We speak with it/him through this microphone (holds a headset). When I wear this... (puts the headset on) Good, now it/he can hear me. And... let me turn on the camera... (turns the camera on) it/he... (waits for camera red light to blink) can see us through this camera. Good, it's on. It/he can understand gestures. Watch this! (looks at Rolland) Rolland, can you go there? – yes, I'll go there. – (looks at participant) It/he can understand us. Watch out! (points at the participant's feet) A step back! (waits for Rolland to reach the indicated place) (looks at Rolland) Thank you, Rolland! Can you go back to your charging station? – Of course, I'll go there! – (looks at participant) It/he thinks its/his charging station is there (points at a carpet). It/he is very intelligent.

Table 4.5: Script for wheelchair demonstration

4.4.7 Purpose Construction

In order to guarantee that the interaction was going to be purposeful, the strategy I adopted was to ask the participants to put themselves in the position of a person who cannot walk and to play that person's role in a morning routine. The wheelchair was verbally represented as a person during the wheelchair demonstration: i.e. as something that can hear and see us not only sensorially (feel) but also perceptually (recognise). In addition, it was demonstrated that it can really understand both what we represent verbally and what we show. In other words, the wheelchair was presented not only as a person (a plausible sender and a plausible addressee), but also as a German speaker. During the

purpose construction phase, participants were asked to consider they were a particular kind of person. It was assumed they knew that the character they were to interpret spoke German since the whole interaction was in German up to that point and Rolland is a German speaker. The character was represented as someone who cannot walk and someone who not only *depends* on *a* wheelchair to move around but also *uses the* wheelchair in their daily routines. This means that the participant and the wheelchair had a standing usage relationship, in which the participant is *Rolland's user* and *a wheelchair user* and in which the wheelchair is *the participant's tool for going to places* and *a tool used by the gait-impaired for going to places*. In other words, though interactants were people, they were represented in the script as *a gait-impaired person/a wheelchair user* and *a tool used by the gait-impaired for going to places* for three reasons: first so that participants could foresee that there will be situations that will instantiate this situation type (register), in the second place so that they had a good understanding of the character they were to interpret and the character they were to interact with, finally so that they realised that this experiment had the potential of helping others and acted their parts as well as they could foresee the future situations that the experiment run represents. In this sense, experiment participants were invited to be seriously engaged as interpreters not only of themselves in a personally undesired future but also of other German speakers in the predictable future. With that in mind, I assumed that the interaction would be purposeful in the sense that participants would make their character's interactional purpose their own.

In diesem Experiment wollen wir herausfinden, wie gut es funktioniert, wenn jemand den Rollstuhl im Alltag benutzen mchte. Stelle dir jetzt bitte vor, du knntest nicht laufen und wrest auf dem Rollstuhl angewiesen. Jetzt geht es darum, einen ganz normalen Tagesablauf zu beginnen. Dabei brauchst du den Rollstuhl, um verschiedene Dinge zu tun. Komm mal mit! Wir fangen das Experiment hier an (berhrt das Bett). Du kannst dich hinsetzen. (wartet bis sich der/die Teilnehmer/in hinsetzt) Gut!	With this experiment, we want to find out how good our technology works when people actually use the wheelchair in their daily routines. Please imagine now that you were not capable of walking and depended on a wheelchair. Your task now is to start a normal daily routine. And for that you need the wheelchair in order to do different things. Come with me! We shall start the experiment here (touches the bed). You can sit here. (waits for participant to sit on the bed) Good!

Table 4.6: Script for making the interaction purposeful

In addition, by telling the participants that the character to be interpreted was *Rolland's user*, I seem to have constructed a tacit social contract between each participant and the wheelchair. This means that a particular gait-impaired human was the user of a particular intelligent wheelchair. If more than one gait-impaired humans and more than one intelligent wheelchairs were in the same physical area, it would be the social contracts between particular humans and particular wheelchairs that would prevent an utterance such as *I would like to go to the kitchen* to be interpreted by all wheelchairs as a command to take this person to the kitchen. With this representation, I assumed participants would feel entitled to ask the wheelchair to take them around with politeness levels

such as *kannst du mich zum Bad fahren* (*can you take me to the bathroom*). If the tacit social contract were not established, such a command might be understandable as rejectable, i.e. open to rejection. If I had not controlled for such a precise register construction, small variations in meaning potential could have made the proportion of interpersonal modalities not represent what is likely to happen when the gait-impaired use intelligent wheelchairs.

Nonetheless, although it seems to be the case that most participants overtook the interactional purposes of future users, two participants seem to have done so in different ways. One participant told me after the experiment run that he "tested" whether the wheelchair understood everything by choosing every time a new structure that he thought the wheelchair might not understand. Indeed his commands vary more than those of other participants, but his commands were not unique to him, they still coincided with what others did. This means he played his role within what other users considered they would do in the same situations.

Another participant played the role of a gait-impaired human using a wheelchair while playing the character of James Bond. He made a gun-shape with his hands throughout the experiment and pointed it to left and right whenever he entered a new room. These gestures were not interpreted by Dr. Dimitra Anastasiou (the wizard-of-oz) as commands to make a turn or to do something else. They were understood as gestures that a wheelchair user made while interpreting James Bond. Though many gestures and statements he made were unrelated to the interaction, the commands that were actually made to the wheelchair by this participant were also not unique to him. They were similar to what other participants did in the same situations. He told he would love to have such a wheelchair if he could not walk and that he felt like in a James Bond movie. I suppose that such acting behaviour[4] might pose some difficulties for wheelchairs that interpret gestures as commands.

4.4.8 Tasks Explanation

With the participants sitting on the bed, I listed the tasks that they should do in their morning routine twice. They were told to do them in that particular order. I told them they could ask me for the next task if they forgot the sequence, but none of the participants forgot the tasks. I attribute this to the fact that the tasks were logically related in the sense that most activities could be understood as a preparation for the next and that the sequence of action was typical for a morning routine. I also told them to use the wheelchair to go around in the apartment and not to walk to places since they were interpreting someone who cannot walk. Example 4.7 contains the script of the instructions.

4.4.9 Last Instructions

Finally, I explained that the closing of the door would be a cue for the experiment to start, that the participant could stop the enactment of the character at any time, that I was going to stay in the apartment the whole time and that, for security reasons, they should not stand up while the wheelchair was moving. I also told them that, if they wanted to stop, I would come and turn the wheelchair immediately off. No participant interrupted the experiment.

[4] In his case, acting as a gait-impaired person acting as James Bond.

Ich werde dir eine Liste mit den Dingen lesen, die du fr dieses Experiment tun sollst, und zwar genau in dieser Reihenfolge. Erst sollst du dir den Mund aussplen. Du hast schon den Becher gesehen. Dann nimmst du etwas zum Essen. Dann wschst du dir die Hnde. Dann holst du dir das Buch und liet es auf dem Sofa; whrenddessen soll Rolland auf der Ladenstation stehen bleiben. Dann wird jemand auf diese Tr klopfen, und du sollst sie ffnen. Sie ist nicht elektronisch wie die anderen. Deswegen muss man sie selbst aufmachen. Also alles noch einmal: erstens dir den Mund aussplen; zweitens essen; drittens dir die Hnde waschen; viertens dir das Buch holen und es auf dem Sofa lesen lass bitte Rolland aufladen, whrend du das Buch liet; fnftens die Tr ffnen, wenn jemand klopft. Und dann kommt das Experiment zum Ende. Du sollst das alle mit dem Rollstuhl machen, ohne zu laufen. Bitte versetze dich genau in die Lage eines Rollstuhlbenutzers. Du kannst also ganz normal mit Rolland kommunizieren, ganz so wie es dir am besten passt. Er ist sehr intelligent! Er kann mich gut verstehen! Ich halte bei mir die Liste von Aufgaben und ich kann sie dir immer zeigen, wenn du willst. Ok?	I'll read a list of things that you should do for this experiment and please exactly in that order. First you should wash your mouth. You've seen the plastic cup. Then you'll take something to eat. Then you'll wash your hands. Then you'll take the book and read it on the sofa; meanwhile Rolland should stay on the charging station. Then someone will knock on this door and you should open it. This one is not electronic like the others. For this reason, you'll need to open it yourself. So, let's go over this once again. First, wash your mouth; second, eat something; third, wash your hands, fourth, take the book and read it on the sofa – please let Rolland recharge while you read it; fifth, open the door when someone knocks. And then the experiment is over. You should do all these things with the wheelchair without walking. Please put yourself really in the position of a wheelchair user. You can communicate with Rolland in the most comfortable way, however you feel is more comfortable to you. It/he is very intelligent! It/he can even understand me! (emphasis on me) I'll keep a list of tasks by me and I can show it to you at anytime. Ok?

Table 4.7: Script for explaining tasks

4.5 Collection

The **German corpus of displacement commands for Rolland** (German-CDCR) was collected during this Wizard-of-Oz experiment that simulates a real-life everyday scenario of wheelchair usage within BAALL. In this experiment, 20 German speakers of both sexes were told to execute 6 tasks using the intelligent wheelchair. For achieving that, they had to drive the wheelchair to 9 destinations through free spoken commands. 13 participants were able to finish the experiment without any hardware failure of the wheelchair. All commands were recorded and manually transcribed twice constituting a corpus of 135 clauses with two transcriptions each. Finally, these clauses were read out loud with pauses in between by a female German speaker in a silent room for better sound quality and recorded again for further processing.

Also... sobald ich die Tr schliee (points to the apartment door), werden die Kameras aufnehmen (points to the cameras) und wir knnen dann mit dem Experiment beginnen. Wenn du mit dem Experiment aufhren willst, sagst du mir Bescheid, aber bitte stehe nicht vom Rolland auf. Ich muss erst ihn komplett ausschalten, bevor du aufstehen kannst. Das ist eine Sicherheitsmanahme, wenn man Experimente mit Fahrzeugen einfhrt. Ich bleibe die ganze Zeit bei dir um den Ablauf zu beobachten, aber bitte tue so, als ob ich nicht hier wre. Ok? Jetzt kannst du mit den Aufgaben beginnen. (schliet die Tr)	Now... as soon as I close this door (points to the apartment door), these camareas will start recording (points to the cameras) and we can start the experiment. If you want to quit the experiment, just tell me, but please do not stand up from Rolland. I need to turn it off completely before you can stand up. This is a security procedure for experiemnts with vehicles. I will stay the whole time next to you to watch the process, but please pretend that I am not here. Ok? Now you can start your tasks. (closes the door)

Table 4.8: Last Instructions

4.5.1 Retrospective Protocol

After concluding the tasks, students were asked to narrate what happened during the experiment. With this approach, I was able to collect how imperative utterances indeed relate to their indicative counterparts. Finally, they were asked to inform what could be improved in the wheelchair. What they said was used as a triangulation support and reported throughout this thesis as arguments in favour of one interpretation of the experiment over others.

4.6 Corpus of Spoken Commands

The **German corpus of displacement commands for Rolland** consists of the video recordings of twenty experiment runs from the perspective of Rolland. Two types of transcription were realised for two different purposes: namely a speech transcription from which speech segments can be extracted for text production (see Chapter 8) and a multimodal transcription with pause marks for multimodal analysis. A multimodal analysis of the second transcription was carried out by Vales (2014). For each speech transcription, linguists produced the corresponding grammatical text most likely intended by the experiment participants.

The resulting text was divided into simple clauses such as *fahr mich in die Küche* (*take me to the kitchen*) and interjections such as *Küche* (*kitchen*), *meine Beine* (*my legs*), or *naja* (*well*). Each simple clause and interjection was analysed regarding its contributions to the interaction. The details of each analysis will be given in the respective chapters where I describe how different contributions were integrated to the interaction.

4.7 Conclusion

In this chapter, I described the design and execution of a Wizard-of-Oz experiment carried out in a laboratory where different participants interacted with a remote-controlled wheelchair assuming that it was intelligent. Experiment runs were recorded, transcribed and analysed in different ways for different purposes.

In the following chapters, I shall describe how each clause and interjection type was further analysed at several steps so that the wheelchair could understand what its users say.

Chapter 5

Ontology Creation

For an intelligent wheelchair to understand an utterance as a symbol composed of other symbols, that is, as a semantic structure, it needs to have a list of all atomic symbols that can be chosen and how they can be put together as constituents of a composite symbol. Such a list of symbols and the corresponding compositional restrictions is known as a **linguistic ontology**.

For producing a **linguistic ontology**, one must analyse a corpus of utterances such as the one reported in Chapter 4 and take note of the phenomena described and the words used for description. Typically, experiment participants describe simple things such as a *kitchen*, locations such as *in the kitchen*, and actions or services such as *taking someone somewhere*. As for semantic composition, represented phenomena such as *the wheelchair taking the speaker into the kitchen* are composed of more elementary phenomena such as the action of *taking someone somewhere*, the *wheelchair*, the *speaker*, and the location *in the kitchen*. In turn, this location is composed of the *kitchen* and a spatial relation of *containment* (Bateman et al., 2009). The structure of such composite phenomena is closely related to the structure of the composite symbols that represent them. For this reason, this isomorphism between experiential structures and semantic structures shall be our guideline for achieving a rank scale according to which semantic composition will be described in such a way that each rank of experiential structure corresponds to a different rank in semantic and grammatical structures (Goals 2.a and 2.b). However, the objects of perception and description are not only the directly observable entities around us.

In the following, I define the objects of interactants' perception and description that fall within the scope of this study, explain the type of description logic adopted for creating the linguistic ontology, and then I list and define the actual symbols added to it.

5.1 Defining the scope of study

In the literature on logics and reasoning, there is no common usage of terms such as **universe**, **world**, and **situation**. As a consequence, there is little to no way to determine whether or not something falls within the scope of a study or not. For instance, what does **context of culture** and **context of situation** mean if we do not have a definition of world and situation? What is the difference

between **material**, **physical**, **ontic**, and **socially constructed** worlds if we do not define these notions? In the following, I shall define all these terms for the purpose of clarity and I shall define the scope of this study in such terms. The basic terms are the following:

Matter that which occupies space, has mass, and can be directly observed.

Material Universe all existing matter as a whole.

Material World the portion of existing matter that has been either observed by interactants and/or reported to them by others, that is, the portion of existing matter assumed to exist by interactants.

Material Situation the portion of existing matter whose relative position to interactants is known to them, that is, the portion of existing matter that participants can point to with their index fingers.

Bodies (greek Physeis) detached portions of matter that can move on their own relatively to other detached portions of matter.

Physical Universe all existing bodies taken together.

Physical World all bodies that have been observed by interactants and/or reported to them by others, that is, those existing bodies assumed to exist by interactants.

Physical Situation all bodies whose relative position to interactants is known to them, that is, those existing bodies that interactants can point to with their index fingers.

Entities (greek Ontoi) bounded portions of matter including not only bodies, but articulated members of bodies, delimited parts of bodies, and groups of entities, as long as they carry some intrinsic or extrinsic attribute as a unit. As a result, the same portion of matter is part of potentially multiple entities.

Ontic Universe all existing entities taken together.

Ontic World all entities that have been observed by interactants and/or reported to them by others, that is, those existing entities assumed to exist by interactants.

Ontic Situation all entities whose relative position to interactants is known to them, that is, those existing entities that participants can point to with their index fingers.

Simple Things potential referents, whether material or not, that is, whether or not they occupy space, have mass, and can be directly observed. Simple things include any mentionable portion of matter, whether they constitute a body or not, whether they constitute an entity or not, as well as any mentionable referent that can be observed only indirectly such as *the distance between two entities* or *the social latter* discussed in Chapter 3.

Socially Construed Universe there is no such thing as a socially construed universe because a universe is, by definition, what exists whether or not humans observed it. Any socially construed universe represented in language is, by definition, a product of imagination, not something observed thus far.

Socially Construed World all simple things that have been observed by interactants directly or indirectly and/or reported to them by others, that is, those simple things assumed to exist by interactants, whether materially or not. For instance, between two entities there is a distance: the distance is not material, but it exists between two entities and can be mentioned by interactants.

Socially Construed Situation all simple things whose relative position to interactants is known to them, that is, those existing simple things that interactants can point to with their index fingers.

Presence simple things may belong to the socially construed situation or not. Present things have a location known to interactants whereas the location of absent things is unknown to at least one of them.

Fame simple things belong to the socially construed world in different degrees. Famous things are known to many people and obscure things are known to few, that is, simple things have a degree of fame in the socially construed world. Therefore, absent things mentioned by name have different degrees of chance of being identified by a random addressee corresponding to their degrees of fame.

Nomenclatory power while simple things belong to socially construed world in different degrees, their names belong to a language also in different degrees. For instance, the actor Brad Pitt's official name is *William Bradly Pitt*. His show business name *Brad Pitt* is widely known and has a high nomenclatory power, but his official name *William Bradly Pitt* is not so known, thus having a low nomenclatory power.

Context of Culture The context of culture of an utterance is a named area in the socially construed world. It can be a continent such as Europe. It can be a country such as Germany. It can be a city such as Bremen. Or it can be the physical offices of an organisation such as a corporation, a university, a school, an assisted-living facility, or some other named place.

Context of Situation The context of situation of an utterance is the socially construed situation.

In this study, all potential referents will be in the socially construed situation. As a result, utterances about absent things will not be understood. Moreover, only directly observable phenomena shall be taken into account. This means utterances about *the distance between two entities* will not be understood by the wheelchair. In addition, all potential referents the wheelchair identifies will be entities, a subset of all potential referents excluding any unbounded portions of matter such as *air* and *water*. In other words, the context of situation for user utterances shall be a socially construed ontic situation. The entities in the socially construed situation will be the only potential referents. Finally, the socially construed situation will be bounded by the walls of an apartment. Everything outside the apartment will not be considered present. The context of culture is a fictional assisted-living facility with other apartments, somewhere unspecified in Bremen, in Germany. Experiment participants were not told of any particular entity in this fictional facility, so they could not refer to anything in this context of culture when talking to the wheelchair.

With this limitation of scope, I shall move on to describing what an ontology is and what is the object of a linguistic ontology.

5.2 SROIQ(D) Description Logic

Symbols are not only composed of other symbols, they also imply other symbols. For instance, all entities represented by the atomic symbols *prey* or *predator* as well as by the composite symbols *prey animal* or *animal of prey* can also be represented by the symbol *animal*. This means that the classes of preys and predators keep logical relations with the class of animals, namely the former are subclasses of the latter. As for the symbols themselves, the classifiers *prey* and *predator* are hyponyms of the classifier *animal* whereas the subclassifier *of prey* in *animal of prey* is a synonym of the classifier *predator*.

Description Logic can be applied for creating models of described phenomena aimed at supporting both composition and inference as described above. Description logic is decidable and can be computed in a finite amount of time. This makes it more adequate for computing than other more powerful logics such as First Order Logic and Common Logic. In its basic form, Description Logic is codenamed ALC (Schmidt-Schauß and Smolka, 1991) and has the following logical structures.

1. An **inventory** is something that contains zero or more **items**.
2. A **class expression** is an either atomic or composite symbol that stands for a different set of items in each inventory. Items in this set are **instances** of the class expression.
3. A **class** is an atomic class expression that stands for a different set of items in each inventory.
4. The **top** class stands for the full set of items in each inventory.
5. The **bottom** class stands for an empty set of items in every inventory.
6. The **union of two class expressions** is a composite symbol that stands for a different set of items in each inventory, resulting from the union of the instances of each class expression in that inventory.

7. The **intersection of two class expressions** is a composite symbol that stands for a different set of items in each inventory, resulting from the intersection of the instances of each class expression in that inventory.

8. A **binary relation** stands for a different set of ordered pairs of items in each inventory. Ordered pairs in this set are instances of the binary relation. The first item of a given ordered pair is the **domain** and the second item is the **range**. Saying that an ordered pair is an instance of a binary relation is the same as saying that the domain is **related to** the range in a given way.

9. Each **role** is associated with one and only one binary relation. A role is an atomic symbol and, for each inventory, it stands for the domains of the instances of the associated relation in that inventory. Since a role stands for sets of items, not sets of ordered pairs, a role is also a class expression.

10. A **role relative to an instance of a given class expression** stands for a subset of the domains of the associated relation where the domain is related to at least one instance of the given class expression.

11. A **role relative to exclusively instances of a given class expression** stands for a subset of the domains of the associated relation where the domain is related to exclusively instances of the given class expression.

12. A class expression **subsumes** another if it always stands for a set of items that contains the set of items represented by the subsumed class expression for any inventory.

The logic used in this research is description logic incremented with the following logical properties and structures codenamed S, R, O, I, Q, and D.

S role transitivity - if item 1 being related to item 2 and item 2 being related to item 3 implies that item 1 is also related to item 3, then the relation is transitive. The role associated with this relation is also transitive.

R generalised role hierarchy - a relation subsumes another if it stands for a set of ordered pairs of items and the other stands for a subset of those ordered pairs for any inventory. If item 1 is related to item 2 in one way and item 2 is related to item 3 in potentially another way, item 1 is indirectly related to item 3, so a chain of direct relations is an indirect relation. A direct relation subsumes an indirect relation if it stands for a set of ordered pairs of items and the indirect one stands for a subset of those ordered pairs for any inventory. A **role hierarchy** is created when the role associated with a subsuming relation is said to subsume the roles associated with subsumed relations. A generalised role hierarchy is created when the role associated with a subsuming relation is said to subsume role chains associated with subsumed relation chains, the simplest case being a role chain comprising a single role.

O nominal - a nominal is a unary class, a class that stands for a single item in any inventory.

I inverse role relation - for any given inventory, if a relation stands for a set of ordered pairs of items, its inverse relation stands for a set of ordered pairs of the same size respecting the rule that for each pair of items in the first, there is an ordered pair in the second in the inverse order. A role is the inverse of another if its relation is the inverse of the other's relation.

Q qualified cardinality restrictions for roles - a role has a qualified cardinality restriction if every single item in the domain set is related to at least, exactly, or at most x items for any inventory, where x is a cardinal number and *at least*, *exactly*, and *at most* are the cardinal number qualifiers.

D data types - open sets of literals such as *strings*, *integers*, and so on.

Given this very brief description of description logic, let me now show how this applies to the creation of linguistic and domains ontologies.

5.3 Linguistic vs domain ontologies

To motivate the distinction between linguistic and domain ontologies, I shall start with the difference between entity names and terms for entity classes. Let us consider the following two examples:

(77) Rolland is a wheelchair.

(78) This wheelchair is called Rolland.

In Example 77 the entity class *wheelchair* is ascribed to the entity called *Rolland*, whereas in Example 78 the entity name *Rolland* is ascribed to the shown instance of *wheelchair*. In a domain ontology, the word *Wheelchair* is associated directly with a class of items and in a domain inventory, the word *Rolland* is associated directly with an item (see Figure 5.1). This is not the case for a linguistic ontology.

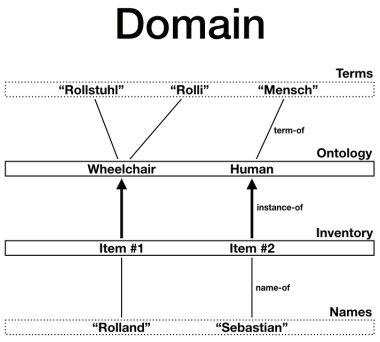

Figure 5.1: A small domain ontology with associated terms and names

In a linguistic ontology, it is the experiential contribution that a linguistic symbol has that is captured, not only the nature of their associations with entities (see Figure 5.2).

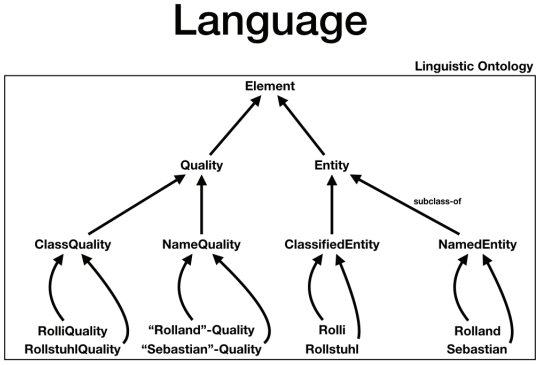

Figure 5.2: A small linguistic ontology for entities and their qualities

For instance, in Example 77, the word *wheelchair* represents a **class quality** that is ascribed to the entity called Rolland. In Example 78, the wording *this wheelchair* represents a **classified entity**. For that reason, these are two different symbols as far as representation is concerned: the first represents a quality and the second an entity. The same applies to the word *Rolland*. In Example 77, it represents a **named entity** whereas in Example 78, it represents a **name quality**. Qualities of entities are not entities themselves. In a linguistic ontology, both the nature of word-entity association and the type of phenomena are differentiated, with precedence to the type of phenomena. Therefore, both a name quality and a class quality are qualities and both a named entity and a classified entity are entities.

Wordings representing entities are different from wordings representing qualities. Wordings representing entities are **nominal groups** and they vary in number, deixis, and case. Tables 5.1-5.3 show different nominal groups for wheelchair.

	singular	plural
nominative	*der Rollstuhl*	*die Rollstühle*
accusative	*den Rollstuhl*	*die Rollstühle*
dative	*dem Rollstuhl*	*den Rollstühlen*
genitive	*des Rollstuhls*	*der Rollstühle*

Example 5.1: Nominal groups for *the wheelchair*

	singular	plural
nominative	*'der Rollstuhl*	*'die Rollstühle*
accusative	*'den Rollstuhl*	*'die Rollstühle*
dative	*'dem Rollstuhl*	*'den Rollstühlen*
genitive	*'des Rollstuhls*	*'der Rollstühle*

Example 5.2: Nominal groups for *that wheelchair*

	singular	plural
nominative	*dieser Rollstuhl*	*diese Rollstühle*
accusative	*diesen Rollstuhl*	*diese Rollstühle*
dative	*diesem Rollstuhl*	*diesen Rollstühlen*
genitive	*dieses Rollstuhls*	*dieser Rollstühle*

Example 5.3: Nominal groups for *this wheelchair*

In contrast, wordings representing qualities are **quality groups** and they vary only in number for class qualities (see Table 5.4).

singular	plural
Rollstuhl	*Rollstüle*

Example 5.4: Quality groups for *wheelchair*

Class and name qualities are not only ascribed to represented entities, they may also be part of the representation of those entities as further restrictions. Let us consider the following dialogue:

(79) A: Play Help!
B: Do you mean the song Help! or the album Help!?
A: The song.
B: Ok, playing the song Help!

In the dialogue of Example 79, there are two playable items called Help! in the domain inventory, one being a song and the other being an album. Both wordings *the song Help!* and *the album Help!* represent named entities that are also classified. Through semantic composition, the named entity represented by *Help!* in the first utterance is represented again with the wording *Help!* modified by the classifiers *the song* or *the album* for the purpose of disambiguation, each of which representing a different class quality. The wording *the song* in the third utterance is elliptical, since it still represents the named entity modified by the classifier *the song*.

The same happens to an entity name. Let us consider the following online article titles:

(80) Obama was here.

(81) A man called Obama was here.

In Example 80 the word *Obama* represents a named entity whereas in Example 81 the wording *a man called Obama* represents a classified entity that carries a name. Through semantic composition *a man* represents a classified entity whereas *called Obama* represents a name quality of the classified entity.

To represent semantic structures of this kind, we can make use of SROIQ(D) Description Logic in the following way. Let **ClassifiedEntity** (CE), **ClassQuality** (CQ), **ClassAscription** (CA), **NamedEntity** (NE), **NameQuality** (NQ), and **NameAscription** (NA) be classes in a linguistic ontology and let **carrier**, **attribute**, **name**, and **class** be roles. These are classes and roles of represented phenomena, which are not necessarily entities. Let us now consider the experiential contribution of each part of a wording.

the	cup
	CE
Modifier	Head

Let A an instance of ClassifiedEntity represented by *the cup*. The logical structure can be represented in the following way:

```
#A ClassifiedEntity
```

For the named entity analysed below:

Help!
NE
Head

Let B be a NamedEntity represented by *Help!*. The logical structure is the following:

```
#B NamedEntity
```

Now let us look at an example with class quality modification.

the	song	Help!
	CQ	NE
Modifier	Modifier	Head

For this example, let C be a NamedEntity represented by *Help!* and D be a class quality represented by *song*. Let class be the role of C relative to D, reading *which is-of-class song*. The resulting logical structure is:

```
#C NamedEntity {
  class #D ClassQuality
}
```

Finally, let us consider a case of name quality modification.

the	man	called	Obama
	CE		NQ
Modifier	Head	Modifier	

Let E be a ClassifiedEntity represented by *a man* and F be a NameQuality represented by *called Obama*. Let name be the role of E relative to F, reading *which has-name Obama*:

```
#E ClassifiedEntity {
  name #F NameQuality
}
```

Finally, when it comes to subclassification as in *prey animals*, the logical structure takes the following shape. Let **subclass** be a role and **SubclassQuality** be a class. Let G be the ClassifiedEntity represented by *animal* and H be the SubclassQuality represented by *prey*. Let subclass be the role of G relative to H, reading *which is-of-subclass prey animal*:

```
#G ClassifiedEntity {
  subclass #H SubclassQuality
}
```

For clauses, we have the following:

this song	is called	Help!
CE		NQ
Carrier	Process	Attribute

Rolland	is	a wheelchair
NE		CQ
Carrier	Process	Attribute

Let A be the ClassifiedEntity represented by *this song*, B be the NameQuality represented by *Help!*, and X be the NameAscription represented by *this song is called Help!*. Moreover, let carrier be the role of X relative to A, and attribute be the role of X relative to B. The following structure represents this composite phenomenon.

```
#X NameAscription {
  carrier #A ClassifiedEntity
  attribute #B NameQuality
}
```

For the wording *Rolland is a wheelchair*, we come to the following logical structure:

```
#X ClassAscription {
  carrier #A NamedEntity
  attribute #B ClassQuality
}
```

As illustrated above, a linguistic ontology is an ontology of represented phenomena, not a domain ontology. Its inventory contains represented phenomena. A single entity in the situation may be represented multiple times and each time it is represented, there is a new represented phenomenon in the linguistic inventory. In addition, not only entities can be represented, but also the qualities of these entities themselves. When it comes to composite symbols, a linguistic ontology contains classes for **figures** such as NameAscription and ClassAscription. In the next section, I give an overview of the types of phenomena required for the linguistic data from the Wizard-of-Oz experiment reported in Chapter 4.

5.4 Upper Model

Logical structures such as the ones in the previous section must be further specified if we want them to be isomorphic to the semantic structure of utterances. For instance, let us consider Examples 82-89.

(82) This wheelchair is called Rolland.

(83) This wheelchair is called Martha.

(84) This assistant is called Siri.

(85) This assistant is called Alexa.

(86) Rolland is a wheelchair.

(87) Martha is a wheelchair.

(88) Siri is an assistant.

(89) Alexa is an assistant.

To differentiate the logical structure for these four name ascriptions in Examples 82-85, let Wheelchair and Assistant be subclasses of ClassifiedEntity and let RollandQuality, MarthaQuality, SiriQuality, and AlexaQuality be subclasses of NameQuality. A logical structure for Example 82 has the following shape:

```
#X NameAscription {
  carrier #A Wheelchair
  attribute #B RollandQuality
}
```

Examples 83-85 follow the same principle. Their logical structure has the same shape. Only the subclass of NameQuality gets replaced.

Now let us look at Examples 86-89. Let Rolland, Martha, Siri, and Alexa be subclasses of NamedEntity and let WheelchairQuality and AssistantQuality be subclasses of ClassQuality. A logical structure for Examples 86 is shown below:

```
#X ClassAscription {
  carrier #A Rolland
  attribute #B WheelchairQuality
}
```

Classes such as Rolland, Martha, Siri, and Alexa are subclasses of the same class, namely NamedEntity. They form a set of experiential classes that can be selected without changing the constituency of a logical structure. Specifying that #A in the logical structure above is not only of class NamedEntity, but also of class Rolland adds the necessary information for this structure to be isomorphic to *Rolland is a wheelchair*, but not to *Martha is a wheelchair*. In other words, this further specification of the logical structure corresponds to a specification of which entity name to choose. As a consequence, there is a bidirectional map between a set of entity names and two sets of classes in the linguistic ontology: namely, the subclasses of NamedEntity and the subclasses of NameQuality. I shall call these subclasses that can be mapped to entity names **nominals**. Classes of entities in the situation can also be bidirectionally mapped to subclasses of ClassifiedEntity and ClassQuality. I shall call these subclasses that can be mapped to entity classes **taxa**.

The most delicate classes of the linguistic ontology realised by lexical choice is a taxonomy and the upper classes realised by grammatical choices are known as an Upper Model. The Upper Model that inspired the one used in this research

is called GUM-3 (Farrar et al., 2005; Bateman et al., 2009). Some classes of the Upper Model were required for this research, but many classes had to be added. In the following, I will only explain the classes that were used in the Upper Model developed for this study.

```
- Phenomenon
  - Sequence
  - Figure
  - Element
    - SimpleThing (ClassifiedEntity, NamedEntity)
    - SimpleQuality (ClassQuality, NameQuality)
    - Circumstance (AbsoluteCircumstance, RelativeCircumstance)
    - Relation
    - Process
```

Any ontology of phenomena starts with a top class **phenomenon** because all items in the corresponding inventories are phenomena. There are three types of phenomena: **elements** are represented by groups and phrases, **figures** by simple clauses, and **sequences** by clause complexes. Elements are divided into **simple things**, **simple qualities**, **circumstances**, **relations**, and **processes**.

A simple thing, as defined in previous sections, is a potential referent, whether material or not, whether detached or not, whether bounded or not. In this study, only bounded things were modelled, that is, only named entities such as *Rolland* and classified entities such as *wheelchair* and *room*. Three quality types were modelled: namely, name qualities such as *Rolland* in *I am Rolland*, class qualities such as *wheelchair* in *I am an intelligent wheelchair* and *room* in *this is a living room*, and subclass qualities such as *living* in *the living room*. Moreover, two circumstance types were modelled: namely, absolute circumstances such as *here*, *there*, *in*, and *home*, as well as relative circumstances such as *in the kitchen* and *in front of the wash basin*. In relative circumstances, spatial relations include *in* as in *in the kitchen* and *in front* as in *in front of the wash basin*. Finally, processes of different kinds were modelled. Examples of and details about processes will be presented closer to the end of this chapter.

Without any further ado, I shall proceed to the listing and definition of the classes in the lower model and the listing of their instances in the situation.

5.5 Simple things

Since the boundary of the socially construed situation was the apartment and since only entities were modelled, all simple things classified for the wheelchair are entities could be found in the apartment at the time.

Let atomic experiential symbols for a given entity class be linguistic symbols with no part that is also a symbol for a more general entity class. Atomic symbols for German include *Tisch* (*table*) and *Waschbecken* (*wash basin*), but does not include *Arbeitstisch* (*desk*, "*working table*") nor *Esstisch* (*dining table*). In English, *desk* is an atomic symbol and *dining table* is not. In German, *Waschbecken* is an atomic symbol even though it is composed of two derivational morphemes, the reason for that being that no one referred to such entities as a *Becken* (pelvis, pool).

For each class of entities that corresponds to an atomic symbol in language, there is a subclass of ClassifiedEntity and ClassQuality in the linguistic ontology. For each subclass of those classes that is associated with a modifying symbol such *Arbeits-* (work) and *Ess-* (dining table), there is a subclass of SubclassQuality in the linguistic ontology. Those are the *domain classes* of the linguistic ontology. Taxa and nominals are the atoms of semantic composition. Taxa correspond to a subset of all potential entity classes in a domain ontology, the subset that is associated with symbols in a language such as *Tisch* and *Ess-* in *Esstish*.

In the following, I list all entity classes that were modelled for this research.

1. **Apartment** there is a single apartment
2. **Room** there are three rooms
3. **Study** there is a single room that is a study
4. **Bedroom** there is a single room that is a bedroom
5. **Bathroom** there is a single room that is a bathroom
6. **LivingRoom** there is a single room that is a living room
7. **Kitchen** there is a single room that is a kitchen
8. **Sofa** there is a single sofa in the apartment
9. **Bed** there is a single bed in the apartment
10. **Table** there are two tables in the apartment
11. **Desk** there is a single table that is a desk
12. **DiningTable** there is a single table that is a dining table
13. **Washbasin** there is a single washbasin
14. **Fridge** there is a single fridge
15. **ChargingStation** there is a single charging station
16. **Book** there are many books
17. **Door** there are many doors
18. **Person** there are three people
19. **Wheelchair** there is one person who is a wheelchair
20. **SeatArea** there is one seat area
21. **Human** there are two people who are humans
22. **Hand** there are four hands
23. **Leg** there are four legs
24. **Mouth** there are two mouths

5.6 Circumstances

In this study, only a few circumstances of entities were represented in spoken language. They were mostly relative locations of entities, that is, locations of an entity relative to another. Relative locations are typically composed of two symbols: one standing for the spatial relation from the entity being located to another entity whose location is taken as the origin for locating entities and the other symbol standing for the origin entity. There are three main relations:

1. **Containment** people can be contained by the apartment and its rooms and humans can be contained by wheelchairs
2. **Support** humans can be supported by the bed, the wheelchair, and the sofa, the wheelchair can be supported by the charging station, and books can be supported by tables
3. **Proximity** people can be close to the doors, sofas, beds, tables, washbasins and fridges
4. **FrontalProjecting** people's location can projected frontally from washbasins, sofas, and fridges if people are close to those objects

However, spatial relations were also represented in general terms without specifying which relation was being considered. This is the case for Examples 91 and 93.

(90) Take me into the kitchen.

(91) Take me to the kitchen.

(92) Take me next to the sofa.

(93) Take me to the sofa.

In such cases, spatial classes for entities were used. A kitchen is a potential container for people and a sofa is a potential obstacle for people and a potential support for humans in particular. For giving meaning to underspecified expressions such as *to the kitchen* and *to the sofa*, three subclasses of Entity were added: namely, **PotentialContainer**, **PotentialObstacle**, and **PotentialSupport**. A Kitchen is a PotentialContainer because *the kitchen* can be part of the wordings *in the kitchen* and *into the kitchen* and because *to the kitchen* can mean the same as *into the kitchen*. A Sofa is a PotentialObstacle because *the sofa* can be part of the wording *next to the sofa* and because *to the sofa* can carry the same meaning as *next to the sofa*. Sofa is also a PotentialSupport because *the sofa* can be part of the wording *on the sofa* for this domain.

One relevant potential location for the book, the wheelchair and the two humans was pre-calculated for each wording. These locations were:

For the wheelchair user

1. **on the bed**
2. **on the sofa**

3. **in the wheelchair**
4. **next to the bed** a position in the wheelchair where the wheelchair user can move onto the bed
5. **in front of the sofa** a position in the wheelchair where the wheelchair user can move onto the sofa
6. **next to the fridge** a position in the wheelchair where the wheelchair user can open the fridge door and reach the fridge contents
7. **in front of the washbasin** a position in the wheelchair where the wheelchair user can use the washbasin
8. **next to the desk** a position in the wheelchair where the wheelchair user can use the desk
9. **next to the dining table** a position in the wheelchair where the wheelchair user can use the dining table
10. **next to the apartment door** a position in the wheelchair where the wheelchair user can open the apartment door
11. **next to the book** a position in the wheelchair where the wheelchair user can pick up the book where ever it is currently in the apartment

For the wheelchair

1. **on the charging station**
2. **next to the bed** a position for which the wheelchair user can move from the bed into the wheelchair
3. **in front of the sofa** a position for which the wheelchair user can move from the sofa into the wheelchair
4. **next to the fridge** a position for which the wheelchair user can open the fridge door and reach the fridge contents
5. **in front of the washbasin** a position for which the wheelchair user can use the washbasin
6. **next to the desk** a position from which the wheelchair user can use the desk
7. **next to the dining table** a position from which the wheelchair user can use the dining table
8. **next to the apartment door** a position from which the wheelchair user can open the apartment door
9. **next to the book** a position for which the wheelchair user can pick up the book where ever it is currently in the apartment
10. **next to the user** a position for which the wheelchair user can move into the wheelchair from where ever he or she is currently in the apartment

For the book

1. **on the shelves**
2. **on the desk**
3. **on someone's hands**

A human and a wheelchair are moving bodies and both a human and a book are moveable bodies. All other entities in this situation cannot move on their own and cannot be carried around. For this reason, the wheelchair needs to keep track of the positions of only humans, wheelchairs, and books for grounding relative locations such as *to the book* in *take me to the book*.

While tracking body positions, the wheelchair needs to expect that books on someone's hands have a dependent position in relation to the person and that a person in a wheelchair has a dependent position in relation to the wheelchair. Any position change of the dominant body corresponds to a position change of the dependent body.

In addition, experiment participants wanted to arrive at different destinations by making commands such as *take me into the bedroom* and *take me into the study*. This means that the actual positions represented by *in the bedroom* and *in the study* were different although wordings such as *the bedroom* and *the study* represented the same room. The intended destination for *in the bedroom* was *next to the bed* and the intended destination for *in the study* was *next to the desk*.

For this reason, in the domain ontology, a bedroom was defined as a room with a bed and a study is defined as a room with a desk. This means that the definition of the room class could be used to determine the entity the user wanted to be next to.

These potential positions relevant for performing daily tasks are the positions that get represented as destinations for movement when humans want to accomplish their daily tasks and also the positions that need to be selected when wheelchair users want to go to a particular room in the apartment. In the following, I will present an overview of the figures construed by wordings in German.

5.7 Figures

In the previous sections, we classified wordings as representations of simple things and spatial locations. Now we move on to describe multiple ways to arrange such elements in figures. The roles described in this section are different from the ones in GUM-3 and they are used instead of the roles offered by the current version of the upper model. Some figures are also new or defined in a different way, so they are also used instead of those offered by the upper model. They are not further specifications of the current version of the upper model.

In descriptions of states of the world, simple things can be represented as a **carrier** of an **attribute**. For instance, in the examples below, a man called Matthias is described as having a location in the past and as having a different location now.

Matthias Matthias	war was	im Schlafzimmer in the bedroom
Carrier		Attribute

Matthias Matthias	ist is	in der Küche in the kitchen
Carrier		Attribute

Similar to the relation between carriers and attributes for name and class ascriptions, this is a relation of **LocationAscription**. Their logical structure can be seen below:

```
#X LocationAscription {
  carrier #A Matthias
  attribute #B RelativeLocation {
    minor-process #B1 Containment
    minor-range #B2 Kitchen
  }
}
```

In descriptions of location changes, simple things can be represented as ***matter*** undergoing ***change***. For instance, in the examples below, Matthias is a portion of matter undergoing a change in location.

Matthias Matthias	ging went	vom Schlafzimmer in die Küche from the Kitchen to the bedroom
Matter		Change

Matthias Matthias	geht will go	von der Küche ins Schlafzimmer from the Kitchen to the bedroom
Matter		Change

These changes in location have an initial location and a final location as shown in examples below:

vom Schlafzimmer from the Kitchen	in die Küche to the bedroom
Initial Attribute	Final Attribute

von der Küche from the Kitchen	ins Schlafzimmer to the bedroom
Initial Attribute	Final Attribute

Generalising over static and dynamic configurations, a simple thing carrying an attribute and a simple thing undergoing change both take the role of ***medium***. This means that all elements that take the roles of either ***carrier*** or ***matter*** also take the role of ***medium***. In other words, the role of ***medium*** subsumes the roles of ***carrier*** and ***matter***.

In the linguistic ontology, ***carrier***, ***matter*** and ***medium*** as well as ***attribute*** and ***change*** are possible roles in a **Figure**. ***Initial attribute*** and ***final attribute*** are roles of a **Change**.

Configurations of roles that ascribe an attribute to a carrier are ***Ascribing Figures*** and those that update an attribute of matter are ***Material Figures***.

There are three types of attributes. Intrinsic attributes are those that are intrinsic to the carrier, that is, that are not another simple thing nor a circumstance such as a spatial location. Name qualities, class qualities and modal qualities such as size and weight are intrinsic attributes.

der that	heißt is	Matthias Matthias
Carrier		Attribute

Example 5.5: intrinsic attribute - name quality

Matthias Matthias	ist is	Krankenschwester a nurse
Carrier		Attribute

Example 5.6: intrinsic attribute - class quality

Matthias Matthias	ist is	groß tall
Carrier		Attribute

Example 5.7: intrinsic attribute - modal quality

Extrinsic attributes come in two forms: they are either another simple thing possessed by or possessing the carrier or they are circumstances in which the carrier finds him or herself. In the case of possessive attributive relational figures, they are either ***OfOwningType*** or ***OfBelongingType***. In the case of circumstantial attributive relational figures, the attributes are circumstantial. Examples below:

die Wohnung the apartment	gehört belongs to	Matthias Matthias
Carrier		Attribute

Example 5.8: possessive attribute - entity - figure of belonging type

Matthias Matthias	besitzt owns	die Wohnung the apartment
Carrier		Attribute

Example 5.9: possessive attribute - entity - figure of owning type

Most circumstantial ascriptions take a full circumstance as attribute as in Example 5.11, but some do not as in Example 5.12. These configurations take a simple thing as **range**. The role of range is equivalent to the role of a minor

Matthias Matthias	ist is	im Wohnzimmer in the living room
Carrier		Attribute

Example 5.10: circumstantial attribute - location

wo where	ist is	das Buch? the book?
Attribute		Carrier

Example 5.11: circumstantial attribute - location

Matthias Matthias	hat has	es it
MinorRange		Carrier

Example 5.12: circumstantial minor range - entity

es it	ist is	auf on	dem Tisch the table
Carrier		Attribute	
			Minor Range

Example 5.13: circumstantial attribute - location

circumstantial minor range - entity

range of a relative circumstance serving as a circumstantial attribute in Example 5.13.

Possessive and ***Circumstantial*** attributive relational figures have respectively a ***SimpleThing*** and a ***Circumstance*** as attributes. ***Circumstantial*** attributive relational figures ***OfHavingType*** have a ***SimpleThing*** as minor range.

5.7.1 Action

Actions are changes in matter performed by a participant. The matter being affected can be the actor him or herself as in Example 5.14.

Matthias Matthias	duscht sich is showering
Actor	
Matter	

Example 5.14: actor-affecting action

The affected matter can also be a goal aimed at by the actor as in Examples 5.15 and 5.16.

However, in German, not all represented classified entities are unique in the situation. For instance, the nominal groups *die Hände* (*the hands*) from

Matthias Matthias	wäscht is washing	seine Hände his hands
Actor		Goal
Agent		Matter

Example 5.15: goal-affecting action

Matthias Matthias	wäscht is washing	die Hände des Babys the baby's hands
Actor		Goal
Agent		Matter

Example 5.16: goal-affecting action

Examples 5.17 and 5.18 does not represent all the hands in the situation. They represent hands that are part of another participant in the process: a part of the actor in Example 5.17 and a part of the goal in Example 5.18.

Matthias Matthias	wäscht sich is washing himself	die Hände the hands
Actor		Actor Part
Agent		Matter

Example 5.17: actor-part-affecting action

Matthias Matthias	wäscht is washing	dem Baby the baby	die Hände the hands
Actor		Goal	Goal Part
Agent			Matter

Example 5.18: goal-part-affecting action

A simple thing taking the role of ***actor***, ***actor-part***, ***goal***, or ***goal-part*** can also take the role of ***matter***, but they not always do. For inferring which simple thing takes the role of matter, the role of ***matter-in*** is specified in the ontology. It is equivalent to an ***actor-in*** an ***ActorAffectingAction***, an ***actor-part-in*** an ***ActorPartAffectingAction***, a ***goal-in*** a ***GoalAffectingAction*** and a ***goal-part-in*** a ***GoalPartAffectingAction***, where ***matter-in*** is the inverse of ***matter***, ***actor-in*** the inverse of ***actor*** and so on.

5.7.2 Service

Services are an understanding of change not as a material process carried out by a single agent, but as a process carried out by two agents, one taking the role of service provider and the other of service client. By separating actions from services, we are able to distinguish the material processes that must be carried out by the wheelchair user from those that must be carried out by the wheelchair for the wheelchair user under his or her command.

There are six types of services regarding which participant or participant part is affected. In the same way as actions, services may affect goals as in Example 5.19 or goal-parts as in Example 5.20.

Matthias Matthias	wäscht is washing	die Hände des Babys the baby's hands	für mich for me
Provider		Goal	Client
Agent		Matter	

Example 5.19: goal-affecting service

Matthias Matthias	wäscht is washing	dem Baby the baby	die Hände the hands	für mich for me
Provider		Goal	Goal Part	Client
Agent			Matter	

Example 5.20: goal-part-affecting service

However, differently from actions, the affected matter can also be a service client as in Example 5.21 or a service client's part as in Example 5.23.

Matthias Matthias	duscht is showering	den Senior the old man
Provider		Client
Agent		Matter

Example 5.21: client-affecting service

Matthias Matthias	wäscht is washing	die Hände des Seniores the old man's hands
Provider		Goal
Agent		Matter

Example 5.22: goal-affecting service

Matthias Matthias	wäscht is washing	dem Senior the old man	die Hände the hands
Provider		Client	Client Part
Agent			Matter

Example 5.23: client-part-affecting service

Finally, the provider can also be represented as the affected matter as in Examples 5.24 and 5.25.

The roles ***goal***, ***goal-part***, ***client***, ***client-part***, ***provider***, and ***provider-part*** are the invert of, respectively, ***goal-in***, ***goal-part-in***, ***client-in***, ***client-part-in***, ***provider-in***, and ***provider-part-in***, which are used for inferring the

Rolland Rolland	kommt is coming	ans Sofa ran to the sofa	für mich for me
Provider			Client
Matter			

Example 5.24: provider-affecting service

Rolland Rolland	dreht turns	die Sitzflache the seat	zum Sofa to the sofa	für mich for me
Provider		Provider Part		Client
Agent		Matter		

Example 5.25: provider-part-affecting service

simple thing serving as matter. In the same way as for actions, the role matter-in has a definition for each type of service. It is specified as equivalent to ***goal-in*** a ***GoalAffectingService***, ***goal-part-in*** a ***GoalPartAffectingService***, and so on. With these definitions, a reasoner can determine which element plays the role of ***matter*** in each figure type.

5.7.3 Action per locution

Actions per locution are a third understanding of agency. In this case, the material process is carried out by two agents, a service provider and a service client. However, differently from services, the one who wants and gets the desired change (the client) is understood as the one doing the action. His or her action is, however, not a simple action because it is not complete when the client finishes uttering a command. It is only complete when the provider finishes carrying out the demanded change. For this reason, though actions per locution are actions, the participant giving the command is not an actor in the sense of the agent who wants a change and performs it. The command giver is a service client.

As explained in the introduction, actions per locution are similar to locutions about services. A service can be represented in locution as in Example 5.26 and it can be performed on locution as in Example 5.27.

ich I	bitte am telling	Roland Roland	darum –	ans Sofa to the sofa	zu kommen to come
Requester		Provider			
Agent		Matter			

Example 5.26: provider-affecting service in command

The "in command" extension *I told* and the "on command" extension *I made* are extensions to a service figure. All service configurations from Section 5.7.2 can be extended in this way.

Actions per locution are similar to "on command"-extended services in that they represent material change as a locution about an executed service. However, they are different from them in the fact that they are not an extended

ich I	bringe am making	Roland Roland	dazu –	ans Sofa to the sofa	zu kommen come
Client		Provider			
Agent		Matter			

Example 5.27: provider-affecting service on command

figure, but a new figure in their own. Examples 5.28 and 5.29 are actions realised by a service client: that is, the client finishes his or her contribution to the material process when he or she finishes making a command.

ich I	bringe am bringing	Roland Roland	ans Sofa to the sofa
Client		Provider	
Agent		Matter	

Example 5.28: provider-affecting action per locution

ich I	gehe am going	zum Sofa to the sofa
Client		
Matter		

Example 5.29: client-affecting action per locution

Actions per locution are primarily about the client and his/her action. This means the service provider might not be represented at all. They come in all variations that are available for services: namely, client-affecting, client-part-affecting, provider-affecting, provider-part-affecting, goal-affecting, and goal-part-affecting. In the same way as for simple actions, "in command" and "on command" extensions do not apply to actions per locution. Finally, the role ***matter-in*** also has a definition for each type of action per locution. It is specified as equivalent to ***goal-in*** a ***GoalAffectingActionPerLocution***, ***goal-part-in*** a ***GoalPartAffectingActionPerLocution***, and so on. Because of this, the participant playing the role of matter is inferred with definition rules in the same way as for simple actions and services.

5.8 Processes

Each figure type from the previous section is further specified regarding whether it also requires an attribute or a change complement. For instance, the service of *coming* (*herkommen*) does not require anything as complement besides the service provider whereas the services of *coming somewhere* (*wohin kommen*) and *going somewhere* (*wohin fahren*) require a change as complement.

For each subfigure, a list of processes was created. For instance, the wheelchair offered the services *coming somewhere* (*wohin kommen*) and *going somewhere* (*wohin fahren*), but not the service of *walking somewhere* (*wohin gehen*). No one in the situation offered the service of walking somewhere, so this process

was not included in the list of potential provider-affecting services with a route. A separate list of processes was create for each specific figure type.

5.8.1 Processual things

Some actions such as the action of *washing something* were not always represented by a verb group such as the bold wording in Example 94. Sometimes they were represented by a combination of an execution verb such as *doing* (*machen*) and an action such as the action of *washing one's own mouth* (*sich den Mund ausspülen*), that is, a *mouth wash* (*Mundspülung*).

(94) Ich möchte **mir** den Mund **ausspülen**.
I want to **wash myself** the mouth.

(95) ich möchte eine Mund**spülung machen**.
I want to **do** a mouth **wash**.

In Example 95, the action of washing one's own mouth is represented as an instance of a class of actions, that is, as a referent. Since simple things include all potential referents, this action is a simple thing. However, it is a simple thing of a particular kind. It is a processual thing in the sense that it is executed, that is, it is the material process that is executed, not some portion of matter affected by and/or controlling the process.

In the linguistic ontology, processual things were further specified in two ways. A subtype of processual things was created for each execution process: *Machbar* (*doable*) for processual things that can be *done* and *Bekommbar* (*gettable*) for those that can be the object of *getting*. In parallel, a subtype of processual things was created for each type of figure they fit in. Lists of processual things were created for each combination of execution process and figure types. For *Mundspülung* (*mouth wash*), *Mund* (*mouth*) represents a subclass quality, specifying the type of actor part affected by such processual things.

5.9 Logical relations

Processes relate to each other in different ways. When processes are represented as a figure, that is, when they are represented by a simple clause, a logical relation between two processes is a connection (**nexus**) between two figures and is realised by clause connectors. In this study, we observed two types of connectors: result binder such as *damit* (*so that*) in *damit ich mir das Buch holen kann* (*so that I can pick up the book*) and purpose binders such as *um* (*to*) in *um mir das Buch zu holen* (*to pick up the book*).

Interrelated figures can be represented in three ways. They can be represented by **free finite** clauses, that is, clauses supposed to be understood directly in the current situation such as *I want to pick up the book* in the sequence *I want to go to the desk, I want to pick up the book*. They may be represented by **bounded finite** clauses in the sense that they will be finite clauses in a particular situation resulting from an action such as *so that I can pick up the book* in the sequence *I want to go to the desk so that I can pick up the book*. Once the speaker arrives at the desk, he or she will be able to state *I can pick up the*

book, a free finite clause. Finally, figures may also be represented by **bounded non-finite** clauses where either auxiliary verbs are removed or finite forms of the process verb are substituted by non-finite forms. For instance, the simple clause *to pick up the book* is non-finite in *I want to go to the desk to pick up the book*.

Moreover, clauses may be **saturated** or **unsaturated** in the following way. The free finite clause *I want to pick up the book* and the bounded finite clause *so that I can pick up the book* are saturated in the sense that all its participants are specified within it. The clause *to pick up the book* is unsaturated in the sense that the actor is a participant of the other figure to which this one is bounded. The clause itself leaves a slot unfilled, which is to be filled by an element of the connected figure.

All sequences of figures in the linguistic data were connections between an action, service, or action per locution and a motivation. The motivation was either a resulting state where the wheelchair user could do an intended action, the action the user wanted to perform, or the fact that the user wanted to perform this action.

Two free finite clauses represent two isolated figures. How a sequence between two isolated figures becomes inferrable is explained in the next section.

5.10 Exchanges

Free finite clauses are defined as clauses that must be understood in the current situation. In our linguistic data, the overwhelming majority of free finite clauses produced by users are demands for a service. These clauses can be divided in two broad groups. The first group consists of clauses representing services by the wheelchair or actions per locution by the user. When a service is represented as in *take me to the washbasin*, it is usually the last material process the wheelchair is required to perform for the user to perform the end action. When an action per locution is represented as in *I want to go to the washbasin*, the corresponding service is usually the last material process demanded from the wheelchair. In this case, the material process that the wheelchair needs to perform is represented, that is, it is **explicit**. The second group of clauses consists of clauses representing actions by the user as in *I want to wash my hands* as well as a few services by the wheelchair such as *go to the charging station*. In these cases, the last material process demanded from the wheelchair is implied by the represented material processes. In the first case, it is the service of taking the user to the wash basin; in the second, it is the service of recharging oneself. The service demanded by these clauses are **implicit**.

Let ***explicative*** and ***implicative*** be roles for free figures in semantic structures called ***Move*** (dialogue move) and let ***interjective*** be the roles of interjections such as *ok*, *thanks*, and *you're welcome*. In a move to demand a service, an explicative figure is the service under negotiation and an implicative figure is a figure that implies the service under negotiation. If explicative and implicative figures belong to the same move, the implicative figure implies the material process in the explicative figure and a sequence is inferred between the two.

In our linguistic data, the wheelchair never offered any service and the user always initiated the service exchange. For this reason, all moves demanding a service from the wheelchair are commands. Other moves follow such as

the wheelchair undertaking of the command, the wheelchair execution of the command, and sometimes a user thanking as in *thanks* (*danke schon*) and a wheelchair welcoming *you're welcome* (*bitte schon*). Each move in a sequence took a different role.

The move where the service being exchanged is specified takes the role of ***request*** and the move where the wheelchair accepts or rejects the request takes the role of ***response***. Any clarification moves - if they had occurred in our data - would have taken place in between the two. Preparation moves such as *may I ask you a favour?* or *what can I do for you?* - if they had occurred - would have happened before the request. Finally, execution moves as well as thanking and welcoming moves happen after the response move.

Let the class ***ExchangeOfService*** be a semantic structure with one or more constituents: preparations, at most one request, clarifications, at most one response, at most one execution, at most one thanking, and at most one welcoming. Let there be two types of moves: ***Utterances*** and ***Performances***. All constituents of utterances and performances are free figures. Taking this as the basic structure of an exchange of service, let's consider the following dialogue.

```
Human: take me to the washbasin!
Robot: ok, I'll take you there.
Robot: ((takes the user to the washbasin))
Human: thanks!
Robot: you're welcome!
```

This exchange of service has the following semantic structure.

```
#E ExchangeOfService {
  request:   #M1 Utterance {
    explicative: #F1 ClientAffectingService {
      provider: #P1 Addressee
      client: #P2 Utterer
      change: #P3 Route {...}
    }
  }
  response:  #M2 Utterance {
    interjective: #I1 Ok
    explicative: #F2 ClientAffectingService {
      provider: #P4 Utterer
      client: #P5 Addressee
      change: #P6 Route {...}
    }
  }
  execution: #M3 Performance {
    explicative: #F3 ClientAffectingService {
      provider: #P7 Performer
      client: #P8 TargetViewer
      change: #P9 Route {...}
    }
  }
```

```
  thanking:  #M4 Utterance {
    interjective: #I2 Thanks
  }
  welcoming: #M5 Utterance {
    interjective: #I3 YoureWelcome
  }
}
```

A dialogue between a wheelchair and an experiment participant consists of multiple exchanges of service such as this one.

5.11 Conclusion

In this chapter, I described the linguistic ontology created based on the linguistic data from the wizard-of-oz experiment. This linguistic ontology was used to support dialogue processing. The largest structures in the ontology are exchanges of services, which are structures comprising a series of dialogue moves. Some of these moves are utterances and are modelled by the linguistic ontology, and some of them are performances, which are not linguistic but are modelled in the same way.

Since both observable phenomena and represented phenomena are modelled in the same way, it is easy to map one onto the other. With this isomorphism, it is easy to determine which service must be carried out and to verify whether the negotiated service was indeed carried out.

Chapter 6

Taxonomy Creation

In Chapter 5, I described the classes of simple things, circumstances, processes, and configurations thereof in the corpus and listed some of their instances. Here I describe the way semantic classes and their instances are mapped to single-word and multi-word expressions found in corpora. The process of taxonomy creation consists of describing both how names and terms are realised by single-word and multi-word expressions and how they are associated with, respectively, actual things and classes of things, circumstances, processes and configurations thereof. In particular, the resulting vocabulary shall not be a mere list of the words spoken in commands, but a list of the expressions that realise taxa and nominals from a linguistic ontology.

Throughout this chapter, I theorise lexis using a systemic and functional approach, filling the theoretical gap in the theory as far as lexis is concerned.

6.1 Language-based concepts

To identify classified entities such as *the glasses* in a situation, a listener must be able to recognise *glasses* in his or her visual senses. If everything goes right in communication, linguistically represented phenomena such as *the glasses* in the command *bring me the glasses* are identified by actual phenomena such as *glasses* in the addressee's visual senses. The glasses perceived by the addressee and represented linguistically in sound waves by the speaker are the same entities in the situation.

Therefore, a linguistic representation of glasses stands for perceived glasses. Linguistic representations are **taxa** or **nominals** as defined in Chapter 5. Cross-linguistic representations equivalent to specific linguistic representations are **concepts**.

6.2 Language-specific terms

Each language has its own representations for phenomena. If participants are talking in German, all representations they construct with German phonemes are German representations of phenomena. From this point on, I shall prefix taxa and nominals from a linguistic ontology with ***deu:*** for German and ***eng:*** for English and I shall prefix concepts with ***lin:*** (cross-linguistic).

Let us consider one example where German has two different taxa for the same phenomenon. For instance, both ***deu:Sofa*** as in *das Sofa* and ***deu:Couch*** as in *die Couch* are redundant to ***lin:Sofa*** as a cross-linguistic representation of a sofa. The concept of sofa is language-based because it is equivalent to linguistic representations in our German corpus. It is also cross-linguistic in the sense that two different languages may have representations of phenomena equivalent to the same concept. For instance, English also has linguistic representations for sofas, namely ***eng:Sofa*** as in *the sofa* and ***eng:Couch*** as in *the couch*. However, a concept is not universal to every language because there are languages such as Ancient Latin and Ancient Greek with no linguistic representations for sofas.

In a situation, a perceived sofa is an instance of ***lin:Sofa***, a sofa represented by *das Sofa* is an instance of both ***deu:Sofa*** and ***lin:Sofa*** and a sofa represented by *die Couch* is an instance of both ***deu:Couch*** and ***lin:Sofa***. A person who wants to refer to a perceived sofa in German must either choose to say *das Sofa* or *die Couch* and the person who wants to identify the sofa linguistically represented must recognise an instance of ***lin:Sofa*** no matter whether the speaker said *das Sofa* or *die Couch*.

Now, if the listener of an utterance wants to refer to the same sofa in his or her turn, he or she has the options in Table 6.1 if the previous speaker said *das Sofa* and the options in Table 6.2 if *die Couch* was said before.

Speaker 1	Speaker 2	
das Sofa	das Sofa	es
'das Sofa	'das Sofa	'das
dieses Sofa	dieses Sofa	dieses

Example 6.1: Dependent options of the second speaker for *das Sofa*

Speaker 1	Speaker 2	
die Couch	die Couch	sie
'die Couch	'die Couch	'die
diese Couch	diese Couch	diese

Example 6.2: Dependent options of the second speaker for *die Couch*

One approach to model this restriction computationally consists of temporarily treating a mentioned entity as an instance of either ***deu:Sofa*** or ***deu:Couch***, that is, entities that become instances of these classes by being referred to as respectively *das Sofa* or *die Couch*. All instances of ***lin:Sofa*** in the situation are similarly tagged for references such as *das andere* and *die andere* (*the other*). This is the way I modelled anaphoric restriction as reported in Chapter 10.

6.3 Expressions derived from terms/names

In the linguistic ontology, I classified linguistic symbols first as representations of entities and qualities and then I subclassified entities and qualities further: entities were subclassified as named entities and classified entities and qualities as name qualities and class qualities. The symbols themselves such as

deu:Rollstuhl, ***deu:Rolland***, ***deu:RollstuhlQuality***, and ***deu:RollandQuality*** were realised by one-word or multiword **expressions**. For instance, the symbol ***deu:Rollstuhl*** can be realised by many expressions including *der Rollstuhl*, *den Rollstuhl*, *dem Rollstuhl*, *des Rollstuhls*, *die Rollstühle*, *die Rollstühle*, *den Rollstühlen*, *der Rollstühle* whereas the symbol ***deu:RollstuhlQuality*** can be realised by a smaller number of expressions including *Rollstuhl* and *Rollstühle*.

Here we can make a connection between terms and names from a domain and linguistic expressions. A term such as *Rolltuhl* is a character string (a literal value) associated with an entity class in a domain. The term *Rolltuhl* can be taken as a **base** from which all the above expressions can be derived, both those realising classified entities and those realising class qualities. The same is true of names such as *Rolland*. Names are character strings associated with an entity and they can also be taken as a base for generation of expressions.

For this thesis, derivation of expressions from terms and names was done manually. No automation was used.

6.4 Multiword expressions

In previous chapters, I used the expressions *prey animal* and *animal of prey* as examples of composite symbols. These expressions realise the atomic symbols ***eng:Animal***, ***eng:PreyQuality***, and ***eng:OfPreyQuality***. In turn, the atomic symbol ***eng:OfPreyQuality*** is realised by a two-word expression, namely *of prey*, that is, a single atomic symbol realised by a two-word expression[1].

In German, the realisation of a symbol by multiple words is particularly difficult because multiword expressions realising a single atomic symbol are not always continuous. Let us consider the multiword expressions in bold in Examples 96 and 97.

(96) Wer **kennt sich mit** Word gut **aus**?
Who **knows** Word well?

(97) Der is einer, der **sich** da**mit** gut **auskennt**?
He is one who **knows** it well.

Expressions such as these represent the cognitive process of knowing how to use some tool in German. They are composed of four words, but they realise a single symbol because no segment of these expressions realise a more general symbol. Moreover, the four words of these expressions are not continuous and they are not in the same order in all clauses. Nonetheless, they could be derived from each other following string-replacement rules (morphology).

In this thesis, the generation of variant expressions was done manually. In Chapter 9, I describe how a combinatory categorial grammar was used (thus can be used) to recognise that words scattered throughout a clause such as these actually realise a single symbol at the semantic stratum.

[1]The fact that the symbol is atomic, not composed of other symbols, does not make the expression realising it indivisible. The expression, not the symbol, comprises two words. In turn, each word defined as indivisible in terms of grammatical composition is represented by a letter or phoneme string. In the same way, this string is divisible at the graphological or phonological stratum even though the word they realise is an indivisible token for grammatical composition.

6.5 Classes of expressions and words

For semantic composition, what counts is the type of phenomenon being described, that is, whether symbols represent an entity or a quality, whether they represent a named entity or a classified entity, and so on. Since symbols are realised by one-word or multiword expressions, it is only natural that an approach favouring semantic composition will also favour classifying full expressions rather than individual words.

Therefore, departing from the tradition of classifying grammatical words with word classes, I classified expressions representing entities as grammatical **nouns**, expressions representing qualities as grammatical **adjectives**, and expressions representing processes as grammatical **verbs**. A set of different expressions such as *Rollstuhl*, *Rollstuhls*, *Rollstühle*, and *Rollstühlen*, which realise the same atomic symbol were taken together as a **lexical item**.

In turn, each grammatical word within an expression was classified as a different noun, adjective, or verb fragment. For instance, the expression comprising the words *kennt*, *sich*, *mit*, and *aus* was classified as a grammatical **verb** and the words within it were classified as, respectively, a **verbal base**, a **verbal reflexive**, a **verbal marker**, and a **verbal particle**. In this way, each expression and each word had its grammatical class: expression classes and word classes.

It is important to emphasize how different such a classification is in practice from a traditional approach to word classes. Expressions such as *Rolland* in *der heißt Rolland* (*it is called Rolland*) are classified as an adjective and the single word within this expression is classified as an adjectival base. The expression *of prey* in *animals of prey* is also an adjective and the words within it are given classes whose names are implementation-specific because no word class label based on terms from the literature is useful at the current stage.

A visualisation of the relations between concepts, taxa, and grammatical expressions can be seen in Figure 6.1.

Recapitulating, concepts are cross-linguistic symbols and taxa and nominals are their linguistic counterparts. Multiple linguistic expressions found in our linguistic data may realise the same linguistic symbol. They are variants of a single lexical item. An expression comprises one or more words. If an expression is a grammatical verb, the words within it are "verbal fragments" and each one of them receives a word class in such terms: verbal base, verbal reflexive, verbal particle, and so on. Finally, some of these word classes are too implementation specific too be worth reporting.

6.6 Systemic Lexis

A vocabulary created in this fashion has two properties: on the generation front, every potential composition of symbols at the semantic stratum will result in a potential composition of expressions at the grammatical stratum if only the correct selection of grammatical features is made; on the analysis front, a combination of recognised expressions at the grammatical stratum is only potential if the combination of realised symbols at the semantic stratum is also potential. From a computational perspective, we gain something in both fronts: a generator can select symbols and combine them without ever needing to backtrack and an analyser/parser can build only lexicogrammatical structures

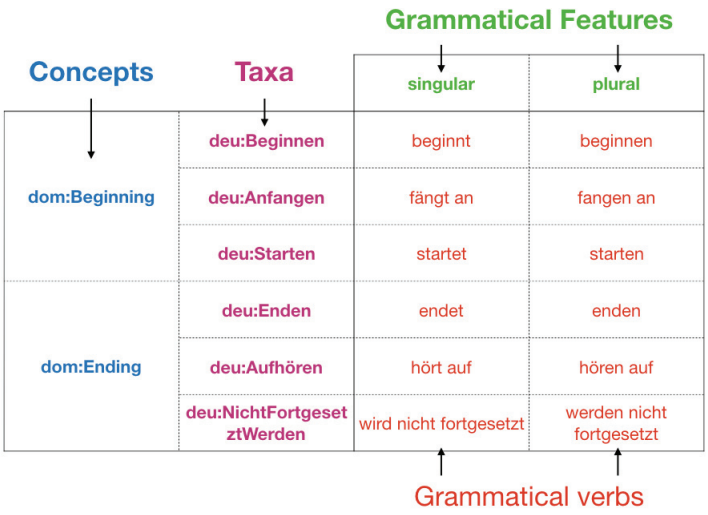

Figure 6.1: Relation between concept, taxa and grammatical expressions

that realise valid semantic structures discarding all semantically incompatible lexicogrammatical structures.

Moreover, this vocabulary is systemic and functional. It is systemic in the sense that concepts can be selected as features in a system (options in a small set) and taxa or nominals can be selected also as systemic features once a concept is selected. At the lexicogrammatical stratum, words are either full symbols or fragments of symbols and they are part of a potentially discontinuous expression. A lexical expression can be selected as a feature in a system corresponding to the columns of the table in Figure 6.1. In turn, a lexical item can be selected as a systemic feature corresponding to the semantic fragment (verbal base, verbal reflexive, verbal particle, and so on). In this way, lexical items are the most delicate level of lexicogrammatical selection because they can be left out as the last options once the full grammatical structure above the word is already defined. A few insertions of functions are required between the selection of a lexical expression and a lexical item in case a lexical expression is realised by more than one word. Every function insertion and ordering that comes after the selection of lexical expressions are lexical, not grammatical. For this reason, the theory presented here counts as a systemic functional theory of lexis.

6.7 Conclusion

In this chapter, I explained how expressions realising indivisible linguistic symbols are catalogued, classified, and grouped into lexical items for lexicogrammatical analysis and generation. Expressions created manually for this thesis could have been automatically generated from terms and names for this domain using rules, but this step was not performed because this was not the focus of

this research.

Part III

Architecture and Implementation

Chapter 7

System Architecture

In the previous three chapters, I described the corpus of wheelchair commands as well as the ontology and the vocabulary created with them, which form together a taxonomy. Here I move on to describe the architecture of the dialogue system that uses this taxonomy. In particular, I describe the blackboard supporting the inter-module communication, the components that take turn in writing onto the blackboard and how the cycle works.

The dialogue system was implemented as a Java application using the DAISIE API (Ross and Bateman, 2009) modified to support a blackboard and a cycle with 14 modules, namely:

1. speech recogniser
2. text producer
3. language analyser (lexicogrammatical analyser)
4. reference integrator
5. figure integrator
6. nexus integrator
7. move integrator
8. move formulator
9. nexus formulator
10. figure formulator
11. referennce formulator
12. language synthesiser (lexicogrammatical realiser)
13. speech producer
14. speech performer (speech synthesiser)

DAISIE dialogue systems follow an agent-based approach to dialogue management, but they also take the complexity of grounding into full consideration. For instance, a DAISIE dialogue system keeps track of proposed and established representations of phenomena as well as manifest phenomena, which allows it to perform grounding subdialogues properly. In particular, when it comes to giving information, a DAISIE dialogue system can keep track of whether a phenomenon was 1) observed by, for instance, the wheelchair, 2) both observed by it and indicated by it in dialogue (proposed representation), 3) or observed by it, indicated by it, and acknowledged by the respondent (established representation), where *not-yet-proposed*, *proposed*, and *established* are alternative representation states. The same stepwise process of updating the representation state is performed when the dialogue system demands information, offers a service, or demands a service. The dialogue system keeps track of representation states, treating information as a state update for representations of phenomena. This is an improvement over previous agent-based dialogue systems that were limited regarding their grounding capabilities. In this thesis, I shall report only on the understanding side of modified DAISIE cycle.

7.1 Blackboard

One of the goals of this thesis was to analyse text lexically and grammatically making sure that grammatical composition enables semantic composition both in terms of experiential semantics and speech acts. This has several impacts on processing.

If an analyser can make decisions of whether a grammatical structure is potential based on whether modifiers, complements, and adjuncts are situationally plausible, the number of potential grammatical structures for the same plain text is drastically reduced. In turn, if a text producer can count on the judgement of such a lexicogrammatical analyser for deciding which of two different ways of incorporating a repair is most likely depending on whether each resulting text is situationally plausible, the number of potential texts for the same recognised speech is reduced.

When further steps of processing are required for making a decision at a previous step, information needs to flow in two directions within a dialogue system. A 'blackboard' is an ideal design pattern for storing editable structures in such cases. A blackboard is a virtual surface where data structures can be written and to which all components can write. One important aspect of blackboards is that every data structure written to it can only be read and updated by components that understand those structures. For this reason, a dialogue system using a blackboard needs to guarantee that all included components can understand and manipulate the same kinds of data structure, creating a technological interdependence between the modules.

In this architecture, candidate recognised segments of speech were represented as a unicode string associated with a numeric confidence value provided by the speech recogniser. In turn, each candidate text was a unicode string resulting from the incorporation of one or more candidate segments of speech to the text produced so far. Each candidate text is analysed lexically and grammatically. For each candidate text, a chart is created with a slot for each substring of the text. Candidate words are added to chart slots corresponding

to substrings that match their spellings. Candidate wordings are built of candidate words and added to the slots that correspond to strings matching their spellings. Since recognised grammatical composition enables semantic composition, each wording corresponds to a symbol. Symbols are integrated one by one with the situation.

Integrating symbols with the situation means identifying mentioned entities in the situation, recognising the role they play in the represented process, roles such as actors, providers, clients, and affected matter, building a logical sequence of events from the current state to the desired state implied by the command, distributing labour between interactants according to their rights and duties, and understanding an utterance as a move in a dialogue flow. All of these establishment states of a command need to be stored somewhere. In this thesis, they are appended to the semantic structures on the blackboard.

7.2 Cycle

There are six components of command understanding after speech recognition: a text producer, a lexicogrammatical analyser, a reference integrator, a configuration integrator, a nexus integrator, and a move integrator. A text producer has the task of producing alternative analysable texts for the same recognised speech so as to overcome disfluencies and incorporate repairs. A lexicogrammatical analyser has the task of analysing candidate texts lexically and grammatically and recognise integratable semantic structures in them. In turn, each integrator has the task of integrating candidate symbols with the situation in order to recognise their full meaning. Figure 7.1 illustrates the cycle.

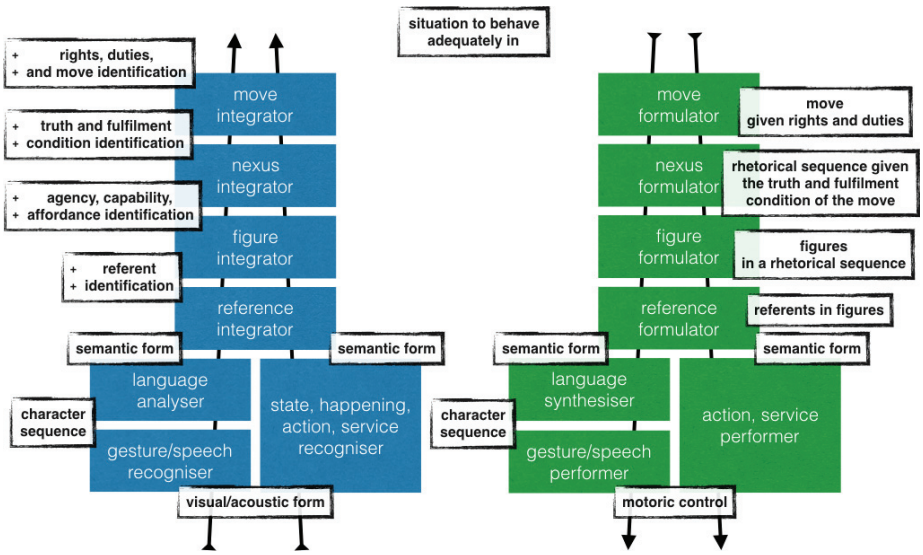

Figure 7.1: Modified DAISIE Cycle

The turn taking by the blackboard works in the following way. Each one of the six components runs as a parallel thread in a java virtual machine. Each one of them is triggered to do their task in a cycle, one after the other, with

50 millisecond pause between each cycle. When triggered, each component first checks whether there is some data waiting for processing. If yes, it checks whether the data is currently being processed by another component. If not, it checks whether all necessary preprocessing was already carried out for its task. And only if this is the case, it starts processing the data.

Such a triggering cycle can result in linear processes between different components where data is processed systematically by a component before being processed by the next. It can also work each time in a different order if there is little need for preprocessing for each task. In practice, for the present domain where utterance meaning is strongly dependent on who can do what, reference must be integrated before configuration and configuration must be integrated before logical inferences most of the time. This results in an almost linear process with frequent interdependence only between the text producer and the lexicogrammatical analyser.

7.3 Conclusion

In this chapter, the architecture of the dialogue system was explained. Emphasis was given to modifications applied to the DAISIE platform, including an added blackboard and a new subdivision of tasks amongst components. In the following chapters, each component will be described individually.

Chapter 8

Text Producer

In this chapter, I describe the component that *consumes* a sequence of recognised speech segments and *produces* lexically and grammatically analysable plain text. This component makes a bridge between speech and text (Section 1.1.1) and operationalises both the notion of repair described in Section 8.2 and the notion of text described in Section 2.1.8.

8.1 Detailed problem

When wheelchair users interact with intelligent autonomous wheelchairs, they talk to it in the way humans usually speak: repeating and rephrasing what they mean in different ways. For instance, Examples 98 and 99 illustrate how speech looks when transcribed by a speech recogniser.

(98) *Roland ich möchte jetzt ein Buch lesen fahr mich bitte ins Schlafzimmer Roland fahr mich bitte ins Schlafzimmer dreh einmal um*

Rolland I want to read a book now please take me to the bedroom Rolland please take me to the bedroom turn around

(99) *Roland fahr bitte zur Tür ich möchte die Tür öffnen fahr mich bitte zur Wohnungsür zur Haustür*

Rolland please go to the door I want to open the door please take me to the apartment door to the house door

As we can see in Examples 98 and 99, the actual commands for the wheelchair to go or take someone somewhere are neither the whole nor a segment of speech, that is, they are not a sound pattern that can be simply recognised in our acoustic impression. Speech is very different from clean spoken commands such as *Roland, ich möchte jetzt ein Buch lesen, fahr mich bitte ins Schlafzimmer* (*Rolland, I want to read a book now, please take me to the bedroom*) for Example 98 and *Roland, fahr mich bitte zur Haustür, ich möchte sie öffnen* (*Rolland, please take me to the house door, I want to open it*) for Example 99. For this reason, an intelligent wheelchair must be able not only to recognise speech, but also to turn the recognised speech into a clean spoken text, which can in turn be understood as a command.

In this chapter, I describe the process of manually transcribing speech and manually writing down the intended commands. In this process, I explain the informal algorithm used by us human transcribers, which served as the basis for implementing the text producer utilised in the final experiment. At the end, I describe the simplified text-producing process implemented.

8.2 Speech as text production

Elisa Vales executed a manual speech transcription of our recordings aiming at capturing events of adding segments to the text produced thus far and replacing or deleting textual segments. As researchers, we can produce different numbers of speech segments depending on the minimal pause length that we consider a speech segment boundary. To minimise the effects of our operationalisation, Elisa Vales transcribed all recordings considering a speech segment boundary every speech pause equal or longer than 250ms. She also annotated the length of the pause in k, where k is such that the pause is equal or longer than $250ms \times 2^k$ and shorter than $250ms \times 2^{k+1}$. As a result, we have access to different number of segments depending on the minimal pause length we want to consider a segment boundary. The length of speech segment boundary in our transcriptions is represented by i in sb_i. The initial boundary at the beginning of a speech turn is called sb_{start} and the final boundary is called sb_{end}.

Example 100 shows a transcribed speech with two segments bounded by pauses of length 1.

(100) (sb_{start}) komm·hier·zu·mir·drehst·du·dich (sb_1) einmal·um (sb_1)

If we treat these speech segments as text production, there were three versions of the produced text at different speech segment boundaries: one at the initial speech segment boundary (sb_{start}), the second at the second boundary and the third text version at the final boundary.

text version	text produced thus far
1	
2	komm·hier·zu·mir·drehst·du·dich
3	komm·hier·zu·mir·drehst·du·dich·einmal·um

Example 8.1: Versions of text produced thus far in Example 100

In Example 100, two clauses are spoken: namely *komm hier zu mir* (*come here to me*) and *drehst du dich einmal um* (*will you please turn around*). However, the clause boundary does not match the speech segment boundary, therefore a dialogue system relying on the assumption that a speech segment is a complete clause or a complete clause complex is not able to understand this example. To solve this mismatch, a text producer can be used to consume these two speech segments and produce a plain text resulting from their concatenation.

Example 101 illustrates a case where incorporation of the second speech segment demands a repair of the text produced thus far.

(101) (sb_2) einmal·umdrehen·bitte·und·einmal·hier·bitte·zum
(sb_3) zur·spüle·und·stopp·stopp (sb_{end})

In Example 101, a typical pattern of repair occurs. In German, words such as *zum* and *zur* result from the contraction of respectively *zu dem* and *zu der*, where *dem* and *der* agree with a word that follows. However, speakers often speak the words *zum* and *zur* before they choose the actual word they are about to speak next. Because of this, they repair their choice between *zum* and *zur* in the following speech segment by speaking the correct one at the beginning of the new segment. To incorporate this repair, a text producer needs to be informed that the concatenation of the two speech segments does not result in a lexically and grammatically analysable text. Only then can it test replacements until it can produce an analysable text.

text version	text produced thus far
1	
2	...·und·bitte·**zum**
3	...·und·bitte·zur·spüle·und·stopp·stopp

Example 8.2: Versions of text produced thus far in Example 101

This example also includes a repair within a speech segment. The wording *und einmal hier* (*and please here*) in the first speech segment is partially replaced by *bitte zum*) (*please to the*) resulting *und bitte zum* (*and please to the*). Since this repair is performed within a speech segment, a smaller unit than a speech segment needs to be used by the text producer. In this thesis, I used the **foot**, a rhythmic unit composed of one stress syllable surrounded by unstressed ones.

The first speech segment reads *// einmal umdrehen / bitte / und einmal hier / bitte zum //* (*// please turn around / please / and please here / please to the //*) where single slashes represent the foot boundaries and double slashes represent tonal curve boundary. Text versions can be seen in Example 8.3.

text version	text produced thus far
1	
2	**einmal**·umdrehen
3	bitte·umdrehen
4	bitte·umdrehen·und·**einmal**·**hier**
5	bitte·umdrehen·und·bitte·zum

Example 8.3: Versions of text produced thus far within a speech segment

To produce an analysable text, a text producer can apply the same strategy to feet as the one applied to speech segments, that is, to treat each foot as a text-producing events on its own, allowing them to be understood as repair the text produced thus far. For such an approach to speech integration to work, the text producer needs feedback from the text analyser, telling whether each potential incorporation of foot results in an analysable text.

8.3 Discourse contributions

Let a **discourse contribution** be a wording that either selects an addressee as in *Rolland* (*Rolland*) or *Rollstuhl* (*wheelchair*), a complete independent clause

such as *fahr mich zur Küche* (*take me to the kitchen*), a fragment of such a clause such as *zur Küche* (*to the kitchen*), or an interjection such as *ok* (*ok*).

In our corpus, we observed that only the currently produced discourse contribution was ever repaired through word replacement such as replacing *zum* by *zur* or *einmal* by *bitte*. This means that a discourse contribution boundary can serve as a limit for repairing the text produced thus far. If we can rely on the assumption that this pattern is also present in other corpora, developing a text producer becomes a drastically simpler task than if the whole text produced thus far is subject to word replacements such as these. In this thesis, I made this assumption and it was confirmed in the final experiment when the dialogue system was tested.

However, there are two other kinds of repair that require treatment: 1) discourse contributions might be replaced as a whole by new ones and 2) a whole chain of nominal groups might be replaced at once by a new one. These cases will be illustrated in Example 102. This example contains the same speech transcription as Example 99, now separated by slashes. Three slashes /// indicate the boundaries of a dialogue turn_2. Two and one slashes // or / indicate the boundaries of feet. The feet are numbered linearly from 1 to 9.

(102) *$///_1$ Roland $//_2$ fahr bitte $/_3$ zur Tür $//_4$ ich möchte $/_5$ die Tür $/_6$ öffnen $//_7$ fahr mich bitte $/_8$ zur Wohnungstür $//_9$ zur Haustür ///*

Rolland please go to the door I want to open the door please take me to the apartment door to the house door

As proposed at the beginning of this chapter, Example 102 is the production of the text in Example 105. Feet 1-6 in Example 102 can be understood as inserting the first version of **phonological forms** a-f in Example 103; feet 7-8 in Example 102 can be understood as replacing the first version of phonological forms b, c, and e in Example 103 by their second version in Example 104. The feet b and c constituted a contribution, which was replaced as a whole and feet c and e constituted a cohesive chain that was also replaced as a whole. Finally, foot 9 in 102 can be understood as replacing the second version of phonological forms c' and e' in Example 104 by their third version in Example 105. This time, c' and e' are still a cohesive chain replaced as a whole.

(103) Version 1: *$///_a$ Roland $//_b$ fahr bitte $/_c$ zur Tür $//_d$ ich möchte $/_e$ die Tür $/_f$ öffnen ///*

Rolland please go to the door I want to open the door

(104) Version 2: *$///_{a'}$ Roland $//_{b'}$ fahr mich bitte $/_{c'}$ zur Wohnungstür $//_{d'}$ ich möchte $/_{e'}$ die Wohnungstür $/_{f'}$ öffnen ///*

Rolland please take me to the apartment door I want to open the apartment door

(105) Version 3: *$///_{a''}$ Roland $//_{b''}$ fahr mich bitte $/_{c''}$ zur Haustür $//_{d''}$ ich möchte $/_{e''}$ die Haustür $/_{f''}$ öffnen ///*

Rolland please take me to the house door I want to open the house door

It is the text in Example 105 that should be taken as a discourse move, i.e. as something that can be understood as an attempt of achieving something with words, not the transcribed speech itself.

8.4 Discourse contributions to ignore

In the previous sections, I presented a method for producing a spoken text based on transcribed speech where one or more feet replace either a full contribution in the text produced thus far or a full cohesive chain. In this section, I present an example where a sequence of feet must be ignored and not added to the text produced thus far. In addition, I also show how the choice of different minimal pause lengths to recognise a boundary for speech segment affects the automation explained in the next section.

To formalise the move-contribution relation, I consider five complementary ways of delimiting speech and text segments: at speech stratum, there is the full utterance produced by a speaker recognised by the listener (**u**); at the semantic stratum, there is the full dialogue move made by the speaker and understood by the listener (**m**). An utterance comprises one or more speech segments separated by pauses (**s**). The smaller the speech pause we take as a speech segment boundary, the more speech segments we encounter. A dialogue move comprises one ore more discourse contributions (**c**). Speech segment are subdivided into feet (**f**) so that segment-internal and cross-segment repairs can be well processed. Example 8.4 from the second run of the Wizard-of-Oz experiment illustrates how these segments map to recordings. On the left side, we can see dialogue moves and discourse contributions found on the text produced thus far. On the right side, we can see utterances and speech segments.

<table>
<tr><td rowspan="4">m_{56}</td><td>$c_{76,R}$</td><td>Roland?
Rolland?</td><td rowspan="2">$s_{0,56}$</td><td rowspan="4">$s_{1,56}$</td><td rowspan="4">$s_{2,56}$</td><td rowspan="8">$u_{23,H}$</td></tr>
<tr><td>$c_{77,R}$</td><td>ich mchte jetzt ein Buch lesen.
I want to read a book now.</td></tr>
<tr><td rowspan="2">$c_{78,R}$</td><td>fahr mich bitte
please take me</td></tr>
<tr><td>ins Schlafzimmer.
to the bedroom.</td><td>$s_{0,57}$</td></tr>
<tr><td rowspan="2">m_{57}</td><td>$c_{79,R}$</td><td>Roland?
Rolland?</td><td>$s_{0,58}$</td><td>$s_{1,57}$</td><td>$s_{2,57}$</td></tr>
<tr><td>$c_{80,R}$</td><td>fahr mich bitte ins Schlafzimmer.
please take me to the bedroom.</td><td>$s_{0,59}$</td><td>$s_{1,58}$</td><td rowspan="2">$s_{2,58}$</td></tr>
<tr><td>m_{58}</td><td>$c_{81,R}$</td><td>dreh einmal um?
just turn around?</td><td>$s_{0,60}$</td><td>$s_{1,59}$</td></tr>
<tr><td rowspan="2">m_{59}</td><td>$c_{82,H}$</td><td>ok. ((starts motion))
ok. ((starts motion))</td><td>$s_{0,61}$</td><td rowspan="2">$s_{1,60}$</td><td rowspan="2">$s_{2,59}$</td><td rowspan="2">$u_{24,R}$</td></tr>
<tr><td>$c_{83,H}$</td><td>ich fahre dahin.
I'll go there.</td><td>$s_{0,62}$</td></tr>
</table>

Example 8.4: Acoustic events in the fifth service exchange of the second WoOz run

H: Human, R: Robot, $u_{i,k}$: Utterance i of Interactant k, m_i: Move i, $m_{i,k}$: Contribution i directed to Interactant k, $s_{k,i}$: Speech segment i delimited by pauses of at least $250ms \times 2^k$

In the informal algorithm applied by researchers, most of the time each

speech segment in a single utterance are taken to add or replace words to previous drafts of the spoken text. Sometimes, though, as in Example 8.4, this is not the case. In this example, Move 57 is a rewording of Move 56 as a whole. This move was probably made because the wheelchair — remote-controlled by the Wizard-of-Oz — took a long time to react. Probably for the same reason, Move 58 was made. However, though Move 57 is a rewording of Move 56, Move 58 is not. Here the user changed his or her control strategy, that is, he or she stopped giving commands representing the whole movement from the current position to the final destination and started giving commands representing directional movements such as turning around, going forward, and turning left or right. Since such a drop is not desirable, the researcher chose to respond to one of the two first moves and not to the third. Because of this, either Move 56 or 57 (or both) was acknowledged by the wheelchair and Move 58 was ignored.

Though two out of three moves were ignored, no speech segment in this example produces a text segment substituted in the text produced thus far. Example 8.5 shows an example of long-distance draft update in which a full chain of referring expressions is replaced. Square brackets indicate the boundaries of grammatical structures above the word projected on to the clause by semantic composition.

m_{56}	$c_{94,R}$	*Roland?* *Rolland?*	$s_{0,92}$
	$c_{95,R}$	*a:[fahr bitte b:[zu c:[r Tr.]]]* *please go to the door.*	
	$c_{96,R}$	*d:[ich mchte e:[die Tr] ffnen.]* *I'd like to open the door*	
$m_{56'}$	$c_{95',R}$ $c_{96',R}$	*f:[fahr mich bitte g:[zu h:[r Wohnungstr.]]]* *please take me to the apartment door.*	$s_{0,93}$
$m_{56''}$	$c_{95'',R}$ $c_{96'',R}$	*i:[zu j:[r Haustr.]]* *to the building door.*	

Example 8.5: Speech segments in the ninth service exchange of the second WoOz run

H: Human, R: Robot, $t_{i,k}$: Turn i of Interactant k, u_i: Utterance i, $a_{i,k}$: Address i to Interactant k, $s_{k,i}$: Speech i delimited by pauses of at least $250ms \times 2^i$

version	addresses 95 and 96
1	
2 m_{56}	a:[fahr bitte b:[zu c:[r Tür.]]] d:[ich möchte e:[die Tür] öffnen.]
3 $m_{56'}$	a':[fahr mich bitte b':[zu c':[r Wohnungstür.]]] d':[ich möchte e':[die Wohnungstür] öffnen.]
4 $m_{56''}$	a'':[fahr mich bitte b'':[zu c'':[r Haustür.]]] d'':[ich möchte e'':[die Haustür] öffnen.]

Example 8.6: Text versions after each speech segment of Example 8.5 (first contribution ignored for space)

In Example 8.5, the speech segment $s_{0,92}$ is an additive revision that is responsible for adding three contributions constituting a move to the previous empty draft: namely the move m_{56} *Roland? fahr bitte zur Tür. ich mchte die*

Tr öffnen. (*Rolland? please go to the door. I would like to open the door.*). This speech segment corresponds to the first versions in Example 8.6.

The following speech event $s_{0,93}$ was taken to contain two substitutive revisions, each one of them affecting two contributions in the previous draft. The first substitutive revision was taken to transform the first draft (first version) into *Roland? fahr mich bitte zur Wohnungstür. ich möchte die Wohnungstür öffnen.* (*Rolland? please take me to the apartment door. I would like to open the apartment door.*) where the underlined segments are the substituted parts of the draft at the semantic stratum (second version).

Following this, another revision took place whereby the apartment door is substituted by a building door. The final textual product is *Roland? fahr mich bitte zur Haustür. ich möchte die Haustür öffnen.* (*Rolland? please take me to the building door. I would like to open the building door.*). The second and third versions of the utterance were labelled respectively $m_{56'}$ and $m_{56''}$.

As we can see in these two examples, the informal process of determining the text spoken by the user was complex in two senses: deciding which dialogue move to take as the command and deciding what text should count as the final version of each dialogue move. In the next section, I present the formalisation of this process and I explain the module that produces texts implemented for this thesis.

8.5 Text producer implemented

The text producer implemented integrates a subset of the repairs found in our corpus. In our data, most utterances comprising two or more speech segments could be turned into grammatical spoken text just by concatenating the speech segments. If the wheelchair had reacted faster, most contribution replacements would probably not have happen. The only frequent word repair was the one where a word such as *zum* or *zur* is replaced by another. For all these reasons, we opted for implementing a text producer that does the following actions.

1. whenever a speech segment is recognised, it consumes the speech segment and offers it as textual segment to be analysed
2. if the textual segment is analysable, it marks the textual segment as established
3. if the textual segment is not analysable, it marks the textual segment as buffered
4. if two or more textual segments are buffered, it offers the lexicogrammatical analyser
   (a) the concatenation of the latest segments
   (b) concatenations of the latest segments ignoring the last incomplete foot of one of the non-final segments
5. the text producer consumes all buffered segments whenever a textual segment is marked as established, whether or not the established textual segment was created out of all buffered textual segments or not

No replacement of contributions and no replacement of cohesive chains was implemented. Nonetheless, the text producer worked for all but two utterances for the initial linguistic data once we discarded the speech segments we considered would not have happened, had the wheelchair reacted in a timely fashion from the first interaction on. The assumption was that the wheelchair would respond faster in the evaluation experiment, which was unfortunately not the case. In addition, we also assumed that the speech recogniser would be able to recognise incomplete text fragments. This was the case for small custom grammars we created, but since our final grammar had too much variation and had to be trimmed down, we had to remove all clause fragments from it before the final evaluation experiment. This means that, even though the text producer was used in the evaluation, it had no impact in the results, whether positive or negative.

8.6 Conclusion

In this chapter, I described the manual speech transcription of our recordings and the informal algorithm used by researchers to manually transform the transcribed speech into what we understand as the text spoken by the user. I also presented our assumption that most complicated repairs were due to the long delay between user utterances and wheelchair responses, which motivated the text producer we implemented. Finally, I described how the text producer works and why it had no impact (positive or negative) in the evaluation experiment.

Chapter 9

Lexicogrammatical Analyser

In this chapter, I describe the component that analyses wordings in terms of lexis and grammar. This component *utilises* the taxonomy described in Chapter 6 and a Combinatory Categorial Grammar described in the following to *consume* a grammatical wording and *produce* a shallow semantic structure where symbols are instances of the symbol classes described in Chapter 5.

9.1 Semantic composition

As presented in the introduction, I established four goals for automatic lexicogrammatical analysis.

1. grammatical composition must enable semantic composition both in terms of represented phenomena and discourse contributions.

2. groups and phrases must correspond to semantic elements whereas clauses must correspond to semantic configurations of elements.

3. a multiword expression such as *dreht sich* (*turns*) must be treated as an instance of a single linguistic symbol

4. a grammatical structure boundary must be recognised inside a written word such as *Küchentisch*.

When describing symbols as structures composed of other symbols, we inevitably need to treat semantic structures compositionally. The additional commitment imposed onto lexicogrammatical analysis in this research is that only words that correspond to a symbol can count as a clause constituent. Let us consider two examples to make this commitment clearer (Examples 106 and 107).

(106) The book is on the table.

(107) The book belongs to the officer.

If a parser considers only part-of-speech tags such as nouns (n), determiners (d), verbs (v), and prepositions (p), these two clauses must have the same grammatical structure because they are both composed of the following tag sequence: d n v p d n. See tagged Examples 108 and 109:

(108) The book is on the table.
d n v p d n

(109) The book belongs to the officer.
d n v p d n

As a result, a strictly formal parser will recognise a prepositional phrase in both Examples 108 and 109: namely, *on the table* and *to the officer*. However, if only wordings that represent a phenomenon can be taken as a grammatical constituent, the parser cannot do this. The wordings *the book*, *the table*, and *the officer* represent classified entities and the wording *on the table* represents the location of a book, a location relative to a table. In contrast, the wording *to the officer* does not represent anything on its own. The officer is represented as the owner of the book and the book is represented as the officer's belonging. The wording *belongs to* represents this relation between the two. In this case, if only wordings that represent a phenomenon such as an entity, a spatial location and a relation can count as a grammatical constituent, there is no prepositional phrase in Example 109. A parser that observes this commitment cannot produce the same grammatical structure for the two clauses. Furthermore, whenever this commitment is observed, grammatical constituents can be directly mapped to semantic elements and a clause can be directly mapped to a semantic figure.

In addition to considering experiential structure as described above, in this thesis lexicogrammatical analysis must also consider compositional symbols realising further specifications of a dialogue move. Let us consider Examples and for illustrating this commitment.

(110) Can you take me to the kitchen?

(111) Can you please take me to the kitchen?

For a symbol such as Example 111 to comprise two symbols, namely the entire Example 110 plus the word *please*, Example 111 must be more specific than Example 110. If this is the case, we recognise the meaning of Example 110 and append a further specification realised by the word *please*.

When both these commitments are observed, it follows that semantic structures are associated with grammatical structures at different ranks and it follows that a parser can produce a directly integratable semantic structure during parsing, that is, semantic structures that do not need to be interpreted nor mapped to any other structure for further processing and that can be accessed by multiple components on a blackboard.

In this approach, semantic elements such as simple things, simple qualities, spatial locations (absolute or relative), and processes are represented by wordings such as groups and phrases and configurations of such elements (figures) are represented by simple clauses. Sequences of such figures are represented by clause complexes. Each step in this composition is a **rank** in a **rank scale** (Halliday and Matthiessen, 2014, p. 5). The correspondences between ranks of grammatical and semantic structure is given in Table 9.1.

Grammatical Ranks	Semantic Ranks
clause complex	Sequence
clause	Figure
group/phrase	Element

Example 9.1: Rank Scale

This functional way of recognising grammatical structures differs from non-functional approaches in aspects relevant for discourse. At the clause rank, it allows us to make a distinction between complete figure representations such as the underlined wording in *gehen wir zum Arbeitstisch um mein buch holen* (*let's go to the office table to pick up my book*) from fragments of figure representations that happen to contain infinitive verbs such as the underlined wording in *versuch mal, mich in die Küche zu fahren* (*please try to take me to the kitchen*). The former is a 'complete' grammatical structure whereas the later is an incomplete one, that is, a fragment of a complete one. This means that a clause (c) from a functional perspective might correspond to a sentence (s) or to a predicate (vp) from a nonfunctional perspective, and that some predicates (vp) might not correspond to any complete grammatical structures and be simply clause fragments that are ignored by the parser.

In addition, by adopting a rank scale, we can postulate that all grammatical constituents must correspond to a semantic constituent or a semantic feature of the associated semantic structure. This makes the task of parsing and generation very straight forward: a parser can build up the grammatical constituents and their corresponding semantic constituents before fitting them in grammatical and semantic composite structures (bottom up); and a generator can take a semantic structure as a specification for a grammatical structure and generate composite grammatical structures as sequences of slots to be filled by grammatical constituents (top down). This simplifies both processes.

Another benefit of a rank scale for parsing is that a clause complex is directly constituted of other clause complexes or simple clauses while clauses are constituted solely of groups and phrases. The ranked correspondence between semantic structures and grammatical structures makes it impossible for simple clauses to contain other clauses since a configuration of semantic elements (what the clause represents) is not a semantic element (what groups and phrases represent). Therefore, the adoption of a rank scale is responsible for grammatical structures becoming not so high as the constituency structures of non-functional grammars. The structure of long clauses such as the ones in Table 9.2 illustrate how flat rank structures are. In turn, this flatness facilitates both semantic parsing and generation.

By modelling grammatical composition so that it corresponds to semantic composition, we are able to recognise some constituents individually as representations of semantic elements: for instance, we can recognise the verbal groups *bringst* (*take*) and *fährst* (*take*) as representations of the material process of taking someone somewhere, the nominal groups *du* (*you*) and *mich* (*me*) as representations of interactants in the situation, and the prepositional phrases in the last column of Table 9.2 as representations of the destination where someone is to take someone else. By associating groups and phrases with the semantic elements they represent, treating groups/phrases as fillers of grammatical frames

clause				
group	group	group	group	phrase
bringst	*du*	*mich*	*bitte*	*in die Küche*
bringst	*du*	*mich*	*bitte*	*ins Bad*
bringst	*du*	*mich*	*bitte*	*ins Badezimmer*
bringst	*du*	*mich*	*bitte*	*ins Schlafzimmer*
bringst	*du*	*mich*	*bitte*	*ins Wohnzimmer*
bringst	*du*	*mich*	*bitte*	*zur Küche*
bringst	*du*	*mich*	*bitte*	*zum Bad*
bringst	*du*	*mich*	*bitte*	*zum Badezimmer*
bringst	*du*	*mich*	*bitte*	*zum Schlafzimmer*
bringst	*du*	*mich*	*bitte*	*zum Wohnzimmer*
fährst	*du*	*mich*	*bitte*	*in die Küche*
fährst	*du*	*mich*	*bitte*	*ins Bad*
fährst	*du*	*mich*	*bitte*	*ins Badezimmer*
fährst	*du*	*mich*	*bitte*	*ins Schlafzimmer*
fährst	*du*	*mich*	*bitte*	*ins Wohnzimmer*
fährst	*du*	*mich*	*bitte*	*zur Küche*
fährst	*du*	*mich*	*bitte*	*zum Bad*
fährst	*du*	*mich*	*bitte*	*zum Badezimmer*
fährst	*du*	*mich*	*bitte*	*zum Schlafzimmer*
fährst	*du*	*mich*	*bitte*	*zum Wohnzimmer*

Example 9.2: Clause constituents representing semantic elements or features

and semantic elements as fillers of semantic frames, the clauses in Table 9.2 can be treated as the grammatical frame *(A) (B) (C) bitte (D)* that is associated with the semantic frame Frame(#A,#B,#C,#D). For the clause *bringst du mich bitte in die Küche* (*would you please take me to the kitchen*), the filled semantic frame would be Frame(Bringen, Du, Ich, In-die-Küche). In-die-Küche is not further analysed here because this chapter focuses on clause structure, not group and phrase structure.

Once we have such an association between grammatical and semantic frames, the process of parsing becomes the action of recognising groups and phrases that represent semantic elements and then to recognise the grammatical frame in which they fit. In parallel, other groups and phrases such as *bitte* should not be understood as representations of any semantic element, but rather as specifying a way of enacting a command (see Table 9.3).

In Table 9.3, the adverbial *doch* presents the command as the opposite of what was asked for beforehand, the opposing adverbial *einfach* presents the command as the same as what was asked for beforehand, *mal* presents the service as a favour, and *bitte* presents the request of service as a typical request for the situation. Any combinations of these adverbials are expected in this position of the clause when the clause is to be understood as a command. However, since they further specify the kind of command that is being enacted, they are optional. Such an optional linguistic variation of further specification needs a different treatment from the previous one of completion. There is no grammatical slot to be associated with a semantic slot. In this case, we are dealing with a kind of dialogue move and further specifications of it in terms of how it relates

clause				
group	group	group	group	phrase
bringst	*du*	*mich*	*doch mal bitte*	*in die Küche*
bringst	*du*	*mich*	*doch mal*	*in die Küche*
bringst	*du*	*mich*	*doch bitte*	*in die Küche*
bringst	*du*	*mich*	*doch*	*in die Küche*
bringst	*du*	*mich*	*mal bitte*	*in die Küche*
bringst	*du*	*mich*	*mal*	*in die Küche*
bringst	*du*	*mich*	*bitte*	*in die Küche*
bringst	*du*	*mich*		*in die Küche*

Example 9.3: Clause constituents realising interpersonal features

to the situation.

In addition, as I discussed in previous chapters, we need to make the distinction between grammatical words as the indivisible parts of a wording and linguistic symbols. Linguistic symbols are composed of one or more grammatical words and they may be scattered throughout a clause. Tables 9.4, 9.5, 9.6 show examples of scattered lexical words.

clause			
part of group$_{\mathrm{A}}$	group	group	part of group$_{\mathrm{A}}$
kommst	*du*	*bitte*	*her*

Example 9.4: Divisible Verb *her kommen* (*coming here*)

clause				
part of group$_{\mathrm{A}}$	group	part of group$_{\mathrm{A}}$	group	phrase
drehst	*du*	*dich*	*bitte*	*zum Waschbecken*
drehst	*du*	*dich*	*bitte*	*zum Waschbecken hin*
drehst	*du*	*dich*	*bitte*	*zur Tür*
drehst	*du*	*dich*	*bitte*	*zur Tür hin*
drehst	*du*	*dich*	*bitte*	*zu mir*
drehst	*du*	*dich*	*bitte*	*zu mir hin*

Example 9.5: Reflexive Verb *sich drehen* (*turning*)

clause				
part of group$_{\mathrm{A}}$	group	part of group$_{\mathrm{A}}$	group	part of group$_{\mathrm{A}}$
lädst	*du*	*dich*	*bitte*	*auf*

Example 9.6: Reflexive Divisible Verb *sich auf laden* (*recharging oneself*)

In such cases, the number of slots in the grammatical frame is bigger than the number of slots in the semantic frame. For instance, *kommst du bitte her* (*would you please come here*) would need three grammatical slots as in *(A) (B) bitte (C)*, but only two semantic slots as in Frame(#CA,#B). Once these slots

are filled, the semantic structure is Frame(HerKommen, Du). For *drehst du dich bitte zu mir hin* (*would you please turn towards me*), the semantic structure is Frame(SichDrehen, Du, Zu-Mir-Hin). For *lädst du dich bitte auf* (*would you please recharge yourself*), it is Frame(SichAufLaden, Du).

In this chapter, I describe how to parse German clauses with a combinatory categorial grammar so as to recognise semantic frames of this particular kind: namely **semantic figures**, which are configurations of semantic elements where one element is an ongoing process and the others are either participants in that process or circumstances in which that process takes place (see Chapter 5). And I show how to do this while acknowledging multiword expressions and further specifications of the discourse contribution. This will make composite semantic structures fit the requirements of further processing steps on a blackboard.

In the following, I describe the custom parsing resource, a Combinatory Categorial Grammar (CCG, Steedman, 2004), created for lexicogrammatical analysis, thereby showing that Combinatory Categorial Grammars are compatible with the listed analytical goals of this thesis.

9.2 Combinatory Categorial Grammars

Since I make the commitment that a semantic frame at the rank of clause must be a semantic figure, the way of implementing a combinatory categorial grammar (CCG) proposed in this chapter is quite different from the standard way CCGs are implemented nowadays for German as well as for any other language.

Usually, in a CCG, words are treated as symbols and they are given either a simple category such as S for sentences, N for nouns, NP for noun phrases, PP for prepositional phrases or a complex category such as NP/N for determiners, PP/NP for prepositions and S\NP, S\NP/NP, S\NP/PP, etc. for verbs. Complex categories such as X/Y and X\Y are functors that consume an argument of type Y and produce a composite structure of type X. A forward slash indicates that the argument must be after the functor and a backslash indicates that the argument must be before the functor. Let us consider the tagged Example 112:

(112)

The	book	is	on	the	table.
NP/N	N	S\NP/PP	PP/NP	NP/N	N

The determiner *The* is an instance of NP/N. It consumes the instance of N *book* after it and produces an instance of NP corresponding to the wording *The book*. This process is called **forward application**. A similar forward application of the NP/N *the* to the N *table* results in the NP *the book*. The application of the PP/NP *on* to the NP *the book* results in the PP *on the book*. In turn, the application of the S\NP/PP *is* to the PP *on the book* results in the S\NP *is on the book*, an incomplete grammatical structure. Finally, the **backward application** of S\NP *is on the book* to the preceding NP *The book* results in the S *The book is on the table*.

Diverging from this tradition, in the present resource all groups and phrases representing participants and their attributes are instances of the same simple category no matter whether they are simple things, simple qualities, or spatial locations. We could do this because, at the semantic stratum, complements are also restricted by the type of semantic element allowed. Since all semantic element types (entities, named entities, classified entities, and so on) belong

to a type hierarchy, they offer a much more precise filtering mechanism than typical simple categories regarding allowed semantic composition. The work left to grammatical categories is solely to respect the rank scale.

Respecting the rank scale is operationalised in three steps: a clause receives the simple category **c** and mentions of elements (groups or phrases) receive the simple category **m**. Words that are not symbols by themselves such as the word *to* in the clause *the book belongs to the officer* receive the simple category **w**.

If we take the current tiny grammars of English and German in OpenCCG (http://openccg.sourceforge.net) as reference, we have the following parallels. First, the category *c* (clause) corresponds roughly to the standard category *s* except for the fact that some sentences in the tiny grammar are not clauses in ours because they are incomplete wordings. Second, the category *m* (mention) corresponds to one of many different categories including *np*, *pp*, *ap*, *adj*, and *n*, but not to all instances of them. An *m* instance corresponds to a complete wording that represents a simple thing, a simple quality, a spatial location, a spatial route only if they are participants in processes.

Table 9.7 shows examples of categories for the constituents of a series of similar clauses.

Category	**A:c/Y:m/X:m**	**B:m**	**C:m**
Semantic Class	A:Figure(P:Process, X:Thing, Y:Route)	B:Thing	C:Route
#1	*fährst*	*du*	*dahin*
#2	*fährst*	*du*	*hierhin*
#3	*fährst*	*du*	*zu mir*
#4	*fährst*	*du*	*zum Bett*
#5	*fährst*	*du*	*zum Sofa*
#6	*kommst*	*du*	*dahin*
#7	*kommst*	*du*	*hierhin*
#8	*kommst*	*du*	*zu mir*
#9	*kommst*	*du*	*zum Bett*
#10	*kommst*	*du*	*zum Sofa*

Example 9.7: Examples of rank tags for the constituents of similar clauses

In Table 9.7, the word *du* (*you*) is given the grammatical category **m** and the semantic class **Thing**, which are associates in the association **B**. In addition, the wordings *dahin* (*there*), *hierhin* (*here*), *zu mir* (*to me*), *zum Bett* (*to bed*), and *zum Sofa* (*to the sofa*) are given the grammatical category **m** and the semantic class **Route**, which are associates in the association **C**. Finally, the words *fährst* (*would roll*) and *kommst* (*would come*) are given the grammatical category **c/m/m** and the semantic class **Figure**. A is the association between the clause and the represented figure whereas X and Y are the associations between the missing clause constituents after *fährst*/*kommst* and the missing elements in the represented figure.

Here we can see another difference between the approach adopted in this thesis and the one usually adopted in the open-source tiny grammars. The task of distinguishing between *du* as a type of complement and *dahin*, *hierhin*, *zu mir*, *zum Bett*, *zum Sofa* as another type of complement is done here at the semantic stratum, not the grammatical stratum. It is the fact that *du*

represents a simple thing and that *dahin*, *hierhin*, *zu mir*, *zum Bett*, *zum Sofa* represent routes that makes them become potential fillers of the semantic slots for, respectively, a simple thing and a route. Traditionally, this distinction is achieved at the grammatical stratum through categories such as *NP*, *PP* and *adv*. As a consequence, *hierhin* (*here*) is traditionally grouped together with *sofort* (*in just a sec*) as adverbial groups and *zum Bett* (*to the bed*) with *in eine Stunde* (*in an hour*) as prepositional phrases, which leads to *sofort* in *ich komme sofort* (*just a sec, I'm coming*) being mistaken for a complement of *komme*. By shifting the complement distinction to the semantic stratum, we can group adverbial groups such as *dahin* and *hierhin* with prepositional phrases such as *zu mir*, *zum Bett* and *zum Sofa* as mentions of routes and differentiate them from *sofort* and *in eine Stunde* as mentions of a time relative to now.

This shift results in a smaller grammatical resource since we do not need to write down different categories for combinations with adverbial groups and prepositional phrases and also in a smaller number of wrong parses since adjuncts are not so often mistaken for complements. Further benefits shall be presented in the following sections when modelling performance processes such as *machen* (*do*) in *eine Mundspülung machen* (*do a mouthwash*) and voice auxiliaries such as *werden* and *bekommen*.

9.3 Words as grammatical complements

The clause constituents in Table 9.7 received the same categories because they were very similar both in semantic and in lexicogrammatical constituency. This does, however, not always happen. Sometimes clause constituents are not complete mentions of semantic elements, they are sometimes fragments of lexical words. For instance, whereas the processes of *brushing one's teeth* (*die Zähne zu putzen*) and *washing one's hands* (*die Hände zu washen*) are represented by single-word lexical verbs, the processes of *washing one's mouth out* (*den Mund auszuspülen*) and *creaming one's face* (*das Gesicht einzucremen*) are represented by two-word lexical verbs. Semantically, though, all these processes have two participants: namely an actor and a goal. In this sense, a clause fragment such as *putze*, *wasche*, *spüle*, and *creme* must have two category components such as *A:m* and *B:m* representing the missing semantic elements, but the last two of them also need a category component representing the missing grammatical words *aus* and *ein* that complete the lexical word *spüle... aus* and *creme... ein*.

In such a grammar, the number of clause constituents is motivated both by the number of semantic elements in a semantic figure and the number of words in the expression representing the process. Because of this, the category for the verbs *spüle* and *creme* need to be different from the category of the verbs *putze* and *wasche* in Example 9.8.

As shown in Table 9.8, whenever a lexical verb has two grammatical words, the one that is not a grammatical verb receives the simple category **w[...]** where **w** stands for a grammatical word and **...** is a lexicogrammatical feature specifying a grammatical term (or lexeme) such as **aus** for *aus* and **ein** for *ein*. The grammatical verb receives a complex category with a component representing the missing lexical constituent such as **w[aus]** or **w[ein]**. Since this missing lexical constituent does not represent any new semantic element on its own, no

A:m	A:c\X:m/Y:m	C:m
A:Thing	B:Figure(X:Thing, Y:Thing)	C:Thing
ich	*putze*	*meine Zähne*
ich	*wasche*	*meine Hände*

A:m	A:c\X:m/P:w[...]/Y:m	C:m	D:w[...]
A:Thing	B:Figure(P:Process, X:Thing, Y:Thing)	C:Thing	D:Process
ich	*spüle*	*meinen Mund*	*aus*
ich	*creme*	*mein Gesicht*	*ein*

Example 9.8: Rank tags for lexical words of with different numbers of terms

semantic composition occurs when this clause constituent is combined with the clause fragment that precedes it. A derivation for a multiword lexical verb is shown in Figure 9.8.

```
ich  spuele                    meinen Mund    aus
---  ----------------------    -----------    --------
A:m  B:c\X:m/w[aus]/Y:m        C:m            w[aus]
A    B(P,X,Y)                  C
     ----------------------------------->
A:m  B:c\X:m/w[aus]                         w[aus]
A    B(P,X,C)
     ------------------------------------------------>
A:m  B:c\X:m
A    B(P,X,C)
<------------------------------------------------------
B:c
B(P,A,C)
```

Figure 9.1: Derivation for a two-term lexical word

Moving on, as described in Chapter 5, in addition to goal-affecting actions shown in Table 9.8, the represented process can also be an actor-part-affecting action as illustrated in Example 9.9.

A:m	B:c \X:m /Y:m /P:w[sich]	P:w[sich]	C:m
A:Thing	B:Figure(P:Process, X:Thing, Y:Thing)	P:Process	D:Thing
ich	*putze*	*mir*	*die Zähne*
ich	*wasche*	*mir*	*die Hände*

Example 9.9: Hypertags for lexical constituents of multiword lexical words

In Example 9.9, the wordings *putze mir* and *wasche mir* are treated as a multiword expression representing the action of doing something to one's own body parts. Two different terms occurred for each of the processes of brushing and washing one's body parts in our corpus: namely *die Zähne putzen*$_a$ and *sich die Zähne putzen*$_b$ for brushing one's teeth and *die Hände waschen*$_a$ and *sich die*

Hände waschen$_b$ for washing one's hands. The grammatical verb *wasche*$_a$ has the category **B:c\X:m/Y:m** whereas the grammatical verb *wasche*$_b$ has the category **B:c\X:m/Y:m/w[sich]** where **w[sich]** is a category component representing a lexical constituent. This constituent agrees with the actor subject in the same way as the grammatical verb does (see Example 113). For this reason, combinatory restrictions due to the agreement between the **w[sich]** component and the actor subject is enforced through grammatical features in the same way as it is standardly done in CCG grammars when enforcing combinatory restrictions between the subject and the grammatical verb.

(113) *ich wasche mir die Hände*
du wäschst dir die Hände
er wäscht sich die Hände
sie wäscht sich die Hände
I/you/he/she wash[es] my/your/him/her self the hands

Finally, in some clauses, the process is mentioned as if it were a simple thing by a noun. When this happens, we further classify the simple thing as a 'processual thing' instead of an 'entity'. Table 9.10 shows how those two types of figure representation differ in terms of combinatory categories.

A:m	**B:c \X:m /P:w[aus] /Y:m**	**C:m**	**D:w[aus]**
A:Thing	B:Figure(P:Process, X:Thing, Y:Thing)	C:Thing	D:Process
ich	*spüle*	*meinen Mund*	*aus*

A:m	**B:c\X:m/Y:m**	**C:m**
A:Thing	B:Figure(X:Thing, Y:Machbar)	C:Machbar
ich	*mache*	*eine Mundspülung*

Example 9.10: Incomplete clauses with different kinds of complements

Instead of a process represented by the bold wording in *ich* ***spüle*** *meinen Mund* ***aus*** (*I'll* ***wash*** *my mouth*), the processual thing is represented by the complement of a performance process underlined in *ich mache* ***eine Mundspülung*** (*I'll do* ***a mouthwash***).

To combine a processual thing with the corresponding performance process such as *machen* and *bekommen*, specific semantic classes are used. The category component *Y:Machbar* indicates that the missing constituent must be a mention of a simple thing of a particular kind: namely, a processual thing and, more specifically, a 'doable' processual thing (Machbar). A mouthwash (Mundspülung) is then classified as a processual thing that a person can 'do' (Machbar) and the parser can therefore restrict the combinations between processual things and performance processes in such a way that only those combinations that truly represent a process in German are allowed.

In short, grammatical complementation can be modelled in the following way: one clause constituent – typically a finite grammatical verb – is given a complex category that produces a clause. The derivation starts at this point and goes on to include all other constituents of the clause, including missing constituents of the lexical verb and the mention of a processual thing that is performed if any.

These combinatory categories are sufficient for parsing as long as the verbal group contains a single verb and all other clause constituents are complements. However, this condition does not always hold. Some clauses do have constituents which are not grammatical complements. In the next section, I shall model these other parts of the clause in terms of combinatory categories.

9.4 Inflections, auxiliaries and adjuncts

In a systemic description of grammar, some meaningful contrasts can only be made if wordings with different grammatical structures are compared and taken to be alternative options in a system. Since meaningful contrasts between wordings must be in some way 'meaningful' (not semantically random), they realise contrasting semantic structures. In this section, I shall go over such comparisons for the linguistic phenomena that were observed in this thesis.

9.4.1 Tense: inflection versus auxiliary

When comparing how a **non-past** relational process is represented in German with how a **past** one is, we notice that two different inflections of the lexical verb are used for relational processes as illustrated by Examples 114 and 115. Tense auxiliaries or words with inflectional feature for tense are written in bold and process words are underlined.

(114) *ich* ***bin*** *im Bett*
I ***am*** *in bed*
I ***will*** *be in bed*

(115) *ich* ***war*** *im Bett*
I ***was*** *in bed*

However, when comparing how a non-past material process is represented with how a past one is, we find something else. The non-past systemic feature is realised by an inflection of the lexical verb for the material process (Examples 116 and 117) whereas the past feature is realised by an auxiliary verb and an inflection of the lexical verb for the material process (Examples 118 and 119).

(116) *Roland* ***fährt*** *mich in die Küche*
Rolland ***is*** *bringing me into the kitchen*
Rolland ***will*** *bring me into the kitchen*

(117) *Roland* ***fährt*** *mit mir in die Küche*
Rolland ***is*** *bringing me into the kitchen*
Rolland ***will*** *bring me into the kitchen*

(118) *Roland* ***hat*** *mich in die Küche gefahren*
Rolland ***brought*** *me into the kitchen*

(119) *Roland* ***ist*** *mit mir in die Küche gefahren*
Rolland ***brought*** *me into the kitchen*

The past auxiliary is either the **haben past auxiliary** (*habe*, *hast*, *hat*, *haben*, *habt*) or the **sein past auxiliary** (*bin*, *bist*, *ist*, *sind*, *seid*). In Examples 118 and 119, the two lexical verbs together constitute a single verbal group where the auxiliary is the second constituent of the clause and the remainder of the verbal group is the last constituent.

Inside the verbal group, the bold verbs in Examples 116-119 function as Finite and the underlined verbs function as Process (Event or State). All non-process verbs function as Auxiliaries and the inflection of all non-finite verbs is an **infinitive**. The verb *fährt* in Examples 116 and 117 is in **present** inflection and the verb *gefahren* in Examples 118 and 119 is in **past-infinitive** inflection (pi).

9.4.2 Tense: inflection or auxiliary plus adjunct

In German, tense is not only represented by inflections and auxiliaries. It is also represented by adjuncts. Examples 120-123 illustrate how the adverbials *gerade* and *gleich* specify the tense further.

(120) *ich* ***bin*** gerade *im Bett*
I ***am*** *in bed*

(121) *ich* ***bin*** gleich *im Bett*
I ***will*** *be in bed*

(122) *Roland* ***fährt*** *mich* gerade *in die Küche*
Rolland ***is*** *bringing me into the kitchen*

(123) *Roland* ***fährt*** *mich* gleich *in die Küche*
Rolland ***will*** *bring me into the kitchen*

Words functioning as Adjuncts differ from those functioning as Auxiliaries in a fundamental way: if we remove the Adjunct, the remainder of the clause is also a clause and this clause has the same meaning except for the contribution brought by the adjunct. In this sense, by removing the adjunct, we produce a less specific clause. This property has a strong consequence for parsing. It means that we can parse the remainder of the clause in the same way we would do if there were no adjunct and model the adjunct as an additional clause constituent that appends a semantic element or a semantic feature to a complete semantic structure. This is not the case for Auxiliaries since Examples 124 and 125 are not grammatically complete despite the fact that they represent a complete figure:

(124) **Roland mich in die Küche gefahren*
???

(125) **Roland mit mir in die Küche gefahren*
???

As we can see in these examples, an auxiliary adds semantic features to the semantic structure in the same way as adjuncts do, but the remainder of the clause is not a clause on its own. This means that parsing clauses with auxiliaries

demands a different approach. We need to recognise grammatically incomplete clauses (**infinitive clauses**) such as the ones in Examples 124 and 125 and assign a complete figure to them. Then we need to recognise the auxiliary as a clause constituent that combines with an infinitive clause and that, if finite, completes the clause grammatically.

In particular, there are two types of adjuncts when it comes to semantic compositionality: namely 'specifiers' and 'subspecifiers'. In the case of tense, inflected verbs and auxiliaries function as Tense_1 because they specify the tense as either past or non-past. In turn, the adjuncts function as Tense_2 because they further specify the non-past tense as either present or future. This means a semantic figure has either one or two tense features in German: an obligatory Tense_1 realised by an inflection or an auxiliary and an optional Tense_2 realised by an adjunct.

	hat	gefahren	fährt	gerade
fährt	ist	gefahren	fährt	gleich
Finite	Finite	NonFinite	Finite	
Process	Auxiliary	Process	Process	Adjunct
Tense_1	Tense_1	–	Tense_1	Tense_2

Example 9.11: Tense in verbal group

In German, there are therefore four primary tenses: past, non-past, present and future. Present and future tenses are realised through semantic composition. In Figure 9.2 I present a combinatory category for auxiliaries that combine with past-infinitive clauses (pi) resulting in a finite clause.

```
Rolland  hat                   mich  ins Bad  gefahren
-------  --------------------  ----  -------  ------------------------
A:m      B:c\Z:m/(B:c[pi]\Z:m)  C:m   D:m      E:c[pi]\X:m\Y:m\Z:m
A        B+past                C     D        E(P,X,Y,Z)
                                      ----------------------------------<
A:m      B:c\Z:m/(B:c[pi]\Z:m)  C     E:c[pi]\X:m\Y:m
A        B+past                C     E(P,X,Y,D)
                               ---------------------------------------<
A:m      B:c\Z:m/(B:c[pi]\Z:m)  E:c[pi]\X:m
A        B+past                E(P,X,C,D)
         ------------------------------------------------------------->
A:m      B:c\X:m
A        B(P,X,C,D)+past
-----------------------------------------------------------------------<
D:c
E(P,A,C,D)+past
```

Figure 9.2: Derivation for a tense auxiliary

In Figure 9.3 I present a combinatory category for tense adjuncts in German.

```
Rolland  f\"{a}hrt                    gerade          ins Bad
-------  --------------------------  --------------  -------
A:m      B:c\X:m/Y:m                 C:c\C:c         D:m
A        B(P,X,Y)+nonpast            C+present       D
         ------------------------------------------<
A:m      C:c\X:m/Y:m                                 D:m
A        C(P,X,Y)+nonpast+present                    D
         ------------------------------------------------>
A:m      C:c\X:m
A        C(P,X,D)+nonpast+present
------------------------------------------------------------>
C:c
C(P,A,D)+nonpast+present
```

Figure 9.3: Derivation for a tense adjunct

9.4.3 Conation: nothing versus auxiliary

In addition to tense, a clause can also represent an action or service as an attempt to achieve something which can be further specified as successfully or unsuccessfully performed (conation). In German, conation is optional. If it is represented, it is realised by an auxiliary as in Examples 126, 127, 129 or realised by an adjunct as in Example 128.

(126) *Roland **hat** versucht in die Küche zu fahren*
Rolland tried to go to the kitchen

(127) *Roland **konnte** in die Küche fahren*
Rolland was able to go to the kitchen

(128) *Roland **ist** immerhin in die Küche gefahren*
Rolland was able to go to the kitchen

(129) *Roland **konnte nicht** in die Küche fahren*
Rolland was not able to go to the kitchen

Differently from tense, conation auxiliaries can have a past inflection and be finite such as *konnte* in Examples 127 and 129 or have a past-infinitive inflection such as *versucht* in Example 126. The verb *zu fahren* of Example 126 is in trial-infinitive inflection (ti) and the verb *fahren* in Example 127 is in success-infinitve inflection (si). The failure auxiliary *konnte... nicht* is a two-word lexical verb and its combinatory category (hypertag) reflects this fact in the same way as multiword lexical verbs in the previous sections did. Moreover, conation can also be realised by adjuncts as in Example 128. Differently from Tense_2 adjuncts, though, Conation adjuncts do not attach themselves to the clause fragment that contains the Tense_1 auxiliary but to clause fragment that contains the Process verb. This difference shall be discussed in the Section 9.6.

Conation is a focus on an actor's behaviour as an attempt to cause a change in the world, which can be a success or a failure. Differently from tense, representing an action or a service as a behaviour with unknown outcome is a non-obligatory option provided by the German grammar. The absence of a

```
hat      :- K:c\Z:m/(K:c[pi]\Z:m)           K+past
versucht :- K:c[pi]\Z:m/(K:c[ti]\Z:m)       K+trial
konnte   :- K:c\Z:m/(K:c[si]\Z:m)           K+past+success
konnte   :- K:c\Z:m/(K:c[si]\Z:m)/w[nicht]  K+past+failure
nicht    :- w[nicht]
immerhin :- K:c[pi]/K:c[pi]                 K+success
```

Figure 9.4: Combinatory categories for conation auxiliaries and adjuncts

conation auxiliary means that the action/service is represented as an occurring change, not as a behaviour of an actor aiming at causing the change. This means that the meaning of a clause always has a tense feature in German, but that it may or may not have a conation feature opposing an occurring feature. All processes represented in German take place unless otherwise specified. Table 9.12 shows the internal structure of verbal groups with conation auxiliaries.

konnte	*fahren*
konnte nicht	*fahren*
Finite	NonFinite
Auxiliary	Process
Tense_1	–
Conation	–

hat	*versucht*	*zu fahren*
Finite	NonFinite	NonFinite
Auxiliary	Auxiliary	Process
Tense_1	–	–
–	Conation	–

Example 9.12: Tense and conation in verbal group

Finally, given the nature of conation, only trial auxiliaries were used in commands. Success and failure auxiliaries are used mainly when people report what happened in the past.

9.4.4 Phase: nothing, adjunct, versus substitution

Phase auxiliaries bind the tense to boundaries of developmental stages. Tense can be attached to the beginning of a displacement as in Example 130 or to the end as in Examples 131 and 132. It can be attached to the beginning of a pause in displacement as in Example 133 or to the end of the pause as in Example 134.

(130) *Roland **ist** in die Küche los gefahren*
Rolland started going to the kitchen

(131) *Roland **ist** in die Küche fertig gefahren*
Rolland finished going to the kitchen

(132) *Roland **ist** in die Küche gekommen*
Rolland finished going to the kitchen

(133) *Roland **hat** gestoppt*
Rolland stopped going to the kitchen

(134) *Roland **ist** in die Küche weiter gefahren*
Rolland resumed going to the kitchen

The way whereby different phases are realised is quite diverse. Sometimes a different lexical verb is chosen such as *gekommen* in Example 132, which represents the end phase of a displacement. Other times phase is realised by adjuncts such as *los*, *fertig*, and *weiter* respectively in Examples 130, 131 and 134 or by substitution as in Example 133. Phase adjuncts are very similar to the conation adjunct *immerhin* as far as combinatory categories are concerned. Substitution, on the other hand, represents a saturated figure where the elements in the figure are dynamically specified depending on the displacement that is taking place.

```
los      :- K:c[pi]/K:c[pi]  K+start
fertig   :- K:c[pi]/K:c[pi]  K+finish
gestoppt :- K:c[pi]\X:m      K(P,X,Y)+stop
weiter   :- K:c[pi]/K:c[pi]  K+resume
```

Figure 9.5: Combinatory categories for conation auxiliaries and adjuncts

Assuming that the lexical verb *stoppen* (*stop*) substitutes the predicate *in die Küche fahren* (*go to the kitchen*), the variables P and Y are respectively *fahren* and *in die Küche* in the lambda expression *K(P,X,Y)* of Figure 9.5.

9.4.5 Contribution: structure and inflection/auxiliary

A discourse contribution within a move is realised by multiple grammatical features. Statements differ from questions in German by constituent order. Examples 135-138 illustrate this opposition:

Statement

(135) *Roland fährt mich in die Küche*
Rolland is bringing me to the kitchen
Rolland will bring me to the kitchen

(136) *Roland hat mich in die Küche gefahren*
Rolland brought me to the kitchen

Polar Question

(137) *fährt Roland mich in die Küche?*
is Rolland bringing me to the kitchen?
will Rolland bring me to the kitchen?

(138) *hat Roland dich in die Küche gefahren?*
did Rolland bring you to the kitchen?

Statements are realised by a **declarative** mood structure, which consists of a constituent order where the Finite is the second constituent. In turn, questions

are realised by an **interrogative** mood structure, which consists of a constituent order where the Finite is the first constituent if there is no unknown element such as *wohin?* (*where?*). Verb inflections for statements and questions are the same in German. For that reason, we can talk about an **indicative** inflection of the verb.

In particular, a speaker usually does not give information about the addressees to the addressees nor do they demand information about themselves. This means that the subjects of indicative clauses can be anyone but a few people taking part in the interaction. Unlike exchanges of information such as statements and questions, exchanges of services tend to have a more restricted structure. The represented process is usually a service and the subject is usually the speaker or the addressee. Examples 139-146 illustrate such structures.

Offer

(139) *soll ich dich in die Küche fahren?*
should I take you to the kitchen?

(140) *möchtest du in die Küche fahren?*
would you like to go me to the kitchen?
do you want to go me to the kitchen?

(141) *musst du in die Küche fahren?*
do you have to go to the kitchen?
do you need to go to the kitchen?

Command

(142) *fahre mich in die Küche!*
bring me to the kitchen!

(143) *fährst du mich in die Küche?*
would you bring you to the kitchen?

(144) *kannst du mich in die Küche fahren?*
can you bring you to the kitchen?

(145) *ich möchte in die Küche fahren!*
I'd like to go me to the kitchen!
I want to go me to the kitchen!

(146) *ich muss in die Küche fahren!*
I have to go to the kitchen!
I need to go to the kitchen!
I must go to the kitchen!

When the client is affected by the service as in the examples above, there are two ways of representing a service: the subject can be the service provider as in Examples 139 and 142-144 (actual services) or the service client as in Examples

140-141 and 145-146 (actions per locution). Offers have a single mood structure where the finite is the first constituent whereas commands can be divided in two groups as far as mood structure is concerned: service commands have a mood structure in which the Finite is the first constituent (Examples 142-144) whereas commands with action-per-locution have a mood structure in which the Finite is the second constituent as if they were a statement (Examples 145-146). In the case of service commands, there is a mood structure in which the subject is implicit (Example 142) and another in which the subject is explicit (Examples 143-144).

Differently from statements and questions, offers and commands are not realised exclusively by mood structure and mode inflections. They are also realised by **mode auxiliaries**. The **oblative** auxiliary *soll* (Example 139) and the **directive** inflection in *fahre* (Example 142) realise respectively offers and commands together with the corresponding oblative and directive word orders. The **oblative** auxiliaries *möchtest* and *musst* (Examples 140-141), the **directive** inflection in *fährst* (Example 143), and the **directive** auxiliaries *kannst*, *möchte*, and *muss* (Examples 144-146) also realise contributions together with mood structure, however, they simultaneously realise a degree of obligation that the provider has to perform the service (modality). In this sense, this second group functions both as Mode and as Modal. Table 9.13 shows three alternative directive structures and Table 9.14 shows the directive inflections/auxiliaries that go together with the mood structures.

	fahre	mich	in die Küche	
	Mood	Remainder		
fährst	du	mich	in die Küche	
kannst	du	mich	in die Küche	fahren
Mood		Remainder		

ich	möchte	in die Küche	fahren
ich	muss	in die Küche	fahren
Mood		Remainder	

Example 9.13: Three directive structures

		kannst	fahren
		möchte	fahren
fahre	fährst	muss	fahren
Finite	Finite	Finite	NonFinite
Process	Process	Auxiliary	Process
Mode	Mode	Mode	–
–	Modal	Modal	–

Example 9.14: Directive inflections and auxiliaries in verbal group

It is therefore a combination of Mood structure at clause rank and Mode inflection/auxiliary at group rank that realises an illocutionary force. For this reason, we need the combinatory categories illustrated in Figure 9.6. The adverbials *einfach*, *doch*, *mal*, and *bitte* function as Mode_2, Mode_3, Mode_4, Mode_5

and so on (traditionally called "Comment") because they further specify the illocutionary force. They only combine with Mode constituents that are semantically compatible. For this reason, a grammatical feature "directive" (d) is added to directive clauses and corresponding grammatical features are added to declarative, interrogative and oblative clauses as well so as to enforce only meaningful combinations between Mode adjuncts with Mode auxiliaries and constituents with inflectional feature for mode.

```
fahre    :- K:c[d]/Y:m/X:m            K(P,X,Y)+directive
faehrst  :- K:c[d]/Y:m/X:m            K(P,X,Y)+directive+mid
fahren   :- K:c[si]\X:m\Y:m           K(P,X,Y)
kannst   :- K:c[d]/(K:c[si])          K(P,X,Y)+directive+low
moechte  :- K:c[d]\X:m/(K:c[si]\X:m)  K(P,X,Y)+directive+low
muss     :- K:c[d]\X:m/(K:c[si]\X:m)  K(P,X,Y)+directive+high
einfach  :- K:c[d]\K:c[d]             K+insistance
doch     :- K:c[d]\K:c[d]             K+change-of-mind
mal      :- K:c[d]\K:c[d]             K+favour
bitte    :- K:c[d]\K:c[d]             K+standard
```

Figure 9.6: Combinatory categories for mode inflection and auxiliaries

In short, discourse contributions are therefore realised by a range of combinatory categories. One or more clause constituents realise mode, that is, they either have an inflectional feature for mode or are a mode auxiliary or adjunct that specify or further specify the illocutionary force.

9.4.6 Voice: inflection versus auxiliary

In all examples up to this point, the complex category producing a clause was always assigned to the grammatical verb in the expression representing the process. This could be done because all represented processes were **operative**. In German, operative material processes are those whose operator is the subject of the clause. Actions such as washing one's hands are operated by the actor, services such as taking someone to the kitchen are operated by the service provider, and action by locution such as going to the kitchen are operated by the service client. However, the subject about which we negotiate information is not always the operator of the material process. It may also be a non-operating affected matter such as a service client, a service goal, or an action goal. In this case, the clause is said to be passive (see Examples 147-148).

(147) *ich **wurde** in die Küche gefahren*
I was brought by Rolland to the kitchen

(148) *ich **bin** in die Küche gefahren worden*
I was brought by Rolland to the kitchen

(149) *ich **wurde** in die Küche gebracht*
I was brought by Rolland to the kitchen

(150) *ich **bin** in die Küche gebracht worden*
I was brought by Rolland to the kitchen

In the examples above, the bold constituent is a finite verb and the underlined constituent functions as the Voice. It is the Voice constituent and not the Process constituent that receives a clause-producing category. The process verb in the examples above is in past-infinitive inflection and it has a rank tag of word. The processes of taking someone somewhere (*Prozess, jemanden wohin zu fahren* and *Prozess, jemanden wohin zu fahren*) are instances of a generic service type 'client-affecting displacement' (Cad) which encompasses all processes that fill a semantic figure containing slots for a service provider (the operator), an affected service client, and a route. Using this semantic class as filter, a Voice auxiliary can combine exclusively with the right type of process.

```
gefahren :- K:w[Cad]                    K
gebracht :- K:w[Cad]                    K
wurde    :- K:c\Y:m/P:w[Cad]/Z:m        Z.{P.{X.{K(P,X,Y,Z)+past+dec.}}}
worden   :- K:c[pi]\Y:m\Z:m\P:w[Cad]    P.{Z.{X.{K(P,X,Y,Z)}}}
ist      :- K:c\X:m/(K:c[pi]\X:m)       X.{K+past+declarative}
```

Figure 9.7: Combinatory categories for voice auxiliaries

With this final addition, we reach the end of the description of the verbal group as far as the linguistic evidence in transcriptions of dialogues and retrospective protocols is concerned. In the next two section, I shall describe in more detail the structural types of the process class names (lexical verbs) and the ways in which conjunctions move elements around in the clause.

9.5 Processes

In the previous section, different clause-producing categories were assigned to the voice constituent depending on the incompleteness 1) of the figure and 2) of the lexical expression representing the process. Lexical expressions are conceived of in this thesis as being composed of **grammatical words**.

Because of the clause incompleteness depends on both figurative and lexical incompleteness, for a voice constituent to be classifiable for clause incompleteness, a process term should be classified both in term of figurative incompleteness and lexical incompleteness.

Table 9.15 shows a list of lexical expressions divided into grammatical words. The first column contains the grammatical words that belong to the inflectional word class of "verbs", the second column contains the grammatical words that belong to the word class of "reflexives", and the last column contains the grammatical words that belong to the word class of "particles". As one can see, this table is sparse since not all lexical expressions contains grammatical words of all kinds. For this reason, a first classification of process terms consists of knowing whether they are "reflexive" and/or "separable" or not. These two lexical features combined with a figure type are what is needed for deciding how many constituents are missing in a clause-producing category for the grammatical verb ($\text{Process}_{\text{Verb}}$).

$\text{Process}_{\text{Verb}}$	$\text{Process}_{\text{Reflexive}}$	$\text{Process}_{\text{Particle}}$
fahr	–	–
fahr	–	–
komm	–	her
fahr	–	–
bring	–	–
wasch	–	–
wasch	dir	–
hol	–	–
hol	dir	–
dreh	dich	–
lade	dich	auf

Example 9.15: Predication structures verbs, reflexes and particles in imperative clauses

9.6 Conjunction

As shown in previous sections, experiential Adjuncts such as Phase, Conation, and Circumstance attach themselves to the Process end of a verbal group whereas interpersonal Adjuncts such as Tense_2 and Mode_2 Adjuncts attach themselves to the Finite end. In particular, I have argued that Tense_2 adjuncts such as *gerade* and *gleich* and Mode_2 adjuncts such as *bitte* further specify Tense_1 and Mode_1 inflections and auxiliaries, thus making combinations between the two restricted. Examples 151-154 illustrate this dependence.

(151) ***fahr*** weiter *in die Küche*
Resume *going to the kitchen*

(152) *ich* ***bin*** weiter *in die Küche gefahren*
I ***resumed*** *going to the kitchen*

(153) ***fahr*** bitte *in die Küche*
please ***go*** *to the kitchen*

(154) **ich* ***bin*** bitte *in die Küche gefahren*
???

This difference becomes yet more evident when we model dependent clauses as in the Examples 155-158

(155) *sodass ich* weiter *in die Küche fahren* ***kann***
so that I ***can*** *resume going to the kitchen*

(156) *um* weiter *in die Küche zu fahren*
to resume going to the kitchen

(157) **sodass ich* bitte *in die Küche fahren* ***kann***
???

(158) **um* bitte *in die Küche zu fahren*
???

Example 156 shows that a Phase Adjunct can append itself to the Process independent of whether a clause is finitive or infinitive. A Mode Adjunct on the other hand cannot. Examples 157 and 158 shows that *bitte* (*please*) can only attach itself to a Mode_1 constituent, which does not exist in dependent clauses.

Similar to illocutionary force, a hypotactic nexus can be realised by a range of grammatical features including a **conjunctive structure**, a **Conjunctive** operator and a **Conjunctive** auxiliary. In German, dependent clauses shows a conjunctive operator such as *sodass* or *um* as the first constituent and, sometimes, an auxiliary such as *kann* as the last constituent. Since dependent clauses are no statements, no questions, no commands, and no offers on their own, they do not have any constituent that function as Mode. This means that hypotactic conjunctions are not adjuncts since the removal of conjunctive constituents does not produce a clause with a less specific meaning. It produces a clause that is grammatically incomplete or that has a different meaning.

Nonetheless, conjunctive auxiliaries can function as Tense_1. For this reason, dependent clause can have a non-past tense as in Example 155 or a past tense as in Example 159.

(159) *sodass ich* weiter *in die Küche fahren* ***konnte***
so that I ***could*** *resume going to the kitchen*

Dependent clauses that have a primary Tense_1 are said to be finite (or "tensed") and dependent clauses that have no primary Tense_1 as in Example 156 are said to be non-finite (or "non-tensed"). The finite constituent is nonetheless 'conjunctive' in the sense that it does not create a complete clause as a way of contributing to discourse ("mode").

Moreover, dependent clauses also vary in the way in which semantic elements are represented. Dependent clauses such as Example 156 depend on other clauses which have semantic elements of their own. They can be chosen when the subject element of the dominant clause is identical with the implicit subject element of the dependent clause. For this reason, Examples 160 and 161 are expected commands to a wheelchair whereas Example 162 is not.

(160) *ich* ***möchte*** *zum Arbeitstisch fahren, um ein Buch zu holen*
I ***'d like to*** *go to the desk, to pick up a book*

(161) ***fahren*** *wir zum Arbeitstisch, um ein Buch zu holen*
let's *go to the desk, to pick up a book*

(162) ****fahr*** *mich zum Arbeitstisch, um ein Buch zu holen*
****take*** *me to the desk, to pick up a book*

When conjunctive operators and auxiliaries in terms of combinatory categories, we should observe that the resulting clauses are 'conjunctive'. In particular, the auxiliaries *kann* and *konnte* are instances of the **können result auxiliary**: *kann*, *kannst*, *kann*, *können*, *könnt* in non-past inflection and *konnte*, *konntest*, *konnte*, *konnten*, *konntet* in past inflection; and clauses with them

```
um       :- K:c\L:c/(M:c[si]\X:m)     K(L,M)
sodass   :- K:c\L:c/M:c[rc]           K(L,M)
kann     :- K:c[rc]\K:c[si]           K+non-past
holen    :- K:c[si]\X:m\Y:m           K(P,X,Y)
```

Figure 9.8: Combinatory categories for conjunctive auxiliaries

receive the grammatical feature **result-conjunctive** (rc). This allows the following combinatory categories to build up the intended grammatical structures:

As a final remark on the development of CCGs, disallowed cross compositions can be avoided by replacing the permissive slash modifier $\times$ (permutative) by the less permissive slash modifiers such as $\times<$ (permutative left) and $\times>$ (permutative right) (Bozşahin et al., 2005). For the purpose of wheelchair automation, such a fine control of combinatorial possibilities is not necessary.

9.7 Conclusion

In this chapter, I described grammatical complements in terms of semantic composition. In particular, I showed how to assign combinatory categories to both grammatical complements that represent semantic elements and grammatical complements that are words in lexical expressions. I also showed how to handle Voice, Conation, and Mode/Modal Auxiliaries as well as Phase and Mode/Modal Adjuncts in terms of combinatory categories. While doing so, I pointed out various benefits of modelling a lexicogrammar with a rank scale in CCG, which include smaller grammars, fewer wrong compositions, and semantic structures that are better for further processing.

In the next four chapters, I shall move on to describing the process of integrating symbols with symbolised observable phenomena in the situation and with the ongoing dialogue as dialogue moves. Viewing an utterance from a listener's perspective, I shall take one step back from the description of composite symbols to the description of situations and I shall describe how uttered symbols can be understood as an attempt to achieve something interpersonal in the situation.

Chapter 10

Reference Integrator

When people describe entities and locations around them to the wheelchair, the wheelchair needs to identify these entities and their locations in the current situation. In this chapter, I describe the component that determines the entity and the entity's current and potential locations described by a speaker. This component tackles the linguistic phenomena described in Section 1.1.2 from both the speaker's and listener's perspectives. For this, it utilises the semantic classes in Chapter 5.

10.1 Entities

Entities in the situation such as the speaker, the addressee, and classified entities such as *the wheelchair*, *the kitchen*, and *the bed* are added to an inventory of present entities. Each entity is a single item in the inventory even if it can be represented in multiple ways. For instance, there is a single item in the inventory for the sofa and the couch because *the sofa* and *the couch* are the same entity. The reference integrator creates a hierarchy of entity sets as a node tree where each node is an entity set and every child node is a subset of its parent node. In turn, the edges between these nodes are linguistic classes of entities, not crosslinguistic classes of entities: for instance, *the sofa* and *the couch* are two different linguistic classes of entities and they correspond to two different edges.

The reference integrator creates a tree adding two edges to the top node for *present* from *absent* entities. Present entities are the entities in the situation whose relative position to the interactants is known by them. In our case, they are limited to the entities in the apartment (see Chapter 5). For each linguistic class of non-persons, a node is added as a child of present entities, each edge corresponding to a linguistic class of entities such as *the kitchen* and *the bed*. Each node in the decision tree contains all instances of the given linguistic classes found in the inventory.

In turn, if a semantic class such as *the table* has subclasses such as *the dining table* (*der Esstish*) and *the 'working' table* (*der Arbeitstish*), a new node is created under that set for each the subclass qualities such as *dining* (*Ess-*) and *'working'* (*Arbeits-*) and all entities that instantiate each subclass is added to the child set.

In a separate branch of present entities, the reference integrator adds present

entities that can be represented as possessed by other entities under the edge *possessed*. A different edge is added to this entity set for every entity possessor and the entities under this edge are the entities possessed by the possessor. Finally, a third branch of the decision tree under present entities is created for entities represented as qualified by a spatial location under the edge *located*. A branch is created for each potential spatial relatum and a subbranch is created for each potential spatial relation to this relatum. The leaf nodes contain all entities in the respective relative location.

The process of determining the entity represented by *my dining table in the kitchen* in the inventory consists of three steps. The reference integrator searches for all present dining tables, which are a subset of present tables, which are a subset of present entities in the decision tree. Then the reference integrator searches for all present entities possessed by the speaker, which are a subset of present possessed entities, which are a subset of present entities in the decision tree. Finally, the reference integrator searches for all present entities in the kitchen, which are a subset of all present entities located relatively to the kitchen, which are a subset of all present entities relatively located, which are a subset of all present entities in the decision tree. These three sets of entities are intersected and the remaining entity - if a single one - is the represented one. A similar process happens to possessors such as the underlined words in *my mouth* (*meinen Mund*) and *to the kitchen table* (*zum Küchentisch*). In both cases, the set present entities possessed by the represented possessors is intersected with the set of present classified entities.

Mentioned entities are temporarily tagged 'recently mentioned' for anaphoric reference and are added to a separate subset of entities next to absent and present entities. This set is subdivided into three subsets: one for *Der* (masculine *it*), *Die* (feminine *it*), and *Das* (neutral *it*). If an entity was mentioned as *das Sofa*, it is added to the subset *Das*. If it was mentioned as *die Couch*, it is added to the subset *Die*. If an entity was represented by a composite symbol such as *der Esstisch* (*the dining table*) or *der Küchentisch* (*the kitchen table*), an extra edge is added under the *Der-Die-Das* sets for the entity class of the head symbol. In this case, the edge *der Tisch* (*the table*) is added under the set *Der* (masculine *it* set), leading down to the set of mentioned tables. With this branch, the reference integrator can determine the recently mentioned table represented by *ihn* (masculine *it*) or *den Tisch* (*the table*) in a given discourse state even though there are more than one entity in the situation that can be represented by these wordings.

Useful entities for the current activity are also temporarily tagged 'useful'. They are added to a subset of entities, namely the set of currently useful entities. This set of currently useful entities can be intersected with other sets whenever multiple entities could be the mentioned one. For instance, when someone knocks on the apartment door, the user starts the activity of opening the door for the new guest. At that moment, the apartment door becomes more useful for this activity than other doors in the situation and it is added to the set of useful entities. When the user tells the wheelchair to take him or her to 'the door', the reference integrator finds multiple present doors and intersects this set with the set of useful entities to determine the door represented by the user.

Persons are added to the list of present things in a different way. Someone's name such as *Rolland*, *Mary*, and *John* and someone's category such as *Wheelchair*, *Ma'am*, and *Sir* can be used to call people that can hear the

conversation. Therefore, their names and categories cannot be easily used for representing them in dialogue without accidentally inviting them to join the interaction. As a result, different semantic classes are available for persons when they are present and absent; if they are present, different semantic classes are available when they are distant and attendant (reactive to calls); if attendant, when they interactants and nearbystanders; if they are interactants, if they are speaking or listening (speaker, listener), and if they are exchanging information and services (sender, addressee) or just standing by (bystander). The systemic network below shows the dependencies of options.

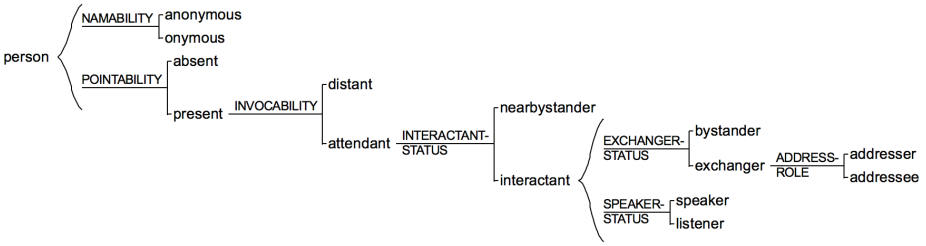

Figure 10.1: Semantic systemic network for exophoric reference to persons

Following this network, the user is added to the unary set of speakers every time he or she is speaking and removed from this set every time he or she stops speaking. Since the speaker is speaking in his or her own behalf, the speaker is also the sender of the message. There are two other people in the room: the wheelchair and the researcher. Both are added to the set of listeners, but which one is the addressee and which one is a bystander is decided based on the contents of the message. If the addressee is specified as in *Rollstuhl, bring mich zum Tisch* (*wheelchair, take me to the table*), the wheelchair is added to the unary set of addressees and the researcher to the unary set of bystanders. If no one is specified, the wheelchair assumes it is the addressee temporarily until it concludes it cannot be the addressee. Since the wheelchair does not need to understand utterances directed to the researcher in the experiment, this strategy was sufficient.

With these categories, the wheelchair is capable of resolving references to the speaker as in *ich*, *mich* and so on, to the addressee as in *du*, *dich* and so on, and to answer to questions such as *Hörst du mich?* (*Do you hear me?*).

10.2 Parts of Entities

In German, someone's mouth can be represented by wordings such as *den Mund* (*the mouth*) in *ich möchte mir den Mund ausspülen* (*I want to wash 'myself' 'the mouth'*) even when there are two or more humans in the situation. As a result, if all present mouths were considered potential referents, the reference integrator would not be able to determine which mouth was mentioned.

In Chapter 5, different figures were proposed for the goal-affecting and actor-part-affecting actions. Goals were defined as entities to be identified in the situation. Actor parts were defined as entities to be identified as parts of the actor. Since the linguistic analyser provides these semantic roles, the reference integrator can identify the goal *meinen Mund* (*my mouth*) and the actor part

den Mund (*the mouth*) in different ways. A goal is identified as described in the previous section. The actor part is identified in two steps. First the corresponding actor is identified. In this case, the actor is the user who spoke this command. Then the set of all parts of this actor is intersected with the set of present mouths. The user's mouth is the only member in the resulting entity set.

10.3 Locations

The set of entities located relatively to non-moving objects is quite stable. For instance, the set of entities in a room changes only when one of the entities moves into or out of the room. Since the wheelchair tracks moving entities, every time the set of entities in a room changes, the motion tracker updates this set. However, the set of entities located relatively to moving objects is very unstable. The entities that are next to the user or to the wheelchair change as they move around in the apartment. For this reason, the tracker of moving entities needs to update the entities located relatively to the moving entities continuously.

Without these updates, wordings such as *the wash basin in the bathroom* (*das Waschbecken im Badezimmer*) would be correctly understood, but wordings such as *the book on the desk* (*das Buch auf dem Arbeitstisch*) and *come to me* (*komm zu mir*) would not.

Moreover, whenever two potential semantic structures are offered by the lexicogrammatical analyser, the reference integrator might be able discard one of the two. For instance, if there is a book on the table and no book on the sofa, the command *I want to read the book on the sofa* (*ich möchte das Buch auf dem sofa lesen*) contains a wording for *the book* and a wording for the place where the reading should take place, namely *on the sofa*. The candidate symbol for the user reading the book that is on the sofa is discarded by the reference integrator because there is no book there.

10.4 Potential Locations

Potential locations of entities are relevant for specifying the destination of a movement. For instance, *on the sofa* is a potential location for a human, not for a wheelchair, but there are potential locations for wheelchairs and humans *in the kitchen*. So if the user tells he or she wants to read a book *in the kitchen*, the wheelchair can conclude the user will stay in the wheelchair whereas, if the user tells the wheelchair he or she wants to read the book on *the sofa*, the wheelchair will need to stop next to the sofa in a position where the user can move onto the sofa by him or herself.

In addition, whenever two potential semantic structures are offered by the lexicogrammatical analyser, the reference integrator might be able discard one of the two. For instance, if there is a book on the table, the command *I want to pick up the book on the table* contains a wording for *the book on the table*. The candidate symbol for the user picking up a book, the full action taking place on the table, can be discarded by the wheelchair because *on the table* is not a potential location for a wheelchair user when picking up a book. One

can say that a person might potentially pick up a book on the shelves while standing on a table and they would be right. Potential locations are drastically different depending on the situation, the special abilities of interactants, their impairments, the activity they are involved in, and so on. For this exact reason, potential locations for each interactant and movable entity had to be predicted for the situation and the activities experiment participants had to carry out.

Within the activities, all potential destinations of interactants and moveable objects were listed. These potential locations were then taken as potential instances of the relative and absolute locations represented in commands. In turn, these were organised in a hierarchy of potential location sets for each class of entity. The final task of the reference integrator is to determine whether there is any potential position applicable to a particular entity for the represented location. Based on this, candidate semantic structures for a command can be discarded.

10.5 Conclusion

In this chapter, I presented how the reference integrator works, how it identifies semantic entities with phenomena and overcomes complexities typical of this task. All linguistic phenomena related to reference listed in the introduction were covered by the tree. The referential phenomena that could not be properly treated with the current approach are listed in Chapter 14.

Chapter 11

Configuration Integrator

In commands to wheelchairs, people describe events in which simple things are washed, opened, eaten, moved around, and the like. The wheelchair needs to recognise who is in charge of performing each of these material processes, that is, who should do the 'hard work' or 'labour', so that it can plan its part in the interaction. In this chapter, I describe the component that *consumes* potentially meant relational and material processes and *produces* a list of the most likely meant relational and material processes. In particular, this component tackles linguistic phenomena related to configurations of elements (Section 1.1.3) by checking whether represented actors can perform the material actions they are represented doing and whether the represented matter affords the material actions under representation. In addition, material actions are further divided as either actions per labour, services per labour, or actions per locution so that specified rights and duties can be used to verify whether the person represented as a service client can demand the represented service from the person represented as a service provider and whether those represented as service providers have to perform the service for their supposed clients.

11.1 Material processes as events

Material processes are defined as processes whereby a portion of matter changes one or more of its attributes. Since only entities[1] were taken into account, all material processes in this study correspond to changes in entity attributes.

In this thesis, figures are classified in two ways. First they are classified regarding the roles they have, then they are subclassified for the entity attribute being changed. All figure types for changes in spatial location are subtypes of ***ChangeInLocation***, a figure in which a medium changes its location. In this way, all material processes, whether they are represented as happenings, actions per labour, services per labour, and actions per locution, can be understood as an **event**, a change in attributes for a medium ignoring all external agents, if any. Event types are used by trackers of material changes in the environment such as the tracker of moving objects. Even if the tracker cannot determine who made the change happen or what caused the change, it can determine whether the change happened or not. This map between figures is relevant in

[1] Bounded portions of matter (see Chapter 5)

human-wheelchair interaction when the wheelchair needs to determine whether a described service or action was completely executed or not.

11.2 Checks

Since multiple candidate semantic structures for a wording might be produced by the lexicogrammatical analyser, one of the tasks of the configuration integrator is to determine whether the configuration is to be considered or to be discarded. In the following, I present all checks implemented.

11.2.1 Affordance check

An affordance check consists of determining whether a medium affords the changes it is described undergoing. For instance, if a medium is described as being moved to the kitchen, the first check would be to check if the medium can move. For instance, in the utterance *I want to pick up the book*, the speaker represents the book as something that can be picked up. In this case, the configuration integrator checks whether the book is a potential medium of the action of picking up something.

For each process, such as the process of picking up something, a class of entities is created. For the action of picking up something (*SichEtwasAbholen*), a class of entities is created for potential mediums of picking up something (*PotMediumOfSichEtwasAbholen*). This class is defined in one of two ways: either all books are defined as entities that can be picked up or a subclass of books is created for the books that can be picked up. In the second case, the instances that can be picked up are added to the subclass of books that can be picked up.

A reasoner is used for checking whether the represented entity afford the specified action or service by checking whether the represented entity is a potential medium for that process. If it is not, the candidate semantic structure is discarded.

11.2.2 Capacity check

When it comes to 'hard work' (labour), the configuration integrator needs to check whether the agents that are supposed to perform labour indeed can perform it.

For each action per labour and service per labour, the configuration integrator needs to check whether the entity supposed to do labour is a potential labourer for the action and/or service. For instance, if the user says *I want to go to the kitchen*, the lexicogrammatical analyser produces two candidate semantic structures: one in which the speaker is an actor of an action per labour and the other in which the speaker is the client of an action per locution. The speaker is a person who cannot go to the kitchen on his or her own, therefore the speaker is not a potential actor of this action per labour. This candidate understanding needs to be discarded.

To accomplish this, for each process involving labour such as the action per labour of going somewhere ($WohinFahren_1$), a class of entities is created automatically for potential actors of going somewhere ($PotDoerOfWohinFahren_1$).

Either all humans are defined as potential doer of going somewhere or the subclass of humans that are not gait-impaired. In this case, the latter option was taken. A reasoner is used to determine that humans that are gait-impaired such as the wheelchair user are not potential actors of such an action. Therefore, the candidate semantic structure for which this would need to be the case is discarded. Only the configuration where the speaker does an action per locution is preserved.

If the material process has a medium other then the agent performing labour, a further specification can be made. Whereas gait-impaired humans can wash their hands, their mouth, and other parts of their bodies on their own, they might not be able to wash the dishes, clean the toilet, the bath, the floor, and so on. Considering semantic composition alone, all those entities are washable and could be the goal of washing. To discard such understandings, the configuration integrator checks whether an agent can be a potential doer for a given type of labour and a given type of medium. In the current case, the user is a potential doer of mouth washing (*PotDoerOfSichDenMundAusSpülen*) and a potential doer of hand washing (*PotDoerOfSichDieHändeWashen*). For each labour that the interactants can perform only for certain types of entities, one such class of potential doers is created and defined accordingly.

If interactants in a situation can only perform the labour to a subset of the entities of a given type, another approach is required. For instance, a wheelchair user might be able to carry some boxes around, for instance, the light ones, but not all boxes. In such cases, functional subclasses such as 'light' boxes and 'heavy' boxes are required for further specifying which boxes the interactants can carry around and which they cannot. In this study, functional subclasses of classified entities were not required for any utterance.

11.2.3 Rights check

Rights can be subdivided into two groups: positive rights and negative rights. Someone's positive right is a social norm whereby other members of society are required to provide this person with goods or services. Someone's negative right is a social norm whereby other members of society are prohibited from interfering with that person's action. In this study, the wheelchair cannot interfere with any user action, therefore negative rights are irrelevant for our concerns.

Positive rights apply to clients of services. If a gait-impaired human tells an intelligent wheelchair that he or she wants to go to the bathroom, it is this person's positive right in this situation that makes this utterance be a demand for a service, not a demand nor an offer of information. In this situation, the gait-impaired human is a potential client for the 'action per locution' of going somewhere.

To automate this understanding, the system creates one entity class for potential clients of each service and action per locution. For the services of taking people somewhere, the potential clients include all gait-impaired humans and excludes all others.

The right checker rely on these inferences to determine that the experiment participant, interpreting a gait-impaired human, has the right to go somewhere by wheelchair whereas the researcher, interpreting a human who can walk, does not have this right. If the user does not have the right, the potential semantic structure is marked as 'discarded'.

In the same way as for capacity, actions per locution and services can be further specified by restricting complement types. For instance, a user may have the right to be taken into his/her own bedroom, but no right to be taken into someone else's bedroom. Though such differences will affect interpretation in apartments shared by multiple people, they do not happen in the scenario nor with the vocabulary taken into consideration. For this reason, they were not implemented. These restrictions could have been included if required with no change in the right-checking module.

11.2.4 Duties check

From the service provider's perspective, an action per locution or a service are events to be carried out by those in charge. The interpretation of services and actions per locution is, however, different. For services, the service provider is determined by the speaker. The task of a duty checker is to verify whether the service provider is indeed required to do what he or she was told to do. For actions per locution, the service provider is sometimes mentioned as in *I want to recharge you*, where *you* is the wheelchair, who is required to recharge itself on its own. However, for all actions per locution, only the service client is always specified. When the service provider is not specified, the task of a duty checker is not only to verify whether the service provider is indeed required to do *what needs to be done by some member of society.*

In the situation studied, there are two people other than the speaker of the commands *Rolland, I want to recharge you* and *Rolland, I want to go to the bathroom*: the researcher and the wheelchair. If the wheelchair has the duty to carry out these material changes in the situation, it should do what needs to be done on its own. However, if the researcher is the one who has this duty, the addressee is required to tell the correct person to perform the material changes. In other words, once a demand is considered legitimate, other members of society need to divide the labour amongst themselves, each one doing his or her part according to their duties. In the present situation, the wheelchair has to perform the material changes on its own.

To implement this understanding, the same approach was adopted as the one adopted for clients, doers, and goals. A class of potential providers was automatically created for each action per locution and service. Per definition, potential providers include intelligent wheelchairs for all actions per locution and services a wheelchair is capable of doing.

11.3 Tacit and spoken contracts

The checks in the previous sections are sufficient if a single wheelchair is used in the situation and a single present human can use the wheelchair. However, as soon as two wheelchairs can hear one potential user speaking or one wheelchair hears one of two potential users speaking, wheelchairs will need to be assigned to users, otherwise they will react to the wrong user and multiple wheelchairs will react to the same user. At least two different approaches can be adopted in such cases: tacit and spoken contracts.

A contract between two parties is a set of rights and duties that each of the parties have only towards each other. Tacit contracts (unspoken contracts) are

a set of rights and duties acknowledged by two parties whether or not they spoke them. For instance, it is reasonable to assume that the person sitting in a shared wheelchair is the current user of the wheelchair and that the commands of this user count whereas the commands of other potential users do not while this person is sitting in the wheelchair. The assignment of a wheelchair to a human in such cases might be tacit. Nonetheless, in this situation, the wheelchair should consider the rights and duties it has in this contract, which do not apply to other potential users in the environment.

The other approach is to have a spoken contract. This can be realised in multiple ways: it can be a dialogue whereby the wheelchair is told who its user is; it can be a configuration procedure whereby the voice id of the user is selected on a graphic interface; it can be any other interaction carried out for associating the wheelchair with a particular user. From this point on, a set of rights and duties will only apply between a particular wheelchair and a particular human.

Such fine grained control of dialogue will be necessary in any deployed product, but it was not necessary for the evaluation experiment proposed in this research. Additional rights and duties applicable to individuals can be added to the implemented checker of rights and duties without any change in architecture.

11.4 Reduced Scope

Several combinations of checks were implemented in different versions of the dialogue system whose evaluation is reported in Chapter 14. When optimising the system for efficiency, we kept only the minimal set of checks for the most likely utterances. These include potential doers (capacity), potential clients, and potential providers. The assumptions made turned out to be fine. The reduction did not cause any misunderstanding in the evaluation. Nonetheless, a deployed wheelchair would require the full set of checks to avoid any misunderstandings of regular spoken language around it.

11.5 Conclusion

In this chapter, I explained how a reasoner was used to determine whether potential alternative semantic structures were plausible or not and to discard those that did not suit the situation. In particular, participant roles were checked. It was checked whether an action or service medium affords the change described and whether the action or service doer is capable of carrying out the change. Then, when it comes to agents, it was checked whether the described entities are potential actors, potential clients, or potential providers. In all those cases, whenever a checker detects an implausible participant role, it discards the corresponding understanding proposed by the lexicogrammatical analyser.

Chapter 12

Nexus Integrator

Users make commands to the wheelchair to take them somewhere not only by telling the wheelchair to take them to those places, but also by telling what they want to do or what they want the wheelchair to do. Both the service of taking the user somewhere and the action the user wants to perform are represented linguistically as configurations of elements. In this chapter, I describe the component that tackles the logical inferences in Section 1.1.4 by mapping a represented configuration of elements onto another through logical relations. This component either *consumes* two configurations to *produce* the logical nexus between them or it *consumes* one configuration to *produce* a logically related configuration. Four particular cases are discussed: the interdependence between the location of the wheelchair and the location of the wheelchair user, actions and services to be performed in a specified location, actions and services that can only be performed in a particular location, and the distribution of 'hard work' between members of the represented group.

12.1 Location interdependency

Commands for the wheelchair to go somewhere such as Examples 163-166 represent one of the wheelchair services.

(163) komm zum Bett
come to the bed
come to the bed

(164) fahr zur Ladestation
go to the charging station
go to the charging station

(165) fahr zum Bett
go to the bed
go to the bed

(166) fahr zum Waschbecken
go to the wash basin
go to the wash basin

These commands are typically uttered in different situations. When users tell the wheelchair to come to bed or to the charging station as in Examples 163 and 164, they are usually not sitting in the wheelchair and the relevant location is that of the wheelchair relative to the user in Example 163 and relative to the charging station in Example 164. In contrast, when users tell the wheelchair to go to bed or to the wash basin as in Examples 165 and 166, they are typically sitting in the wheelchair and they want the wheelchair to take them there. The relevant final location is not that of the wheelchair relative to the object, but that of the user. The final wheelchair location is such that the final user location works for the action the user wants to perform, be it an action performed while sitting in the wheelchair or the action of moving from the wheelchair to bed or to the sofa.

Nonetheless, if the user makes a command for the wheelchair to go to the bed or to the wash basin while sitting elsewhere, the wheelchair cannot interpret these utterances as commands to take the user somewhere. In other words, this dependency between locations only applies while the user is sitting in the wheelchair.

To implement this dependency, the relevant potential final positions of the user were manually listed and manually mapped to the corresponding positions of the wheelchair. Both the potential positions of the wheelchair and those of the user were manually ascribed classes such as 'in front of the wash basin' and 'next to the wash basin'. These positions and the resulting map are loaded by the wheelchair and the mapping becomes active whenever the user is sitting in the wheelchair. In this way, when the user asks the wheelchair to come to bed, the wheelchair position relative to the user is retrieved, but when the user tells the wheelchair to go to bed while sitting in it, the position chosen is the wheelchair position relative to the bed corresponding to the user position relative to the bed. In this way, the wheelchair can choose one position in case the user wants to move into it and a different position when the user wants to move onto the bed.

Representations of transitive services such as taking the user to the wash basin and transitive actions such as taking the book to the sofa also trigger implicatures. The wheelchair can only take the user somewhere if the user is sitting in it and users can only take the book somewhere if they are holding the book. For this reason, if a user tells the wheelchair to take them somewhere while sitting on the bed or sofa, the wheelchair first needs to go to them, they need to move into the wheelchair, and only then can the wheelchair take them somewhere. The same applies to taking the book somewhere. Users first need to be taken to the book and pick it up before they can take the book somewhere else. In turn, taking the book somewhere else implies a service of the wheelchair.

To model these dependencies, three trackers of movable entities were created, one for the wheelchair, one for the user, and one for the book. The model works as a stack. If the user is sitting in the wheelchair, his/her position is determined by the wheelchair's. If the book is on the wheelchair, its position is also determined by the wheelchair's. And if the user is holding the book, its position is determined (for simplicity) by the user's, which in turn might be determined by the wheelchair's. A motion of the book corresponds to a motion of the user and a motion of the user corresponds to a motion of the wheelchair. So the precondition for the book to move is for the user to be holding it and a precondition for the user to move within the apartment is for the user to be

sitting in the wheelchair.

In the implementation, a very simple planner was used. The planner creates a story ending in the represented action or service and adding events that accomplish preconditions: these events include wheelchair services and user actions that the wheelchair can count on such as picking up a book or moving into or out of the wheelchair. The planner is hardcoded and would need to be replaced for more sophisticated planning.

12.2 Actions/services to be done in a specified location

When the user tells the wheelchair he or she wants to perform an action at a particular location such as reading a book on the sofa (Example 167), the precondition for the user to perform this action is for the user to be at the given location.

(167) ich möchte das Buch auf dem Sofa lesen
I want to the book on the sofa read
I want to read the book on the sofa

For this to be a command, the user cannot be already on the sofa. This being the case, this precondition will correspond to a postcondition of a necessary movement by the user. Therefore, the wheelchair will need to plan the sequence of actions and services necessary for the user to be there. It can do this in the way described in the previous section, namely by using a planner. This is the way the understanding of such commands was implemented in this study.

12.3 Actions/services that can only be done in a particular location

The actions to open the door and to pick up a book (Examples 168 and 169) can only be done when the user is next to, respectively, the door and the book. Therefore, the precondition for the user to perform these actions is for the user to be in a location relative to these objects.

(168) ich möchte die Tür öffnen
I want to the door open
I want to open the door

(169) ich möchte mir das Buch holen
I want to me the book pick up
I want to pick up the book

The door location relative to the apartment does not change over time and the book location is tracked by the wheelchair. Therefore, if the user is not next to these entities, the precondition for these actions corresponds to a postcondition of a movement by the user. Once the necessary position of the user is determined, the planner can determine the sequence of actions and services necessary for the user to be there.

In the implementation, goal-affecting actions triggers implicatures whenever the goal is not an actor part and the actor is not already next to the goal. This approach works for this morning routine because all goal-affecting actions in this category required physical contact between actor and goal. If remote actions were included in the scenario such as turning on a TV with a remote control or the ceiling lights with a light switch, a subclassification of goal-affecting actions into contact actions and non-contact ones would be necessary. This was not the case for the actions in the morning routine.

In addition, services and actions that do not affect a separate entity such as recharging oneself and washing one's own hands do not require movement to reach the entity to be affected. However, they may require using a facility such as a charging station as in Example 170 or a wash basin as in Example 171.

(170) lade dich auf
recharge yourself .
recharge yourself

(171) ich möchte mir die Hände waschen
I want to me the hands wash
I want to wash my hands

For services, the wheelchair needs to be next to a facility to use it whereas, for actions, it is the user who needs to be next to the facility. In the implementation, a reasoner is used to determine which kind of facility is required for each kind of action. The location relative to the facility is inferred based on definitions associated with a class expression. For instance, washing one's own hands is a class expression that defines a figure type. In turn, a place of this type of figure is a class expression that defines a necessary actor location, which is described as a location next to the wash basin. With these inferential rules, the reasoner can determine that the actor needs to be at the wash basin for washing his or her own hands. Rules like these ones were defined for each action and service type that requires a facility.

Finally, the user's and the wheelchair's relative positions inferred as preconditions for the actions and services are the destinations represented, respectively, in Examples 172-175

(172) fahre mich zur Tür
take me to the door
take me to the door

(173) fahre mich zum Schreibtisch
take me to the desk
take me to the desk

(174) fahre zur Ladestation
go to the charging station
go to the charging station

(175) fahre mich zum Waschbecken
take me to the wash basin
take me to the wash basin

For these destinations to be the necessary locations for performing actions and services, the wheelchair takes only potential actor/provider locations that are preconditions for potential actions or services in the situation into consideration. In the implementation, this was achieved by using a predefined list of user and wheelchair positions that corresponded to the actions and services that the user and the wheelchair could perform in the morning routine.

12.4 Distribution of labour

When the user and the wheelchair are represented as a team (Example 176) performing actions, they perform actions per labour, not per locution, because no external agent needs to carry out the intended changes. However, labour is distributed amongst team members. For instance, the labour of going to the door is carried out by the wheelchair and the labour of opening the door is carried out by the user.

(176) fahren wir zur Tür, um die Tür zu öffnen.
go let's to the door to the door open.
let's go to the door to open the door.

To distribute labour, a capacity check was adopted. If the speaker can perform the represented labour by him or herself, the labour is understood as equivalent to an action per labour. If the speaker cannot perform the described labour, it is understood as equivalent to a service per labour. Once labour is distributed in such a fashion, Example 176 becomes equivalent to Example 177.

(177) fahr mich zur Tür, damit ich die Tür öffnen kann.
take me to the door so that I the door open can.
take me to the door so that I can open the door.

Distribution of labour was implemented with capacity checks. However, these checks had a high computational cost because they demanded refreshing the reasoner whether or not labour needed to be distributed. When the code was optimised for the evaluation experiment, these checks became good candidates for deactivation. Instead of reimplementing this module in a more reasonable fashion, due to time pressure and the low frequency of this phenomenon, I simply deactivated the checks altogether. Utterances such as these did not occur in the evaluation experiment.

12.5 Multiple contributions in a single move

Discourse contributions are classified according to the logical relation between the represented process and the service under negotiation. Contributions that represent the service being exchanged are classified as *explicative*, those that represent a process that implies the service being exchanged are classified as *implicative* and interjections such as *ja* (*yes*) and *ok* (*ok*) are classified as *interjective*. A dialogue move is realised by one or more interjections and clauses that imply, or represent the same service under negotiation (see Example 178).

(178) fahr mich zur Tür. ich möchte die Tür öffnen.
take me to the door. I want to the door open.
take me to the door. I want to open the door.

In the implementation, whenever an explicative and an implicative clause were uttered for the same service, the two of them were understood as a single move. In the case above, a single command. Interjections and accompanying clauses for the same service were also understood as a single move.

12.6 Conclusion

In this chapter, I described how logical inferences were realised with a simple planner for dependent movements and a reasoner for relating actions and services with the necessary location of actors and service providers. I also described how labour can be distributed amongst members of a team with capability checkers and how two clauses were integrated in a single move. This encompasses all inferential phenomena necessary for the collected utterances in the Wizard-of-Oz experiment.

Chapter 13

Move Integrator

The same utterance by the same person may mean different things depending on the situation in which the interactants find themselves and the tasks that they assume they agreed to do. In this chapter, I describe the component that keeps track of the ongoing exchanges in the current situation as well as the potential moves that can be performed at the current moment, thus tackling the linguistic phenomena described in Section 1.1.5. This component is responsible for updating the discourse state by integrating an utterance as a move in an exchange taking into consideration other non-verbal moves that were integrated so far.

13.1 Initiating an exchange

As long as utterances are composed of only one clause and no implicatures occur, Halliday and Matthiessen's model works adequately. The grammatical symptoms of a speech function will be fully seen at the clause rank. However, as soon as utterances are composed of two or more independent clauses, each independent clause has the grammatical symptoms of its own contribution for the dialogue move. Each one could realise the dialogue move on its own, but they do this in parallel, though realising together a single dialogue move. In contrast, if a logical relation between the two figures is represented, the sequence realises a dialogue move, not all of its constituents in parallel.

Because each independent clause realises a dialogue contribution on its own and can realise a dialogue move on its own, one can choose between making an explicative contribution such as *fahr mich zur Tür* (*take me to the door*) and an implicative contribution such as *ich möchte die Tür öffnen* (*I want to open the door*). As a consequence, the contribution features 'explicative' and 'implicative' are mutually exclusive, thus belonging to the system of CONTRIBUTION TYPE. These two types of contribution are realised by semantic restrictions. Explicative imperative contributions have either a service per labour as process and the addressee as service provider or an action per locution as process and the speaker as service client. Implicative imperative contributions have an action per labour as process and the speaker as actor.

The alternative potential semantic structures recognised by the lexicogrammatical analyser have a feature for the type of contribution based on lexicogram-

matical features including mood as word order, word features, auxiliaries, and adverbials, and modality as word features, auxiliaries, and adverbials. The task to check whether the semantic restrictions specified above apply is left for the integration process.

The move integrator checks whether the actor is the speaker in implicative imperative clauses, whether the service provider is the addressee in explicative imperative clauses, and whether the implicative imperative clauses imply the service represented by the explicative imperative clause if two clauses are present.

13.2 Continuing an exchange

In the linguistic data, all exchange-initiating moves were commands. There were no preparations for a command such as *can you do me a favour?* or *can you help me?*. For this reason, in the implementation, once the exchange is initiated, commands are added to a new exchange as a request (see Section 5.10). All continuing moves such as the command undertaking, execution, thanking, and welcoming are added to an ongoing exchange.

For this reason, the move integrator keeps track of the ongoing exchanges of service. The wheelchair, user, and book trackers fire circumstantial attribute change events and the move integrator listens to those events and integrates the execution whenever the service is complete.

In a first implementation, all discarded understandings of a command were compared to the observed circumstantial attribute change so that the wheelchair could understand utterances such as *zum Bett* (*to the bed*) uttered at room entrance as evidence that the user thought the wheelchair had misunderstood the previous utterance. By doing this, the wheelchair would have been able to respond *ich weiß* (*I know*). However, the wheelchair delay when answering to requests was so long (20 seconds) that this feature became unreasonable. During the optimisation phase, I removed all inferences by the wheelchair whether the user assumed it had misunderstood the previous commands. In the evaluation experiment, the wheelchair simply ignored all commands the user made while it was executing a command, even process phase commands such as *stop* due to the long delays.

13.3 Conclusion

The move integrator implemented in this research initiates an exchange whenever a command for the wheelchair to take the user somewhere is realised and it keeps track of ongoing exchanges of service. This move integrator implemented is capable of integrating all utterances in service exchanges between a wheelchair and a user in our linguistic data. However, how well the more fine-grained move integration works in practice can only be evaluated empirically if efficiency related issues with the current implementation are tackled first.

Part IV

Evaluation and Conclusion

Chapter 14

Evaluation

In Chapters 5-6, I described the creation of a taxonomy based on a corpus of spoken commands and, in Chapters 7-13, I described the dialogue system and its components for the automatic understanding of spoken commands. Here I report the evaluation of the dialogue system in two criteria: how frequently user commands are properly processed and how frequently a user produces at least one command that is properly processed when trying to execute a task.

Speech recognition failures were ignored in this evaluation because the third-party speech recognition module and the used microphone are not products of this study. Even though users spoke the utterances through a microphone and their utterances were automatically recognised and processed by the wheelchair during the experiment, for the purpose of evaluation, non-recognised utterances were manually transcribed and given again to the wheelchair in written form after the experiment run to test whether they would have been understood in the given situation had they been recognised.

This chapter is divided into three parts: 1) the description of the experiment and employed success criteria, 2) the results and 3) a discussion of specific failures focusing on what is still needed in which component for the wheelchair to understand the non-covered utterances.

14.1 Experiment

Experiment runs took place in the CREATE laboratory of the Institut für Anglistik, Amerikanistik, und Romanistik (IfAAR) at RWTH Aachen University. The evaluation experiment was carried out in a virtual environment representing the laboratory BAALL. This virtual environment was implemented as a JAVA application and it was displayed in an operating system window. For the experiment, a Windows tablet with headsets and a microphone was placed on a central table in the laboratory. A camera was placed on a tripod facing down onto the tablet. The camera faced what participants see including the tablet screen, but not the participants themselves.

The virtual environment window displayed BAALL viewed from above, the wheelchair in the living room and the participant sitting on the bed. The participant could talk to the wheelchair through a microphone and could listen to the wheelchair through headphones. All services by the wheelchair were

graphically displayed through animations. Both the wordings that the system recognised and those that it produced were programmatically logged.

Figure 14.1 is a screen shot of the virtual environment. There are two internal walls dividing the apartment into three open spaces. Each open space is further divided into two areas delimited by floor type. The top right area is the bedroom, the bottom right area is the office, the bottom central area is the bathroom, the bottom left are is the kitchen, and the top left area is the living room.

Figure 14.1: A bird-view of BAALL

The apartment is viewed from above and not from the perspective of the wheelchair user. The wheelchair user is represented on screen as a green avatar. All avatar's motions such as moving into and out of the wheelchair as well as pick up and laying down the book somewhere are performed automatically without participants' intervention. They are also animated. The motion of the wheelchair takes place on predefined paths, which are segments within a directed graph of paths between positions. The wheelchair always chooses the shortest path from that graph. Some positions in that graph of paths have the same identity as functional positions where actions and services can be performed by labour or that their preconditions can be reached by labour. These functional positions were the only positions in the graph that are accessible linguistically, that is, the only positions that can be represented and/or inferred. All other positions were only relevant for the subdivision of a trajectory into directed path shapes, which are useful either for performing a smooth motion or for reusing path segments across different routes. In this sense, intermediary positions in a smooth displacement exists whenever the shape of the path changes and whenever two different routes between linguistically accessible positions begin and end sharing a path segment such as moving through a door. No paths were created for routes that were not represented and no paths were created which do not fit any route. Therefore, the underlying path graph is 100% linguistically based since each path segment was represented at least once linguistically.

In the preparation for trial runs, the tablet to be used could not run the speech recogniser with the original grammar due to memory limitations. The grammar had to be reduced (see Section 14.1.2). In trial runs, the dialog system showed it was too slow for being on any practical use. For this reason, it required speeding up (see Section 14.1.1). In the following, I explain how the dialog system was sped up, how its recognition grammar was reduced, and what was done in each step of each experiment run.

14.1.1 Speeding up

Each dialog system component was produced separately and tested separately through unit tests. When put together, the response delay of the wheelchair was over 300 seconds for each utterance in a fast computer. However, I did not have access to a Vocon 3200 licence for Mac OS X but only for Windows 32-bit and the computers I had access to ran either Mac or Windows 64-bit[1]. The only Windows 8 32-bit machine I could find in the market was a low-end tablet with 1.33 GHz processor and 2GB ram. In that tablet, the response had a delay of over 600 seconds. In other words, the wheelchair spoke nothing and did nothing after the user said *Roland, komm her* (*Rolland, come here*) for over 10 minutes. This delay prevented any testing of the system.

Since this thesis concerns the automation of the understanding of spoken commands, the modules that needed testing were the language analysis and the meaning integration steps. It is not an aim of this thesis to deliver a completely functional product that can be sold in the market without any changes. Because of this, the amount of engineering effort applied to speed up the system needed be only that which is necessary to make those components testable in the new experiment. For that reason, a series of pilot studies were carried out to verify the maximum delay tolerance of participants and to test strategies to extend that tolerance.

My initial attempt to increase delay tolerance was to play a bell sound every time the wheelchair recognised speech, expecting that experiment participants would understand that the speech was recognised and would wait. That did not turn out to be the case in the first pilot study. The two participants of the first pilot experiment started and kept on speaking continuously without an interval or large pauses as the bell kept on ringing. Given that each utterance added additional parallel processes, the delay of the wheelchair reaction was actually multiplied by the number of parallel processes initiated. More than 25 processes were started by each participants, what would result in a delay of over 5 hours if the participant said nothing more for that period of time. For this reason, I chose to deal with the timing issue.

The second attempt to increase delay tolerance of experiment participants for response delay was to add "status messages" in different points and test them with users. This was an iterative process of adjusting and testing again. In the end, I achieved the longest delay tolerance with a framing of the system's capabilities during the instruction and two status messages. The framing consisted of telling the participants prior to the experiment that the first bell sound meant that the wheelchair recognised what they said, that the wheelchair would tell them when it starts to think about the relevant positions and that it would tell them when it starts to plan the way to the destination. That process would take 40 seconds. Moreover, I also told participants that, when the wheelchair arrives at a location, it needs about 40 seconds to relocate itself and notice the changes in the environment and that it would ring a bell again when it is ready for a new command. The added status messages were *ich denke über die relevanten Positionen nach* (*I'm considering the relevant positions*) and *ich plane den Weg* (*I'm planning the way*). The first status message was set to be triggered at the start of the reference handling process, whether or not a location is mentioned – *komm her* (*come here*) does not contain a Place reference,

[1] I had no permission to install a 32-bit versions of Windows in them.

but the status message is triggered anyhow – and the second message was set to be triggered at the start of figure integration even though the actual paths are decided by another component running in parallel to the dialogue system. In other words, the status messages do not always correspond perfectly to what is actually going on at each stage, but they are temporally located in such a way as to maximize delay tolerance.

To increase tolerance yet more, I separated imperative responses into two speech bursts: one for *ok* (*ok*) and another for the following contribution such as *ich fahre dich dahin* (*I'll take you there*). As a result, the *ok* (*ok*) contribution could be anticipated in 10 seconds in the fast notebook. These changes enabled the wheelchair to occupy the speech channel in approximately regular intervals. In that way it could claim the dialogue turn while creating an expectation of when the next speech burst should come. In addition, contributions were selected so that the wheelchair was perceived by participants as holding the turn. Moreover, participants were told in advance about the sequence of contributions that they should expect the wheelchair to make, what is likely to have reinforced the turn-holding effect of the selected contributions.

After this, there was the issue of parallel commands reducing processing speed. Not answering to them quickly implied that the wheelchair would not claim the turn and that more commands would be made with the expectation that the wheelchair did not hear or understand anything at all. Because of that, commands made while the wheelchair is understanding prior commands and trying to respond to them could not be processed. To alert participants that they need to wait, I added a blockage of commands at the language analysis module. Once a command was recognised, the system was locked for new commands until that command was fully processed. If a new speech burst is recognised during this period, the wheelchair was programmed to say *Moment!* (*Wait a moment!*).

The rest of the work was invested in reducing processing time. The main issue was the fact that the system contained a single ontology with every phenomenon that the wheelchair remembered, whether inserted by programmers, observed, or represented linguistically. Every time a new phenomenon was represented linguistically, the wheelchair added that representation and all representations to be discarded to the assertion box. After each addition and deletion, the inferences were recalculated. Additions were not a real issue since the adopted reasoner HermIT deals well with progressive additions. Time increase appeared when assertions were removed and inferences needed to be recalculated. Because of that, I substituted adding candidate representations and removing them by further specifications that a representation was grounded.

For instance, prior to the optimisation, candidate identities of represented things were associated with a represented thing through the owl property ***can-be-identical-to***. When grounding occurred, these potential relations between represented things were removed and the most relevant observed thing was associated with the represented thing via the owl property ***is-identical-to***. After the optimisation, candidate identities were associated with an element in the same way through the owl property *can-be-identical-to*. During grounding those properties were not removed. Instead the most relevant identity associated with the represented thing through the owl property *is-identical-to* and the less relevant identities were associated with the represented thing through the owl property *could-also-be-identical-to*. These last two properties were made sub-

properties of *can-be-identical-to*. This further specification did not cause the recalculation of inferences that was causing a slow down in processing.

This approach of replacing removals of ontological assertions by further specifications was adopted wherever it improved speeds. In other points of the code where the ontology was serving as a simple data store and no reasoning was being performed, additions and removals of ontological assertions were simply substituted by object-oriented code.

Another type of improvement made was to precalculate responses. For instance, whether a contribution is imperative is verified very often. Since the wheelchair had simple classification of contributions without any logical definitions, no module required a reasoner to infer whether a particular contribution is imperative or not. A hash map from classes to all its subclasses suffices. Using such specificities, a module can preload all classes of contribution that are imperative in advance and keep them in a hash set. Therefore checking whether a contribution is imperative can have a cost of O(1). In other words, the wheelchair does not need to refresh the reasoner for checking whether a contribution is imperative.

Another set of changes consisted in taking advantage of the compositionality of structures. Every time a component asserted a property such as *eum:provider*, other components could count on the fact that this property was asserted and not inferred. Therefore, they did not need to use the reasoner for them. Other properties such as *eum:providerIn* were inferred. Because of that, every time a module needed to navigate in that direction, they needed to use a reasoner. There is, however, no intrinsic need to use an ontology for most of these small inferences. The same behaviour could also be achieved through a well-designed simple object-oriented structure.

Finally, many submodules of the code were deactivated or removed because they were slow. For instance, in the reference handler, I implemented a module that imagined and tracked fictional entities such as the apartment building in which the apartment is located and fictional relations such the fictional possessive relation between a user and a book represented as *mein Buch* (*my book*). The maintenance of fictional entities and relations was taking approximately 15 seconds. Since they were represented twice in the initial experiment, I opted to ignore them altogether instead of optimising this submodule, which could be done with a few days of refactoring.

Taking advantage of all such cases, I was able to reduce the delay for the fast notebook from 300 seconds (s) to less than 50s between the bell sound and departure (response delay) and from 200s to less than 30s between arrival and the second bell sound. The delays were approx. twice as long in the tablet, varying significantly depending on the number of references and the number of inferential nexuses that needed to be integrated.

In short, after reviewing the code, it can be said that there are very few places in the code where complex reasoning take place. The most prominent of them is reference handling for which, known entities must be taken as identical to represented entities. Here an assertion box is useful. However, using a reasoner over a large a-box is a very expensive solution for most problems, which can usually be solved with a t-box alone or simply no reasoner. From code inspection, I assume a much more drastic improvement in speed might be achieved if small a-boxes are created only when a complex reasoning is deemed necessary and only for the relevant instances and relevant classes. Since such

an optimisation procedure would require a larger engineering effort than the available time resources I had, they were not performed.

14.1.2 Reducing speech recognition grammar

Another issue of running the dialogue system completely in a low-end 32-bit Windows tablet consisted in the limits of physical memory of the machine. The code is fully implemented in Java with the exception of the speech recogniser and the speech synthesiser. The speech recogniser is Vocon3200. A Java native interface (JNI) was implemented to access a windows library generated with C++. The operating system for this library needs to be 32-bits and the memory that the library consumes needs to be contained within the limits established by the Java Runtime Environment. The main issue that resulted from these limitations is that the size of the grammar had to be limited. Otherwise, more memory would be allocated than what was programmatically available. Increasing the memory space of the native interface would not be a good solution because other components also needed space. For this reason, the speech recogniser grammar had to be reduced.

The reduction criteria were the following: as mentioned in Chapter 8, all speech segments containing clause fragments that were meant for progressive text construction were removed.

The commands to reorient the wheelchair were only made during the Wizard-of-Oz experiment because the final positions of the wheelchair were not functional enough. Since this time the experiment was done in a virtual environment and the avatar was able to do all the actions it needed to do, these commands would be unlikely in the new setting, which turned out to be the case. For this reason, they were removed from the restriction grammar. Similarly, utterances such as *pass auf* (*watch out*) and *meine Beine* (*my legs*) were also unlikely in a virtual environment, so they were also dropped.

Moreover, there were clauses described by Elisa Vales as 'self-talk' that were not dealt with in any of the following components such as *[jemand hat] an die Tür geklopft* (*[someone] knocked on the door*). They were already ignored as if nothing was said at the integration steps. Because of this, they were also removed from the recognition grammar.

Finally, different ways of representing the same set of utterances lead to different memory allocations. For this reason, some refactorings had to be made to make sure that Vocon3200 structures in memory were as small as possible. This process consisted of testing different grammars with the same linguistic potential and checking which one passed the memory threshold.

With these final adjustments, all major commands to go somewhere or to take the user somewhere and all labour representing contributions were kept recognisable.

14.1.3 Differences between experiments

The purpose of the interaction from the participant's perspective is the same in both experiments. However, the purpose of the experiment is different. This time I did not aim at collecting a representative set of utterances, but at testing how well the software automates the understanding of commands.

Another change from the previous experiment to this one is that a virtual environment adds more distance between the present situation and the future situation in which someone will use an intelligent wheelchair, potentially reducing the purposefulness of the experiment from the participant's perspective. For instance, the image of BAALL is not BAALL itself. It represents BAALL. Because of that, when starting this experiment, I cannot start the instructions saying that what the participant sees is an apartment that was built to test technologies for people with motoric disabilities because that is not the case. The instructions, if they are to make sense and prime participants in a similar way, need to include that this is a view from above of an apartment that exists in Bremen, which was built to test technologies for people with motoric disability. It is the apartment that was physically built for testing technology, not the virtual environment. The virtual environment simulates that apartment for testing a dialogue system. As a result, all instructions had to be adapted.

In addition, physical things such as a fridge can be opened to reveal its contents to participants in BAALL, enabling a joke about there being beer in the fridge. Those comments were made to help participants put themselves in the position of the future wheelchair user. In a virtual environment viewed from above, it would be less of a surprise for a participant if a beer bottle were to be represented inside the fridge and it would be harder for an instructor to claim that the presence of a beer bottle is unexpected since anything is possible and controlled by the researcher in a virtual environment. For this reason, such comments needed to be removed from the instruction or substituted by something new.

Finally, an integration of two representations is necessary in the virtual environment. For instance, the participant is not on the bed when the experiment starts. A green avatar is. Because of this, the experiment instructor needs to tell participants that 'the green object on the screen is their avatar' and that 'they should put themselves in the position of a person who cannot walk (that the avatar represents) and needed to go around an apartment (that the virtual environment represents)'. In this sense, the participant collaborates with the voice of the avatar for the representation of the future situation and not with his or her physical behaviour. In an analogy with cinema, the difference between participating in these two experiments is similar to the difference between acting for a live-action movie and voicing a character for an animated movie. As a consequence, in the evaluation experiment, the instructions could not be the same as the ones in the previous experiment and needed to account for those differences in the participant's interpretation of a future wheelchair user.

In the following sections, I shall describe the steps of an experiment run.

14.1.4 Recruitment

Experiment runs were limited to a maximum of 30 minutes including instructions. A payment of 8 euros was offered so that recruitment became feasible. A total of 10 bachelor and master students took part in the experiment, of which 5 were male and 5 female.

The invitation process consisted of a researcher coming out of the laboratory into a study alley and verifying whether there were students doing other activities than studying (e.g. waiting for appointments, talking to peers, navigating on Facebook). Participants were not asked which language their mothers

speak/spoke with them (see Section 4.3) during invitation and were not selected on such criteria. Participants were invited to join the experiment in German and, if they understood the invitation and were confident that they could control a wheelchair that understands German, they were recruited. Some students did not understand what the researcher said and claimed they did not speak German, sometimes in German and sometimes in English. They were not recruited. Other students could understand the invitation, but were not confident that they could do the task, so they did not take part in the experiment either. Other rejections were justified on the basis of lack of interest in participating for the offered payment or no time for that in their schedules. Those who took part in the experiment either scheduled an appointment or joined the experiment immediately.

14.1.5 Apartment tour

The following script was read out loud for the students immediately after they signed a consent form ('Einverständniserklärung'). The image of BAALL viewed from above was visible on the screen of the tablet as the text was read.

14.1.6 Wheelchair Presentation

After presenting the apartment to participants, the researcher presented Rolland. This time no demonstration was made. There was a time interval of approximately 5 years between the two experiments and, since then, dialogue systems became a default functionality of smart phones. Possibly influenced by experience with deployed dialogue systems, the pilot studies showed that participants knew what to do and believed the system would work without any demonstration.

14.1.7 Purpose Construction, task explanation, last instructions

The remainder of the instruction script is almost identical to the one of the previous experiment. For this reason, I shall not report it again here. See Chapter 4 for more details.

14.1.8 Limitations

Before reporting the number of utterances correctly understood, it is relevant to set our expectations to the right level because the Wizard-of-Oz experiment conducted for data collection is quite different from the evaluation experiment.

In the first place, participants viewing a wheelchair and an avatar from above will perceive different phenomena from participants moving around an apartment in a wheelchair. As a result, they may represent different phenomena in their utterances. I have the impression that this happened and made the coverage smaller than what it would be for an experiment in a physical environment. Examples of this will be presented in the discussion.

In the second place, in the virtual environment the wheelchair will always reach a functional position when taking the user somewhere, an achievement that is not replicable in physical environments. For this reason, users are unlikely

Dies ist ein Labor in Bremen. Dies ist einfach eine Karte vom Labor von oben gesehen. Das Labor ist eine Forschungswohnung, die in Bremen tatsächlich existiert. Die wurde gebaut, um Technologien für Menschen mit motorischen Behinderungen zu testen. Sie ist wie eine normale Wohnung aufgeteilt. Hier ist ein Ort, wo man schlafen kann (zeigt das Bett). Hier sind Bücher (zeigt das Regal). Dieses Buch ist hier für das Experiment (zeigt das Buch *Merlin* auf den Tisch). Ich erkläre dir gleich, was deine Aufgaben sind. Hier sollte ein Becher mit Mundspülung gezeigt werden (zeigt den Ort auf dem Waschbecken). Für das system gibt es eins da, auch wenn wir es nicht sehen. Dieser Raum wurde so gebaut, dass man mit einem Rollstuhl reinkommen kann. Siehst du? (deutet mit zwei Finger) Es gibt eine Tür da und eine da (zeigt auf die Türen), aber die sind auf. Hier ist eine Dusche und hier ist eine Toilette (zeigt auf die Dusche und Toilette). Hier gibt es noch einen Tisch (zeigt den Esstisch). Hier sollte Essen gezeigt werden (zeigt einen Ort auf dem Tisch)! Der Rollstuhl kann sehen, dass es Essen hier gibt. Es wird einfach nicht gezeigt. Das ist der Kühlschrank (zeigt den Kühlschrank). Die Wohnung ist noch nicht komplett, es fehlen noch Sachen. Dieser Raum ist z.B. noch leer (zeigt den leeren Raum). Er ist tatsächlich leer dort.

This is a lab in Bremen. This is actually a map of the lab viewed from above. This is also a research apartment that really exists in Bremen. It was build to test appliances for people with motor deficit. It is organised like a normal apartment. Here is a place where we can sleep (points at the bed). Here there're some books (points at the shelves). This book is here for the experiment (points at the book *Merlin* on the desk). I'll tell you what your tasks are in a moment. A cup with mouth wash should be shown here for the experiment (points at a place on the wash basin). For the system there is one there even if we don't see it. This room was built in such a way that we can enter with a wheelchair. See? (makes a two-finger swing gesture) There's a door there and another there (points at them), but they are open. Here there's a shower and here there's a toilet (points at them). Here there's a table (points at the dining table) and here we should see food (points at a place on the table)! The wheelchair can see that there's food here, it's just not displayed. This is the fridge (points at the fridge). The apartment is still not complete. It still needs some extra things. For example, this room is empty (points at the empty room). It is really empty over there in Bremen.

Example 14.1: Script for the apartment tour

to make reorientation commands such as *ein Stückchen weiter nach vorne* (*a bit more to the front*) once they arrive at the final destination, which I did not have to treat. This means that part of the interaction necessary for a functional wheelchair was not covered in this experiment.

In the third place, since the wheelchair will not collide with the avatar nor look like it is about to collide with the avatar, all commands such as *pass auf!* (*watch out!*) and *meine Beine!* (*my legs!*) are so unlikely to happen that they could even be removed from the recognition grammar without any consequence to coverage. Together with reorientation commands, this means that we are likely to overestimate the coverage of all commands given to wheelchairs if we take the experiment results as representative without considering that it is covering only commands to take the user somewhere.

Und das ist Rolland (zeigt den Rollstuhl). Dieser Rollstuhl ermöglicht Menschen mit Behinderungen, sich in der Wohnung zu bewegen. Er kann uns hören und das sehen, was um ihm ist; man spricht mit ihm durch diesen Mikrofon (haltet Kopfhörer mit Mikrofon). Und er kann deinen Avatar auch sehen, wenn er rein kommt (zeigt den Avatar). Er ist ganz langsam weil das ganze System in einem Tablet läuft. Er wird viel schneller, wenn er mit riesigen Datencenters verbunden ist, wie Handys z.B.. Nach dem man ein Befehl macht, braucht er ungefähr 30 Sekunden, um über die relevanten Positionen nachzudenken, dann braucht er noch 30 Sekunden, um den Weg zu planen, und, wenn er ankommt, brauch er wieder 50 Sekunden, um sich wieder zu orientieren, bevor er ein neues Befehl annehmen kann. Wenn er versteht, was du sagst, klingt eine Glocke, und, wenn er wieder bereit ist, ein neues Befehl anzunehmen, klingt die Glocke wieder. Er kann uns gut verstehen aber er is gaaanz langsam. Er denkt seine Ladestation ist da wo er gerade steht (zeigt den Rolltuhl).	And this is Rolland (points at it/him). This wheelchair enables people with disabilities to move inside the apartment. It/he can hear us and see what is around it. We speak with it/him through this microphone (holds a headset). And it/he can also see your avatar when it comes into the bedroom (points at the avatar). It/he takes very long to respond because the whole system is running in a tablet. It/he will become much faster when it/he is connected to enormous datacenters like your cell phone, for instance. When one makes a command, it/he needs 30 seconds to think about the relevant positions, then it/he needs another 30 seconds to plan the route, and when it arrives at the destination, it/he needs yet another 50 seconds to get reoriented before it can accept another command. When it/he understand what you said, a bell will ring, and when it is ready to accept another command, the bell will ring again. It/he can understand us well, but it is veeery slow. It/he thinks its/his charging station is there where it is currently (points at the wheelchair).

Example 14.2: Script for wheelchair presentation

Finally, it is also expected that regional differences between German speakers in Bremen and German speakers in Aachen might result in uncovered lexical items and command styles.

14.1.9 Success Criteria

Since speech recognition and text production will be ignored, only five components of text understanding will be evaluated:

1. Lexicogrammatical Analyser
2. Reference Integrator
3. Figure Integrator
4. Nexus Integrator
5. Move Integrator

These components were evaluated as a group, not individually, for understanding of utterances and for task completion. If a user makes two commands

that are not understood and a third command that is, the user is able to complete the task of demanding and receiving a service from the wheelchair. In this case, one out of three utterances are successful and one out of one task is successful.

However, when speech recognition does not work, users tend to repeat the same utterance multiple times. If we count the number of utterances that would have worked if speech recognition were to work properly, we might have the following situation. The user makes the same command three times. The third time the utterance is recognised, a bell rings, but the wheelchair does not understand the utterance and tells the user it did not understand it. The user makes another command five times and it is never recognised. Then the user makes another command that is both recognised and understood. If we counted each time a user makes a command that would have been understood by the wheelchair had the speech recognition worked, we come to one out of nine utterances. This would have been a misrepresentation of what would actually happen if the wheelchair had recognised all commands because the user would not have repeated the same command after the wheelchair told him or her that it did not understand it. For this reason, we counted only different commands from the same user ignoring repetitions due to malfunctioning speech recognition.

In addition, when evaluating which component required improvement, I separated the coverage into multiple dimensions: I checked whether a command that was not understood would have been understood if it was made slightly differently. For instance, if the command *kannst du mich in die Küche fahren* (*can you take me to the kitchen*) is understood, but the command *könntest du mich in die Küche fahren* (*could you take me to the kitchen*) is not, the components doing reference, figure, and nexus integration do not need to be changed. Only the one doing move integration does. It requires a new class of imperative clauses. For this reason, the utterance coverage was split into reference coverage, figure coverage, nexus coverage, and move coverage.

14.2 Results

A total of 23 tasks of demanding services were observed. 17 were successful and 6 were unsuccessful. Among the unsuccessful tasks, 5 contained at least one utterance that the wheelchair would have understood had it recognised it and only 1 task contained no utterances that would have been understood had they been recognised. The components being tested had, therefore, a task success rate of approximately 96% for activities in a morning routine when ignoring speech recognition failures.

The experiment data confirms that the modules implemented for this research are general enough to achieve a large utterance coverage: 100% reference coverage, 80% figure coverage, 100% reference coverage, 90% move coverage. For the utterances covered by the linguistic resource, that is, the utterances that could be lexicogrammatically analysed, integration components achieve 100% recall and 100% precision for integration.

This state-of-the-art coverage was achieved with a Wizard-of-Oz experiment in which only 20 instances of commands were collected for each task, which serves as evidence that the development strategy utilised in this research is adequate for developing an initial version of a dialogue system whenever collecting

a corpus of utterances is expensive.

14.3 Discussion

When move and figure coverage was insufficient, this was reflected in lexicogrammatical analysis in missing vocabulary or grammatical structures, and also reflected in the coverage of speech recognition. When they were sufficient, contributions were properly analysed regarding vocabulary and grammar. For instance, a missing expression for processes of a given type in the vocabulary corresponded with a missing type of process in the ontology and missing affordances, capacity, rights, and duties at the figure integrator. This tight coupling between recognisable utterances, lexicogrammatically analysable utterances, and situationally integratable ones had the consequence that all occurring utterances that could not be integrated could also not be lexicogrammatically analysed.

In the following, I shall present all cases that were not covered and make a short remark as for the effort to cover them given the current architecture.

14.3.1 Figure coverage

Throughout the experiment, 11 figure types were realised. As explained previously, when establishing coverage, I shall count whether figures were covered and not whether the actual wordings were covered. For instance, participant 8 said the following:

(179) */// ich* **hätte** *gern eine Mundspülung /// [...] // eine Mundspülung haben ///*
/// I'd like to have a mouth wash /// [...] // have a mouth wash ///

In such cases, the first contribution was not grammatically identical to the last, but they shared the same figure. The collocation of the scope *eine Mundspülung* (*a mouth wash*) with the process *haben* (*have*) was not predicted because it did not happen in the Wizard-of-Oz experiment. The only collocation that was recognised was that of *eine Mundspülung* (*a mouth wash*) with *machen* (*make*), what resulted in the non-coverage of these two wordings. For that reason, even though those were two different wordings that were not covered, thus not recognised by the speech recogniser nor analysed by the lexicogrammatical analyser, they were not covered for the same reason, the absence of the figure *eine Mundspülung haben* (*having a mouth wash*) in the original linguistic data, thus also in the implementation. Had any instance of this figure happened the linguistic corpus, both would have been covered.

If we count instances in such a way, the number of figure instances in this experiment is 49, of which 39 were covered and 10 were not. This amounts to a coverage of 80%. In the following, I shall describe what was not covered, explain why they were not covered, and evaluate how much effort it would take to increase coverage given the current architecture.

mein Buch lesen, mein Buch holen, mein Buch holen

Contributions such as *ich möchte mein Buch lesen* (*I'd like to read my book*) and *ich möchte mein Buch holen* (*I'd like to pick up my book*) contain a reference to the speaker's book. In the experiment situation there were no books

that belong to the speaker. However, in the fictional situation enacted by the interactants, the human assumed that the book belonged to him or her. This fictional possessive relation between the human and the book was then used for reference.

Covering this instance is trivial. The reference integrator implemented can already cope with fictional possessions when the mentioned classified entity is unique in the situation ignoring who owns it. Whenever this happens, a fictional possessor edge is added to the set of possessed entities and the entity is added as possessed by that entity in fiction, that is, interactants are pretending this entity belongs to that possessor. This functionality was deactivated during this experiment in order to speed up the software and make it testable. If the routines were active, these instances would have been properly dealt with and 86% of the instances would have been covered.

in die Küche gehen, [am Tisch] essen

Other contributions such as *in die Küche gehen* (*[I'd like to] go to the kitchen*) and *[am Tisch] essen* (*eat at the table*) were not covered because the lexical items { *gehen* } and { *essen* } did not occur in the Wizard-of-Oz for these figure types. The figure of *in die Küche gehen* is a SimpleClientAffectingDisplacement and the service deu:WohinGehen3 had not been included as a taxon next to deu:WohinFahren3. Similarly, the figure of *[am Tisch] essen* is a SimpleActorAffectingAction and the action deu:Essen1 had also not been included.

Adding new lexical items and taxa to the system takes approximately 30 minutes per cycle plus 5 minutes per lexical item in the cycle, including adding test cases, compilation and rerunning the system. Adding two lexical items would take approx. 40 minutes for a researcher that knows were the Morph.xml and the Deu.owl files are. This process can be significantly improved and sped up if a taxonomy manager is implemented and developers can add taxa and lexical items through an graphical programming interface such as Google Flow.

eine Mundspülung haben

One contribution was not covered because the collocation between *eine Mundspülung* (*a mouth wash*) and *haben* (*have*) did not occur in the Wizard-of-Oz experiment. This collocation represents an action by an actor of doing a mouth wash that is to be enabled by a service where the actor of doing a mouth wash is a client of the action-enabling service. The contributions *ich hätte gern eine Mundspülung* and *ich würde gern eine Mundspülung haben* oppose *ich würde gern eine Mundspülung machen* in the sense that the second does not impose the restriction that the actor must be the client of a negotiated service. Both processes fit a SimpleActorAffecingScopedFigure, which is already covered by the figure integrator. To cover this address, one would need to add a new lexical item as in the previous examples. As for the taxa, not only the new taxon deu:EtwasHaben1S should be added next to deu:EtwasMachen1S, a collocational restriction deu:Habbar1S should also be added next to the restriction deu:Machbar1S, both of which need to be superclasses of deu:Mundspülung. In this way, both the collocational restriction and the objects could be added.

von dem Bett ins Bad, vom Bett ins Badezimmer, das Buch vom Schreibtisch holen

Currently, only destinations are further processed. Routes with both origin and destination add a further complexity to the system since they represent the first state of a series of state contrasts. In this sense, in addition to predicting those routes, it is also necessary to handle origins in both reference and figure integration. Implementing such a change should not take more than 2 days of work (10 hours) for a researcher who can run the system locally, whether or not they know already where to change the code.

sich etwas zum Essen machen

Finally, creative actions such as *sich etwas zum Essen machen* (*making something to eat*) were not dealt with in this research because they did not happen in the Wizard-of-Oz experiment. Reference integration must consider that the created goal of the action does not exist in the present world and figure integration should add that raw material should be a prime matter for creation. Therefore, the labour-enabling service would be to take the user to this raw material.

Adding this figure to two integration modules and to the analysis resource as well as the lexical items should also not take more than 2 days of work (10 hours).

14.3.2 Move coverage

For the figure of *eine Mundspülung machen* (*doing a mouth wash*), 17 types of move were covered (see Table 14.3) and, for the figure of *mich zum Waschbecken fahren* (*taking me to the wash basin*), a total of 22 types of move were covered (see Table 14.3).

These sets of nano-classes coincide for processes of the same general type. For instance, service-implying actions by the speaker always have the same potential set of nano-classes as in Table 14.3 whereas services by the addressee always have the same potential set of nano-classes as in Table 14.3.

Because of this, when counting the number of different contributions by the same user, I counted an instance of a new contribution whenever a new figure is represented and whenever a new nano-class of move was instantiated for an already represented figure. As a result, the series of contributions below has two figures and three different contribution:

(180) */// fahr mich ins Badezimmer // ich möchte mir den Mund ausspülen /// ich möchte mir den Mund ausspülen /// fahr mich bitte ins Badezimmer // ich möchte mir den Mund ausspülen /// fahr mich bitte ins Badezimmer ///*

/// take me to the bathroom // I would like to wash my mouth /// I would like to wash my mouth /// please take me to the bathroom // I would like to wash my mouth /// please take me to the bathroom ///

In Example 180, the figure of *sich den Mund ausspülen* (*washing one's mouth*) is represented three times with the same wording type. In contrast, the figure of *den Kunde ins Badezimmer fahren* (*taking the client to the bathroom*) is represented three times but for two different contributions. In the first

Service-implying action	Service
ich würde gerne eine Mundspülung machen	fahr mich zum Waschbecken
ich würde gerne jetzt eine Mundspülung machen	fahr mich bitte zum Waschbecken
ich würde jetzt gerne eine Mundspülung machen	fahr mich mal zum Waschbecken
jetzt würde ich gerne eine Mundspülung machen	fahr mich einmal zum Waschbecken
ich möchte eine Mundspülung machen	fahr mich einfach zum Waschbecken
ich möchte jetzt eine Mundspülung machen	fahr mich zum Waschbecken bitte
jetzt möchte ich eine Mundspülung machen	bitte fahr mich zum Waschbecken
ich möchte gerne eine Mundspülung machen	fährst du mich zum Waschbecken
ich möchte gerne jetzt eine Mundspülung machen	fährst du mich bitte zum Waschbecken
ich möchte jetzt gerne eine Mundspülung machen	fährst du mich mal zum Waschbecken
jetzt möchte ich gerne eine Mundspülung machen	fährst du mich einmal zum Waschbecken
ich muss eine Mundspülung machen	fährst du mich einfach zum Waschbecken
ich muss jetzt eine Mundspülung machen	fährst du mich zum Waschbecken bitte
jetzt muss ich eine Mundspülung machen	kannst du mich zum Waschbecken fahren
ich will eine Mundspülung machen	kannst du mich bitte zum Waschbecken fahren
ich will jetzt eine Mundspülung machen	kannst du mich mal zum Waschbecken fahren
jetzt will ich eine Mundspülung machen	kannst du mich einmal zum Waschbecken fahren
eine Mundspülung machen	kannst du mich einfach zum Waschbecken fahren
	kannst du mich zum Waschbecken fahren bitte
	mich zum Waschbecken fahren
	bitte mich zum Waschbecken fahren
	einfach mich zum Waschbecken fahren
um eine Mundspülung zu machen	
sodass ich eine Mundspülung machen kann	

Example 14.3: Nano classes of move for different figures

time it is represented, the word *bitte* is not present. In the other two times the word *bitte* is present in the middle of the clause. Because adding a *bitte* (*please*) changes the politeness level, we can say that the nano-class of move for the fist clause is different from that for the other two clauses. As a result, this passage has three different contributions.

Using this strategy for counting contributions, I come to 52 instances of nano-classes of contribution. From these, 48 instances were covered and 4 were not. This amounts for a coverage of 92%. The 4 instances that were not covered instantiate two different nano-classes. In the following, I shall describe these two moves, then I shall explain why they were not predicted and estimate the cost of adding them to the system.

ich bin im Bett, ich bin im Bett, ich sitze im Bett

In both experiments, when asking the wheelchair to come to them, some users made reference to the bed. Contributions were either *directive requestive explicative* as in *komm zum Bett* (*come to bed*) or they were representations of destinations such as *zum Bett* (*to bed*) or things such as *Bett* (*bed*). The destination, independent of whether it was represented or not in the previous clause, could be further restricted by restrictive relative clauses such as *hier wo ich bin* (*here where I am*) and *hier wo ich sitze* (*here where I'm sitting*). This means

that the clauses that oppose *ich bin im Bett* (*I am in bed*) in this situation are actually *komm her, hier wo ich bin* (*come here, here where I am*), *komm zum Bett, hier wo ich bin* (*come to bed, here where I am*), and *zum Bett, hier wo ich bin* (*to bed, here where I am*). If we assume that this is the case, the primary illocutionary force of this clause would be not only *directive requestive*, but also **relative-referential restrictive** in the sense that it restricts a reference to a thing whose social value in the situation is to function as relatum of a location to be taken as a destination of movement.

This kind of illocutionary force was not predicted in the evaluation experiment because it did not occur in the Wizard-of-Oz experiment. Implementing the understanding of this nano-class of move would be trivial for the current architecture given the fact that socially valued things are already taken into consideration. The change would consist of adding a new reference integration according to which the attributive figure is not taken as something that is stated by the speaker, but rather counted upon for identifying the referenced bed. In other words, this clause would restrict a reference to the bed where speaker is. In turn, the bed would be a useful thing for the current activity (see Section 1.1.3).

These contributions do not have a simple *affirmative requestive* force. It would be absolutely inadequate for the wheelchair to respond to the user's request *ich bin im Bett* (*I am in bed*) by saying *ok* (*ok*) or *ich weiß* (*I know*). Moreover, in some contexts of discourses, the contribution *ich bin im Bett* (*I am in bed*) could be an *element-interrogative response*. The created Example 14.4 would include such a case.

```
1: Human  Roland, kommst du bitte her!
2: Robot  wo bist du
3: Human  ich bin im Bett
4: Robot  ok, ich komme

1: Human  Rolland, will you please come here!
2: Robot  where are you?
3: Human  I am in bed
4: Robot  ok, I'm coming
```

Example 14.4: Created dialogue with clarification question

Alternatively, one could understand such clauses as both a statement that the user is in bed and as a command for the wheelchair to come to the user or to bed. In that case, there would be a primary illocutionary force and a secondary force: i.e. this contribution would be a statement of a represented figure (a request to add a state to the model of the present situation) and a command to do an implied service (a request to add a service to path from the present situation to a better situation). This understanding implies that the shared model of the situation does not include the position of the wheelchair user. Because of this, the wheelchair could respond to a request with two illocutionary forces – a primary and a secondary – such as *ich bin im Bett* (*I am in bed*) by saying both *ok, ich komme zu dir* (*ok, I'm coming to you*) or *ich weiß, ich komme zu dir* (*I know, I'm coming to you*). The first contributions of each response *ok* (*ok*) and *ich weiß* (*I know*) would collate with the primary illocutionary force whereas

the second contributions would collate with the secondary illocutionary force.

This alternative understanding of a contribution with two illocutionary forces does not compete with the understanding of the contribution with one. On the one hand, this understanding might be necessary if users indeed give new information together with a command. This has not been seen in the evaluation experiment, but it might happen. On the other hand, having only a double-force understanding might result in 'reactive-style' dialogues in which the wheelchair emphasises that the represented figure is no new piece of information by saying *ich weiß* (*I know*) repeatedly. In this sense, if we value the 'simplicity' of interaction, one way of guaranteeing both coverage and simplicity would be to support both understandings in parallel and then to integrate the most likely one for each given situation.

ich hätte gern eine Mundspülung

The second and last non-predicted nano-class of move was not predicted because it was not specified in the taxonomic end of the semantic network. See Table 14.5.

	NanoTaxonA	NanoTaxonB
etwas haben	*ich würde gerne etwas haben*	*ich hätte gerne etwas*
etwas tun	*ich würde gerne etwas tun*	**ich täte gern etwas*
etwas machen	*ich würde gerne etwas machen*	**ich ???? gern etwas*

Example 14.5: Two taxa that are subclasses of the same nano-class of illocutionary force

Table 14.5 shows that there are two ways of realising the same nano-class of move for the Process verb *haben* (*have*), namely *würde ... haben* or *hätte ...*, which is not the case for the other process verbs. Since there was no instance of an action or execution verb *haben* (*have*) in the Wizard-of-Oz experiment, this way of realising this nano-class of move could not be predicted.

Including this is quite strait-forward for the current architecture. It consists of adding a new semantic feature for mode and a new CCG category for *hätte* to the analysis resource. The actual nano-class is already implemented, only the nano-taxon B (a subclass of the nano-class) was not considered so far.

14.4 Conclusion

In this chapter, I described the evaluation experiment, reported the results, and discussed the few examples that were not covered. The experiment showed that the dialogue system had state-of-the-art rates of recognition. In particular, by separating the coverage issues in two dimensions - figures versus moves - I was able to collapse multiple understanding failures into categories in a way that allows us to determine what needs to be added to the dialogue system for these utterances to be properly understood. Once a new type of figure or a new type of move is added, all combinations of move and figure become readily available, a property that makes such a dialogue system easy to extend and improve.

Chapter 15

Conclusion

In this chapter, I present achieved goals and open issues as far as dialogue systems for intelligent wheelchairs are concerned.

15.1 Achieved goals

In this thesis, I aimed at achieving two goals:

1. Demonstrating how to recognise the user's intent relying on what the user meant by the words chosen, the situation the interactants are in, and the ongoing discourse of interaction, making use of only symbolic processing.

2. Demonstrating how to create a language-based taxonomy of simple things, locations and processes that can be integrated into a rule-based understanding module.

The first goal was broken down into the following subgoals:

1.a. Resolving reference to actual things in the situation and actual and potential positions of those things respectively in current and subsequent situations.

1.b. Assigning participant roles to mentioned things in the described processes.

1.c. Building a logical sequence of events from the current state of things in the situation to the end state most likely to be intended by the user given the utterance, situation, and discourse from a need ranking and planning perspective.

1.d. Accounting for personal rights and duties as well as potential misunderstandings by the user of what the wheelchair is doing or did.

The second goal was broken down into the following subgoals:

2.a. Assuring that grammatical composition enables semantic composition both in terms of experiential semantics and speech acts.

2.b. Assuring that each grammatical structure rank corresponds to a semantic structure rank, thus making groups and phrases correspond to semantic elements and clauses correspond to semantic configurations of elements.

2.c. Assuring that a multiword expression such as *dreht sich* (*turns*) is treated as a form of a single term associated with a single class of events.

2.d. Assuring that the boundaries of grammatical structures are not limited to the boundaries of written words[1].

2.e. Developing the theory of lexis in SFL with multiword expressions.

2.f. Developing the theory of logical inference in SFL.

2.g. That CCG can be employed to create a rank structure with multiword expressions, in particular, covering reflexive, divisible, and prepositional verbs in German.

2.h. That CCG can be employed to suggest illocutionary forces.

All of these goals were achieved. With the current architecture, I showed how to determine the user intent for all situations the interactants were in during a morning routine. This included resolving reference to present entities, descriptions of their current locations and potential destinations, determining the role each participant takes in the described processes, reasoning about the purpose of the described processes for establishing an end state, planning the desired changes from the present state to the end state, and accounting for the rights and duties of interactants during the process as well as for misalignments of representation understanding.

I also showed how to create a taxonomy of simple things, locations, and processes based on linguistic data from a Wizard-of-Oz experiment and how to utilise this taxonomy in a CCG parser, and subsequent integration steps. In particular, I demonstrated how to classify wordings in a way that grammatical composition is constrained by semantic composition regarding both experiential and interpersonal meaning. Grammatical ranks and semantic ranks were aligned, thus making groups and phrases correspond to semantic elements and clauses correspond to semantic configurations of elements.

When it comes to lexis, I proposed a method to determine the words that integrate of a multiword expression such as *dreht sich* (*turns*) and demonstrated how types of elements and figures can be associated with them. Moreover, with a custom implementation of a CCG parser, boundaries of grammatical structure were not limited to the boundaries of written words.

When it comes to the theory of lexis in SFL, in this thesis I explained how semantic features in a semantic network can be realised by atomic symbols and how atomic symbols relate to multiword expressions. With this addition, the theory of lexis in SFL becomes compatible with multiword expressions. In the same way, logical inference was treated as a systemic option: the speaker either representing the service being negotiated or representing an action that implies this service in one of finite ways.

[1] Which is relevant for understanding *der Küchentisch* (*the kitchen table*) as a reference to the same table as the one referred to in *der Tisch in der Küche* (*the table in the kitchen*)

And when it comes to practical implementation, I showed how to employ a CCG to a ranked structure with multiword expressions. I also showed how to employ a CCG to suggest composite categories of move for an utterance.

15.2 Limitations

This study is limited in several dimensions. All potential referents were in the socially construed situation. As a result, utterances about absent things cannot be understood. Moreover, only directly observable phenomena were taken into account. This meant utterances about *the distance between two entities* such as *der Abstand ist mir zu groß* (*the distance is too large for me*) cannot be covered with this approach as it is.

In addition, all potential referents the wheelchair identifies were entities, a subset of all potential referents excluding any unbounded portions of matter such as *air*, *water*, *food*, and so on. In other words, the context of situation for user utterances was a socially construed ontic situation. The dialogue system is currently also not capable of understanding reference to an exact, approximate, or general quantity of entities such as *ten apples*, *about ten apples*, and *some apples* nor to general references to entities in the world or the universe such as *all known apples*/*every known apple* or *all apples*/*every apple* or multiple references to entities such as *a different apple* and *each apple*. In the current system, the entities in the socially construed situation are the only potential referents and a single reference can be realised by a nominal group.

Finally, the socially construed situation was bounded by the walls of an apartment. Everything outside the apartment was not considered present. When an intelligent wheelchair is used in practice, users will need to talk about entities in a larger situation such as the building where the apartment is in, the city district, and the city. Expected entities will include the nearest supermarket, pharmacy, post office, and so on. A larger situation will definitely add new challenges that are not covered in this thesis.

Bibliography

Barbara Abbott. *Reference*. Oxford University, Oxford, 2010.

Jan Alexandersson, Bianka Buschbeck-Wolf, Tsutomu Fujinami, Elisabeth Maier, Norbert Reithinger, Birte Schmitz, and Melanie Siegel. Dialog Acts in VERBMOBIL-2. Technical report, May 1997.

James F. Allen and Mark G. Core. Draft of DAMSL: Dialog Act Markup in Several Layers. 1997.

Dimitra Anastasiou. Gestures in assisted living environments. In *Proceedings of the 9th International Gesture Workshop (GW '11)*, pages 6–9, Athens, May 2011. Springer Berlin Heidelberg.

Dimitra Anastasiou, Desislava Zhekova, and Cui Jian. Speech and gesture interaction in an ambient assisted living lab. In *Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments (SMIAE '12)*, pages 18–27, Jeju, July 2012.

Hannah Arendt. *The Human Condition*. University of Chicago, Chicago IL, 1958.

John L. Austin. Ifs and Cans. *Proceedings of the British Academy*, 42:107–132, 1956.

John L. Austin. *How to do things with words*. Oxford University, London, 1962.

John A. Bateman. Systemic-functional linguistics and the notion of linguistic structure: unanswered questions, new possibilities. In Jonathan J. Webster, editor, *Meaning in context implementing intelligent applications of language studies*, pages 24–58. London/New York, 2008.

John A. Bateman, Joana Hois, Thora Tenbrink, and Robert J. Ross. Technical Report – GUM-Space. Technical report, University Bremen SFB/TR8 Spatial Cognition, Bremen, 2009.

Cem Bozşahin, Geert-Jan Kruijff, and Michael White. Specifying Grammars for OpenCCG: A Rough Guide, March 2005. URL `https://goo.gl/6uzTIS`.

Philippe Bretier and M. David Sadek. A rational agent as the kernel of a cooperative spoken dialogue system: implementing a logical theory of interaction. In *Proceedings of the Workshop on Intelligent Agents III, Agent Theories, Architectures, and Languages (ATAL) of the 7th European Conference on Artificial Intelligence (ECAI 96) in Budapest*, pages 189–203, Berlin/Heidelberg, 1997. Springer-Verlag.

Mark G. Core. Analyzing and predicting patterns of DAMSL utterance tags. In *Proceedings of the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, pages 18–24, Stanford, CA, March 1998.

Mark G. Core and James F. Allen. Coding dialogs with the DAMSL annotation scheme. In *Proceedings of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, Cambridge MA, November 1997. University of Rochester.

Daniel Couto-Vale and Arndt Heilmann. Scalable ecologically valid translation experiment design. In *Proceedings of TRA&CO Symposium*, Germersheim, April 2016.

Daniel Couto-Vale and Vivien Mast. Using Foot-Syllable Grammars to Customize Speech Recognizers for Dialogue Systems. In *TSD '12 Lecture Notes in Artificial Intelligence vol. 7499*, Brno, Czech Republic, 2012. Springer.

Daniel Couto-Vale, Elisa Vales, and Rumiya Izgalieva. Towards a Description of Symbolic Maps. In *Proceedings of the Eighth International Natural Language Generation Conference*, pages 83–92, Philadelphia, 2014.

Scott Farrar, Thora Tenbrink, John A. Bateman, and Robert J. Ross. On the role of conceptual and linguistic ontologies in spoken dialogue systems. In *Proceedings of the Symposium on Dialogue Modelling and Generation*, pages 1–12, June 2005.

John R. Firth. The technique of semantics. *Transactions of the Philological Society*, 35:36–72, 1936.

John R. Firth. Personality and language in society. *Sociological Review*, 42: 37–52, 1950.

John R. Firth. Modes of meaning. In Geoffrey Tillotson, editor, *Essays and studies 1951: being the volume four of the essays and studies collected for the English Association*, pages 118–149. John Murray, London, 1951a.

John R. Firth. General linguistics and descriptive grammar. *Transactions of the Philological Society*, 50:69–87, 1951b.

John R. Firth. Ethnographic analysis and language with reference to Malinowski's views. In Raymond Firth, editor, *Man and culture: an evaluation of the work of Bronisław Malinowski*, pages 93–118. Routledge and Kegan Paul, London, 1957a.

John R. Firth. *Papers in linguistics 1934-51.* Oxford University, London, 1957b.

Birte Glimm, Ian Horrocks, Boris Motik, Giorgos Stoilos, and Zhe Wang. HermiT: An OWL 2 Reasoner. *Journal of Automated Reasoning*, 53(245), 2014.

Herbert Paul Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics*, pages 41–58. New York, NY, 1975.

Michael A.K. Halliday. Grammar, society, and the noun (1966). In Jonathan J. Webster, editor, *On Language and Linguistics*. 1966a.

Michael A.K. Halliday. Lexis as a linguistic level (1966). In Jonathan J. Webster, editor, *On Grammar*, pages 158–172. Continuum, London, 1966b.

Michael A.K. Halliday. Spoken and written modes of meaning (1987). In Jonathan J. Webster, editor, *On Grammar*, pages 323–351. Continuum, London, 1987.

Michael A.K. Halliday. Language and the order of nature (1987). In Jonathan J. Webster, editor, *On Language and Linguistics*, pages 116–138. 2003.

Michael A.K. Halliday. Categories of the theory of grammar (1961). In Jonathan J. Webster, editor, *On Grammar*, pages 37–94. Continuum, London, 2005.

Michael A.K. Halliday and Christian M.I.M. Matthiessen. *Construing experience through meaning: a language-based approach to cognition*. Continuum, London/New York, 1999.

Michael A.K. Halliday and Christian M.I.M. Matthiessen. *Halliday's Introduction to Functional Grammar*. Routledge, London/New York, 4 edition, 2014.

Ruqaiya Hasan. The grammarian's dream: lexis as most delicate grammar. In Michael A.K. Halliday and Robin P. Fawcett, editors, *New developments in systemic linguistics theory and description*, pages 184–211. London, 1987.

Geoffrey Horrocks. *Generative Gramamr*. Routledge, New York, 2013.

Gail Jefferson. Side sequences. In David N. Sudnow, editor, *Studies in Social Interaction*, pages 294–338. The Free Press, New York, 1972.

Gail Jefferson. Notes on some orderlinesses of overlap onset. In V. D'Urso and P. Leonardi, editors, *Discourse Analysis and Natural Rhetoric*, pages 11–38. Language and Literature, Padua, Italy, 1984.

Gail Jefferson. Notes on 'latency' in overlap onset. *Human Studies*, 9:153–183, 1986.

Susanne Jekat, Alexandra Klein, Elisabeth Maier, Ilona Maleck, Marion Mast, and J Joachim Quantz. Dialogue Acts in VERBMOBIL. Technical report, 1995.

Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, August 1997. URL `http://www.stanford.edu/ jurafsky/manual.august1.html`.

Bronisław Malinowski. The problem of meaning in primitive languages. In *The meaning of meaning*, pages 1–386. February 1923.

Bronisław Malinowski. *Argonauts of the western pacific*. George Routledge & Sons, London, 1932.

Bronisław Malinowski. *Coral gardens and their magic*, volume 2. George Allen & Unwin, London, 1935.

Bronisław Malinowski. *Magic, Science and Religion and Other Essays*. 1948.

Rui Mao, Chenghua Lin, and Frank Guerin. Word embedding and WordNet based metaphor identification and interpretation. In *Proceedings of the th Annual Meeting of the Association for Computational Linguistics*, pages 1222–1231, Melbourne, Australia, July 2018.

Vivien Mast and Diedrich Wolter. A probabilistic framework for object descriptions in indoor route instructions. In Thora Tenbrink, J. Stell, Antony Galton, and Z. Wood, editors, *Proceedings of the Conference on Spatial Information Theory (COSIT 2013) published under Lecture Notes on Computer Science*, pages 185–204. Springer, Scarborough, September 2013a.

Vivien Mast and Diedrich Wolter. Context and vagueness in REG. In *Proceedings of the Workshop on Referring Expression Generation "Bridging the gap between cognitive and computational approaches to reference" at the Annual Meeting of the Cognitive Science Society (CogSci 2013)*, Berlin, July 2013b.

Vivien Mast, Cui Jian, and Desislava Zhekova. Elaborate descriptive information in indoor route instructions. In *Proceedings of the th annual conference of the cognitive science society*, Austin, 2012.

Vivien Mast, Daniel Couto-Vale, and Zoe Falomir. Enabling grounding dialogues through probabilistic reference handling. In *Proceedings of the RefNet Workshop on Psychological and Computational Models of Reference Comprehension and Production*, Edinburgh, August 2014a.

Vivien Mast, Daniel Couto-Vale, Zoe Falomir, and Fazleh Elahi. Referential grounding for situated human-robot communication. In *Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue (DialWatt/SemDial 2014)*, Edinburgh, August 2014b.

George A Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

Nuance VoCon 3200. *Nuance VoCon 3200 Embedded Development System - Developer's Guide.*

William O'Grady. The Syntax of Idioms. 16:279–312, May 1998.

Petya Osenova and Kiril Simov. Treatment of multiword expressions and compounds in Bulgarian. pages 1–6, May 2014.

Bijan Parsia, Peter Patel-Schneider, and Boris Motik. OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition). Technical report, December 2012.

Steven Pinker. *The Language Instinct.* Penguin, London, 1995.

Rachel Reichman. *Plain speaking: a theory and grammar of spontaneous discourse.* PhD thesis, Bolt Beranek and Newman Inc Cambridge MA, Harvard University, June 1981.

Robert J. Ross and John A. Bateman. Daisie: Information State Dialogues for Situated Systems. In *Proceedings of the 12th International Conference on Text, Speech and Dialogue*, pages 379–386, Pilsen, September 2009.

Robert J. Ross, Hui Shi, Thora Tenbrink, and John A. Bateman. Modelling illocutionary structure: combining empirical studies with formal model analysis. In *Proceedings of the International Conference on Computational Linguistics and Intelligent Text Processing (CICLing '10) in Iaşi, Romania, published under Lecture Notes on Computer Science (LNCS 6008)*, pages 340–353. Berlin, March 2010.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735, 1974.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn taking for conversation*. In *Studies in the organization of conversational interaction*, pages 7–55. Academic Press, New York • San Francisco • London, 1978.

M. David Sadek. Communication Theory = Rationality Principles + Communicative Act Models. In *Proceedings of the Workshop on Planning for Interagent Communication of the 12th National Conference on Artificial Intelligence sponsored by the Association for the Advancement of Artificial Intelligence (AAAI 94)*, page not numbered, Seattle WA, July 1994.

M. David Sadek. Design considerations on dialogue systems. In *Proceedings of the Workshop on Interactive Dialoge in Multi-Modal Systems (IDS 99) organised by the European Speech Communicaiton Association (ESCA)*, pages 1–15, Kloster Irsee, June 1999.

M. David Sadek, A. Ferrieux, A. Cozannet, Philippe Bretier, Frank Panaget, and Jacky Simonin. Effective human-computer cooperative spoken dialogue: the AGS demonstrator. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, pages 546–549 Vol. 1, Philadelphia PA, October 1996. IEEE.

M. David Sadek, Philippe Bretier, and Frank Panaget. Artimis: Natural dialogue meets rational agency. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence (JCAI 97)*, pages 1030–1035 Vol. 1, Nagoya, August 1997. Morgan Kaufmann.

Emanuel A. Schegloff. Sequencing in conversational openings. *American Anthropologist*, 70(6):1075–1095, 1968.

Emanuel A. Schegloff. The relevance of repair to syntax-for-conversation. *Syntax and Semantics 12: Discourse and Syntax*, 12:261–286, 1979.

Emanuel A. Schegloff. Preliminaries to preliminaries: "can I ask you a question?". *Sociological Inquiry*, 50:104–152, 1980.

Emanuel A. Schegloff. The routine as achievement*. *Human Studies*, 9:111–151, 1986.

Emanuel A. Schegloff. Presequences and indirection: applying speech act theory to ordinary conversation. *Journal of Pragmatics*, 12:55–62, 1988.

Emanuel A. Schegloff. To Searle on conversation: a note in return. In *On Searle on Conversation*, pages 113–128. 1992a.

Emanuel A. Schegloff. Repair after next turn: the last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5):1295–1345, 1992b.

Emanuel A. Schegloff. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29:1–63, 2000a.

Emanuel A. Schegloff. When "Others" Initiate Repair. *Applied Linguistics*, 21 (2):205–243, 2000b.

Emanuel A. Schegloff. Accounts of Conduct in Interaction: interruption, overlap, and turn-taking. In Jonathan H. Turner, editor, *Handbook of Sociological Theory*, pages 287–321. London, 2001.

Emanuel A. Schegloff. On the organization of sequences as a source of "coherence" in talk-in-interaction. In Bruce Dorval, editor, *Conversational Organization and its Development*, pages 51–77. Nordwood, 2002a.

Emanuel A. Schegloff. Recycled turn beginnings: a precise repair mechanism in conversation's turn-taking organisation. In Graham Button and John R.E. Lee, editors, *Talk and social organisation*, pages 70–85. Clevedon, PA, 2002b.

Emanuel A. Schegloff. Experimentation or observation? Of the self alone or the natural world? *Behavioral and Brain Sciences*, 27(2):171–172, 2004.

Emanuel A. Schegloff. Word repeats as unit ends. *Discourse Studies*, 13(3): 367–380, May 2011.

Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. The preference for self-correction in the organisation of repair in conversation. *Language*, 53(2): 361–382, 1977.

Manfred Schmidt-Schauß and Gert Smolka. Attributive Concept Descriptions with Complements. 48:1–26, February 1991.

John R. Searle. What is a speech act? In Maurice Black, editor, *Philosophy in America*, pages 221–239. London, 1965.

John R. Searle. Indirect Speech Acts. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics*, pages 59–82. New York, NY, 1975a.

John R. Searle. A taxonomy of illocutionary acts. In Keith Gunderson, editor, *Mind and Knowledge*, pages 344–369. University of Minnesota, Minneapolis, 1975b.

Hui Shi and Thora Tenbrink. Telling Rolland where to go. In Kenny R. Coventry, Thora Tenbrink, and John A. Bateman, editors, *Spatial Language and Dialogue*, pages 177–190. Oxford University Press, Oxford, 2009.

Luigi Squillante. Towards an empirical subcategorization of multiword expressions. pages 1–5, 2014.

Stanford NLP Group. Stanford Parser, July 2012. URL `http://nlp.stanford.edu:8080/parser/`.

Statistisches Bundesamt. Zensus 2011: Ausgewählte Ergebnisse, May 2013.

Mark Steedman. A very short introduction to CCG, 1996. URL `http://www.inf.ed.ac.uk/teaching/courses/ics/papers/ccgintro.pdf`.

Mark Steedman. Where does compositionality come from? In *Proceedings of the Workshop on Compositional Connectionism in Cognitive Science at the IAAA 2004 Fall Symposium*, Washington D.C., October 2004.

Thora Tenbrink and Hui Shi. Negotiating spatial goals with a wheelchair. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 103–110, Antwerp, September 2007.

Thora Tenbrink, Elena Andonova, and Kenny R. Coventry. Negotiating spatial relationships in dialogue: The role of the addressee. In *Proceedings of the 12th SEMDIAL workshop (LONDIAL)*, pages 1–8, London, June 2008.

Thora Tenbrink, Robert J. Ross, Kavita E. Thomas, Nina Dethlefs, and Elena Andonova. Route instructions in map-based human–human and human–computer dialogue: A comparative analysis. *Journal of Visual Languages & Computing*, 21(5):292–309, 2010.

Paul J Thibault and Theo Van Leeuwen. Grammar, society, and the speech act: Renewing the connections. *Journal of Pragmatics*, 25(4):561–585, 1996.

David R. Traum and James F. Allen. Discourse obligations in dialogue processing. In *Proceedings of 32the nd annual meeting on Association for Computational Linguistics*, Las Cruces, NM, June 1994.

Amy B. Tsui. A taxonomy of discourse acts. In *English conversation*, pages 44–61. Oxford, 1994.

Elisa Vales. *Users' view of the robot in HRI: robot-as-companion or robot-as-tool*. PhD thesis, University Bremen, Bremen, December 2014.

Constanze Vorwerg and Thora Tenbrink. Discourse factors influencing spatial descriptions in English and German. *Spatial Cognition*, 5:470–488, 2007.

Hendrik Zender, Geert-Jan Kruijff, and Ivana Kruijff-Korbayová. Situated Resolution and Generation of Spatial Referring Expressions for Robotic Assistants. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 1604–1609, Pasadena, May 2009.

Index