D6.2: Annotators Handbook

Nate Blaylock, Bettina Fromkorth, Ciprian Gerstenberger, Ivana Kruijff-Korbayová (ed.), Oliver Lemon, Pilar Manchón, Anja Moos, Verena Rieser, Carmen del Solar, Karl Weilhammer

Distribution: Public

TALK
Talk and Look: Tools for Ambient Linguistic Knowledge
IST-507802 Deliverable 6.2
February 13, 2006

Project funded by the European Community under the Sixth Framework Programme for Research and Technological Development

The deliverable identification sheet is to be found on the reverse of this page.
<table>
<thead>
<tr>
<th>Project ref. no.</th>
<th>IST-507802</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project acronym</td>
<td>TALK</td>
</tr>
<tr>
<td>Project full title</td>
<td>Talk and Look: Tools for Ambient Linguistic Knowledge</td>
</tr>
<tr>
<td>Instrument</td>
<td>STREP</td>
</tr>
<tr>
<td>Thematic Priority</td>
<td>Information Society Technologies</td>
</tr>
<tr>
<td>Start date / duration</td>
<td>01 January 2004 / 36 Months</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Security</th>
<th>Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contractual date of delivery</td>
<td>M24 = December 2005</td>
</tr>
<tr>
<td>Actual date of delivery</td>
<td>February 13, 2006</td>
</tr>
<tr>
<td>Deliverable number</td>
<td>6.2</td>
</tr>
<tr>
<td>Deliverable title</td>
<td>D6.2: Annotators Handbook</td>
</tr>
<tr>
<td>Type</td>
<td>Report</td>
</tr>
<tr>
<td>Status &amp; version</td>
<td>Final version February 13, 2006</td>
</tr>
<tr>
<td>Number of pages</td>
<td>112 (excluding front matter)</td>
</tr>
<tr>
<td>Contributing WP</td>
<td>USAAR</td>
</tr>
<tr>
<td>WP/Task responsible</td>
<td>USE, UCAM, UEDIN</td>
</tr>
<tr>
<td>Other contributors</td>
<td>Nate Blaylock, Bettina Fromkorth, Ciprian Gerstenberger, Ivana Kruijff-Korbayová (ed.), Oliver Lemon, Pilar Manchón, Anja Moos, Verena Rieser, Carmen del Solar, Karl Weilhammer</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Evangelia Markidou (Anne Bajart)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EC Project Officer</th>
<th>The TALK Project Co-ordinator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prof. Manfred Pinkal</td>
</tr>
<tr>
<td></td>
<td>Computerlinguistik</td>
</tr>
<tr>
<td></td>
<td>Fachrichtung 4.7 Allgemeine Linguistik</td>
</tr>
<tr>
<td></td>
<td>Postfach 15 11 50</td>
</tr>
<tr>
<td></td>
<td>66041 Saarbrücken, Germany</td>
</tr>
<tr>
<td></td>
<td><a href="mailto:pinkal@coli.uni-sb.de">pinkal@coli.uni-sb.de</a></td>
</tr>
<tr>
<td></td>
<td>Phone +49 (681) 302-4343 - Fax +49 (681) 302-4351</td>
</tr>
</tbody>
</table>

The partners in TALK are:  
Saarland University  
University of Edinburgh HCRC  
University of Gothenburg  
University of Cambridge  
University of Seville  
Deutches Forschungszentrum fur Künstliche Intelligenz  
Linguamatics  
BMW Forschung und Technik GmbH  
Robert Bosch GmbH

Keywords: transcription standard, corpus annotation, annotation guidelines

For copies of reports, updates on project activities and other TALK-related information, contact:

The TALK Project Co-ordinator  
Prof. Manfred Pinkal  
Computerlinguistik  
Fachrichtung 4.7 Allgemeine Linguistik  
Postfach 15 11 50  
66041 Saarbrücken, Germany  
pinkal@coli.uni-sb.de  
Phone +49 (681) 302-4343 - Fax +49 (681) 302-4351
Copies of reports and other material can also be accessed via the project’s administration homepage, http://www.talk-project.org

©2006, The Individual Authors.

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>1</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>2 Corpora for Annotation</strong></td>
<td>3</td>
</tr>
<tr>
<td>2.1 USAAR/DFKI</td>
<td>3</td>
</tr>
<tr>
<td>2.1.1 SAMMIE-1 Corpus</td>
<td>3</td>
</tr>
<tr>
<td>2.1.2 SAMMIE-2 Corpus</td>
<td>3</td>
</tr>
<tr>
<td>2.2 UCAM</td>
<td>3</td>
</tr>
<tr>
<td>2.2.1 SACTI-1 Corpus</td>
<td>4</td>
</tr>
<tr>
<td>2.2.2 SACTI-2 Corpus</td>
<td>6</td>
</tr>
<tr>
<td>2.3 USE: The MIMUS Corpus of Multimodal WOz Experiments</td>
<td>6</td>
</tr>
<tr>
<td>2.3.1 Subjects</td>
<td>6</td>
</tr>
<tr>
<td>2.3.2 Tasks</td>
<td>7</td>
</tr>
<tr>
<td>2.3.3 Procedure</td>
<td>7</td>
</tr>
<tr>
<td>2.3.4 User–Wizard Interactions</td>
<td>9</td>
</tr>
<tr>
<td>2.3.5 Data</td>
<td>9</td>
</tr>
<tr>
<td><strong>3 Transcription Standards</strong></td>
<td>12</td>
</tr>
<tr>
<td>3.1 English Transcription Standards</td>
<td>12</td>
</tr>
<tr>
<td>3.2 German Transcription Standards</td>
<td>16</td>
</tr>
<tr>
<td>3.2.1 Structure</td>
<td>16</td>
</tr>
<tr>
<td>3.2.2 Overlapping Speech</td>
<td>17</td>
</tr>
<tr>
<td>3.2.3 Orthography</td>
<td>17</td>
</tr>
<tr>
<td>3.2.4 Punctuation</td>
<td>17</td>
</tr>
<tr>
<td>3.2.5 Symbols</td>
<td>17</td>
</tr>
<tr>
<td>3.2.6 Articulatory Noises</td>
<td>21</td>
</tr>
<tr>
<td>3.3 Spanish Transcription Standards</td>
<td>21</td>
</tr>
<tr>
<td>3.3.1 Spelling</td>
<td>21</td>
</tr>
<tr>
<td>3.3.2 Orthography</td>
<td>21</td>
</tr>
<tr>
<td><strong>4 Annotation Standards</strong></td>
<td>25</td>
</tr>
</tbody>
</table>
4.1 Task-completion annotation [UCAM] ................................................. 25
   4.1.1 Annotation Purpose ....................................................... 25
   4.1.2 Annotation Instructions ............................................... 26
4.2 Wizard Understanding Status [UCAM] ............................................ 27
   4.2.1 Annotation Purpose ..................................................... 27
   4.2.2 Annotation Instructions ............................................... 27
4.3 Grounding Act Annotation [UCAM] ............................................... 27
   4.3.1 Annotation Purpose ..................................................... 27
   4.3.2 Annotation Instructions ............................................... 27
4.4 Annotation Standards for Multimodal Natural Command Language Dialogues (NCLDs) [USE] ................................................................. 32
   4.4.1 Annotation Purpose ..................................................... 32
   4.4.2 Annotation Levels and Annotation Instructions ...................... 35
4.5 Domain object names and database search results [USAAR] ................. 45
   4.5.1 Annotation Purpose ..................................................... 45
   4.5.2 Annotation Instructions ............................................... 45
4.6 Nativized Pronunciation of Foreign Words [USAAR] ............................ 46
   4.6.1 Annotation Purpose ..................................................... 46
   4.6.2 Annotation Instructions ............................................... 46
4.7 Syntactic Features [USAAR] .................................................... 53
   4.7.1 Annotation Purpose ..................................................... 53
   4.7.2 Annotation Instructions ............................................... 53
4.8 Discourse-Entity Features [USAAR] .............................................. 69
   4.8.1 Annotation Purpose ..................................................... 69
   4.8.2 Annotation Instructions ............................................... 69
4.9 Clarification Requests [USAAR] .................................................. 89
   4.9.1 Annotation Purpose ..................................................... 89
   4.9.2 Annotation Instructions ............................................... 89
4.10 Task-Layer Features [USAAR] ................................................... 102
   4.10.1 Annotation Purpose .................................................... 102
   4.10.2 Annotation Instructions .............................................. 102
Summary

This document describes the standards for the human annotation and transcription of the corpus data used in the TALK project for research in the various work packages. We only included human annotation standards; automatically-derived annotations will be described in Deliverable 6.5 “Annotated Data Archive”.

Since corpus annotations serve the purpose of investigations for specific research purposes in the individual work packages, we have defined several annotation standards which are particular to individual research issues. In order to maximize reuse, however, we have (a) defined a common transcription standard for each language (English, German and Spanish); and (b) defined a common data representation format to ensure that all of the annotated data is globally storable and sharable: the format is based on the Nite XML Toolkit [CEH+03]. As the data representation applies to all project data, not just that which is human annotated, it will be described in detail in Deliverable 6.5.

Besides the transcription standards for English, German and Spanish, this document includes annotation guidelines for the following annotation tasks: task-completion annotation, wizard understanding status, grounding act annotation for the SACTI corpus, annotation of multimodal natural command language dialogues in the MIMUS corpus; and the annotation of domain object names and database search results, nativized pronunciation of foreign words, syntactic features, discourse-entity features, clarification requests and task-layer features in the SAMMIE-1 and/or SAMMIE-2 corpus.
Chapter 1

Introduction

This document describes the human annotation and transcription standards used in the TALK project for research in the various work packages. As this deliverable is aimed towards human annotation standards, we do not include data annotations made by automatic means (e.g., recording information states in a dialogue system). Automatically-derived annotations will be described in Deliverable 6.5 “Annotated Data Archive”.

Unlike other tasks which are fairly tightly defined, annotation is very open-ended. In fact, what is annotated is often closely associated with the particular experimental question being asked. For this reason, standards often do not exist for a particular feature that one wants to annotate. In such cases, a standard is produced in conjunction with the annotation.

Since investigating individual research questions requires different annotation schemes, it is not appropriate that the TALK project define and enforce a common annotation standard with which all partner data is annotated. Such an approach is impractical and would waste project resources, since annotations are oftentimes useless unless performed on data collected for a particular research issue. Hence we have defined several annotation standards which are particular to individual research issues. In order to maximize reuse, however, we have (a) defined a common data representation to ensure that all of the annotated data is globally storable and sharable; and (b) defined a common transcription standard for each language. As the data representation applies to all project data, not just that which is human annotated, it is described in detail in Deliverable 6.5.

The structure of the remainder of this document is as follows: In Chapter 2, we briefly describe the corpora we use for annotation. In Chapter 3, we detail transcription standards for each language used. Finally, in Chapter 4, we describe each type of annotation and instructions for annotation using the standard.
Chapter 2
Corpora for Annotation

In this chapter, we briefly describe the data which has been, or is to be annotated using the standards in this document.

2.1 USAAR/DFKI

USAAR and DFKI have jointly produced two Wizard-of-Oz corpora in an MP3 domain, in the SAMMIE (Saarbrücken Multimodal MP3 Player Interaction Experiment) series of experiments.

2.1.1 SAMMIE-1 Corpus

The SAMMIE-1 corpus was gathered as a monomodal (speech) Wizard-of-Oz experiment in German. Human wizards were given access to a large database of information about music albums and an MP3 player tool for creating and manipulating playlists and simulating music playback. Subjects were given a series of tasks to perform using the MP3 player. More information about the SAMMIE-1 corpus can be found in Deliverable 6.1 [BLY+04].

2.1.2 SAMMIE-2 Corpus

The SAMMIE-2 corpus extended the SAMMIE-1 setup to a multimodal setting. The wizard was also given access to facilities of displaying tables or text on the subject’s screen, which were computed given the search context. More information on the SAMMIE-2 corpus can be found in [KKBG+05].

2.2 UCAM

UCAM has produced three SACTI data collections (Simulated ASR-Channel – Tourist Information), which are distributed in two corpora SACTI-1 and SACTI-2. The first two collections are speech-only and are both part of the SACTI-1 corpus. In the SACTI-2 corpus the participants could interact via speech and a multi-modal map interface.
Both corpora were conducted in the tourist information domain, which is part of the in-car scenario. The users were given a map and asked to perform a number of tasks such as finding a hotel within a particular price range, or finding a restaurant of a particular type. An example map is shown in figure 2.1.

![Map](image)

Figure 2.1: The map used for the tourist information data collection (SACTI)

The main features of SACTI-1 and SACTI-2 are compared in Table 2.1. Only manual annotations are described in the Annotators Handbook.

### 2.2.1 SACTI-1 Corpus

The SACTI-1 corpus contains human-human dialogues. The major part of them was recorded in a "simulated automated speech recognition (ASR) channel" [SWY04] [WY04a] and a small portion was recorded as direct conversation. The participants could only communicate via speech. The simulated speech recognition error rate was varied during the recordings from no errors to high. Apart from the logging information of the recording setup, orthographic transcriptions, grounding acts and the understanding status of the wizard have been annotated. For further information on SACTI-1, please consult the respective Manual [WY04b].
<table>
<thead>
<tr>
<th></th>
<th>SACTI-1</th>
<th>SACTI-2</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>speech</td>
<td>speech and interactive map</td>
<td></td>
</tr>
<tr>
<td>Time resolution</td>
<td>1 sec</td>
<td>0.001 - 0.5 sec depending on what was recorded</td>
<td></td>
</tr>
<tr>
<td>speech files</td>
<td>speech segments/turns</td>
<td>speech segments/turns, full dialogue recordings for wizard and user</td>
<td></td>
</tr>
<tr>
<td>Different Tasks</td>
<td>24 standard tasks</td>
<td>24 standard tasks + 6 new tasks</td>
<td></td>
</tr>
<tr>
<td>Error rates</td>
<td>Non, Low, Med, Hi</td>
<td>Non (all tasks)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(all standard tasks)</td>
<td>Med (standard tasks), Hi (new tasks)</td>
<td></td>
</tr>
<tr>
<td>Users per wizard</td>
<td>3 users</td>
<td>6 users</td>
<td></td>
</tr>
<tr>
<td>Corpus contents</td>
<td>states sequence</td>
<td>states sequence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>turns</td>
<td>turns</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>user clicks</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>wizard clicks</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>wizard buttons</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>user endpointed turns</td>
<td>M/A</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>wizard endpointed turns</td>
<td>M/A</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>manually transcribed and segmented wizard turns</td>
<td>M/A</td>
</tr>
<tr>
<td></td>
<td>wizard understanding status</td>
<td>-</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>grounding acts</td>
<td>-</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Task-completion</td>
<td>Task-completion</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Comparison of SACTI-1 and SACTI-2 (Annotation mode: M=manual, A=automatic)
2.2.2  SACTI-2 Corpus

The SACTI-2 corpus [WWY04] contains human-human dialogues in a "simulated automated speech recognition (ASR) channel." The interaction is via speech and an interactive map. A lot of effort has been put in getting a good time resolution for mouse clicks with respect to the speech recordings. The user interface was slightly varied during the recordings. User and wizard behaviour responded to these variations. The simulated speech recognition error rate was no errors for one recording day and medium for all other recording days. Logging information of the recording setup and orthographic transcriptions are available with the corpus. The data of SACTI-2 is intended to be useful for investigating how speech and mouse clicks on a map can be used together in a conversation.

2.3  USE: The MIMUS Corpus of Multimodal WOz Experiments

USE has gathered a Wizard–of–Oz corpus in an in–home domain. The MIMUS (MultIModal, University of Seville) corpus is the result of a multimodal WoZ series of experiments in Spanish. The objective of these experiments is to extend an existing spoken dialogue system by adding new input and output modalities. In order to support this multimodal functionality, the dialogue system will deal with both graphical and spoken input, as well as with a combination of the two:

- “Open the door” (spoken input)
- Click (graphical input)
- “Open this” + Click (multimodal input)

The experiments investigate users’ speech and pen multimodal integration patterns on a system application that controls several types of devices within a home environment: lights, blind, radio, heating, alarm, main door, security camera, and telephone. The interactions between users and the human wizard have been recorded from different perspectives, in an attempt to gather information for the reconfiguration of the basic system.

Other goals include creating a corpus of natural language in the home domain, observing modality preference in relation to task and system familiarity, analysing task completion time and inter–modality timing, and exploring integration patterns used by subjects while performing multimodal multitasking (the ability to carry independent tasks at the same time through different input channels).

2.3.1  Subjects

There are two groups of informants, all of whom can be described as completely naïve subjects. A primary group is formed by 16 wheel–chair bound subjects; a secondary group includes 7 subjects without disabilities. The profile of those within the primary group can be stated in the following terms: subjects are wheel–chair users, have full mobility on upper limbs and no speech impediment. Those within the secondary group respond to the following characteristics: they have only limited (or full) mobility on lower limbs and no speech impediment. The physically impaired users involved in these
experiments belong to two associations for disabled people, namely “FAMS–COCEMFE Sevilla”\(^1\) and “ASPHEBH”\(^2\). Some of the disabilities that they exhibit are: poliomyelitis, paraplegia, sclerosis and spina bifida. Informants’ ages range between 19 and 54 years old. There is a total of 7 women and 16 men among the subjects. All are native speakers of Spanish. They show varying levels of computer expertise.

### 2.3.2 Tasks

Subjects were alone in a room especially prepared for the experiments. There were a number of tools: a Tablet PC (with a touch-screen) and a pen (used to mark graphical input on the system’s GUI), a microphone, speakers, two cameras (a digital camera and a web camera), and the abovementioned devices (connected to X10 modules). Experiments were divided into two parts. In addition to this, Experiment 1 was accomplished in two phases: A and B.

For the initial stage of the first experiment (1.A), naïve users were given just enough information to perform the tasks, but were given not precise instructions as to how to proceed with the system. Task instructions were displayed by the system on the lower part of the computer screen. A total of 10 tasks were delivered in a consecutive way during this part of the experiment. They were rather basic, and focused on providing the user with a general survey of the available functionality (e.g., “TASK 3: You just remembered today is your mum’s birthday and you want to get in touch with her.”). Tasks during the second stage of the first experiment (1.B) were more complex than in the former stage (e.g., “TASK 6: It is getting dark and you just heard a noise in the outside. You want to have some light in the patio and check what is going on there with the security camera.”). There were 12 tasks for this phase.

The main objective for the second stage of the experiment is to analyse the behaviour of a naïve wizard who has just gone through the experiments 1.A and 1.B. This is the reason why users adopted the wizard’s role. Naïve wizards can have preferences as to what they would have liked to hear or see while acting as users; it is worth mentioning that they did not know that the user with whom they were interacting was not naïve. Figure 2.2 shows a screen shot of the GUI used for the experiment.

### 2.3.3 Procedure

Informants were first interviewed about technology-related skills, graphical interfaces, speech technology, and home systems. They were then introduced to the system by an experimenter. Information as to how to perform tasks was given, very general instructions concerning interaction with the system were provided, and questions were answered until the user was ready to interact with the system on his own. Then the experimenter left and the subject was left alone in the room. As far as the subjects were concerned, they interacted with an intelligent multimodal dialogue system and no other human was involved.

Upon tasks completion, users were interviewed about their interaction with the system, modality preferences, tasks, and problems that they might have encountered. After this, subjects were delivered a new set of tasks. The post-experimental survey followed again. It was after this that informants learnt about

---

\(^1\) Federación de Asociaciones de Personas con Discapacidad Física y Orgánica de Sevilla (Seville Association’s Federation of People with Disabilities).

\(^2\) Asociación de Padres de Hijos con Espina Bífida e Hidrocefalia (Parents Association of Children with Spina Bífida and Hydrocephaly).
Figure 2.2: Screen shot of Experiment’s GUI
the dynamics of the Wizard–of–Oz experiment. Only then did they adopt the wizard’s role. For the last part of the experiment, naïve wizards received general guidelines about the functionality of the wizard’s platform. The user was not naïve, but informants did not know. The experiment lasted about one hour and a half per informant. The pictures in Figure 2.3 show different stages of the experiments.

2.3.4 User–Wizard Interactions

The interaction between subject and system was recorded from different perspectives. A digital camera recorded the progression of the experiment. A web camera captured the subjects’ face as they performed tasks. The touch–screen activity was logged.

The wizard was out of sight, but he could hear what users said and see whatever actions they carried out on their touch–screen. The wizard pretended to understand everything, except for a number of simulated recognition errors that had been previously stated in the experiments’ outline. In response to the subject’s actions, the wizard could produce spoken output (via synthetic speech), display messages, play audio and/or video files (to simulate the camera and telephone), and update device states (via X10 commands). Figure 2.4 shows multimodal user–wizard interactions.

2.3.5 Data

This section provides a summary of the WoZ data collection performed at the University of Seville:

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Domain</th>
<th>Input Modes</th>
<th>Output Modes</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIMUS</td>
<td>h-sm³</td>
<td>home</td>
<td>speech+graphics</td>
<td>speech+graphics</td>
<td>Spanish</td>
</tr>
</tbody>
</table>

Annotations are currently being carried out (see chapters 3 and 4 for a description of the manual annotations). These are some of the preliminary data collection statistics:
Figure 2.4: Multimodal User–Wizard Interactions
<table>
<thead>
<tr>
<th>Name</th>
<th>#Tasks</th>
<th>#Users</th>
<th>#Dialogues</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIMUS</td>
<td>32</td>
<td>23</td>
<td>73</td>
</tr>
</tbody>
</table>
Chapter 3

Transcription Standards

In this chapter, we describe the transcription standards used for annotated data. We first describe the standards for English, then for German, and finally for Spanish.

3.1 English Transcription Standards

In general, the goal of transcription is to capture exactly what is said – not to make a speaker’s meaning clear. Thus, transcribers should transcribe exactly what they hear. For example, do not correct grammatical mistakes, do not remove hesitation words, and do not modify what is said to make its meaning clearer. Attention to detail is absolutely crucial!

Use all lower-case letters, even for proper nouns, for individual letters, and for “i”.

Each convention below has been copied directly from the LDC RT-03 manual [LDC03], including LDC section numbers. Use British English spellings. Spellings are checked at www.dictionary.com.

[3.2.1.2] Spelling:

Transcribers use standard orthography, word segmentation and word spelling. All files must be spell-checked after transcription is complete. When in doubt about the spelling of a word or name, annotators consult a standard reference, like an online or paper dictionary, world atlas or news website. Also, please consult the provided materials, including the map and ”Town information” for spellings.

[3.2.1.3] Contractions:

- Annotators limit their use of contractions to those that exist in standard written English, and of course only when a contraction is actually produced by the speaker.

- Annotators must take care to transcribe exactly what the speaker says. The table below, while not comprehensive, shows some examples of how to transcribe common contractions.
<table>
<thead>
<tr>
<th>Complete Form</th>
<th>Spoken As</th>
<th>Transcribed As</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have</td>
<td>I’ve</td>
<td>i’ve</td>
<td></td>
</tr>
<tr>
<td>cannot</td>
<td>can’t</td>
<td>can’t</td>
<td></td>
</tr>
<tr>
<td>will not</td>
<td>won’t</td>
<td>won’t</td>
<td></td>
</tr>
<tr>
<td>you have</td>
<td>you’ve</td>
<td>you’ve</td>
<td></td>
</tr>
<tr>
<td>could not</td>
<td>couldn’t</td>
<td>couldn’t</td>
<td></td>
</tr>
<tr>
<td>should have</td>
<td>should’ve</td>
<td>should’ve</td>
<td>should of, shoulda</td>
</tr>
<tr>
<td>would have</td>
<td>would’ve</td>
<td>would’ve</td>
<td>would of, woulda</td>
</tr>
<tr>
<td>it is</td>
<td>it’s</td>
<td>it’s</td>
<td>its</td>
</tr>
<tr>
<td>its (possessive)</td>
<td>its</td>
<td>its</td>
<td>its</td>
</tr>
<tr>
<td>Marvin (possessive)</td>
<td>Marvin’s</td>
<td>marvin’s</td>
<td></td>
</tr>
<tr>
<td>Marvin is</td>
<td>Marvin’s</td>
<td>marvin’s</td>
<td></td>
</tr>
<tr>
<td>Marvin has</td>
<td>Marvin’s</td>
<td>marvin’s</td>
<td></td>
</tr>
<tr>
<td>going to</td>
<td>gonna</td>
<td>going to</td>
<td>gonna</td>
</tr>
<tr>
<td>want to</td>
<td>wanna</td>
<td>want to</td>
<td>wanna</td>
</tr>
</tbody>
</table>

**Note:** Annotators should take care to avoid the common mistakes of transposing possessive its for contraction it’s (it is), possessive your for the contraction you’re (you are), and their (possessive), they’re (they are) and there.

Annotators should transcribe exactly what they hear using standard orthography. If a speaker uses a contraction, the word is transcribed as contracted: they’re, won’t, isn’t, don’t and so on. If the speaker uses a complete form, the annotator should transcribe what is heard: they are, is not and so on.

For non-standard contractions like ”gonna” and ”wanna” annotators should spellout the entire word: going to, want to.

### [3.2.1.4] Numbers:

All numerals are written out as complete words. Hyphenation is used for numbers between twenty-one and ninety-nine only.

- ”twenty-two”
- ”nineteen ninety-five”
- ”seven thousand two hundred seventy-five”
- ”nineteen oh nine”

### [3.2.1.5] Words and compounds:

- In general, annotators should be conservative about use of hyphens. For instance:
  - ”an overly complicated analysis”
  - not: ”an overly-complicated analysis”
- However, in some cases, a hyphen is required:
Compounds can be tricky. When in doubt, annotators should consult a dictionary and talk to their language team leader.

### [3.2.2.2] Filled pauses and hesitation sounds:

Filled pauses are non-lexemes (non-words) that speakers employ to indicate hesitation or to maintain control of a conversation while thinking of what to say next.

- Each language has a limited set of filled pauses that speakers can employ.
- Annotators use the standardized spellings shown in the table below for filled pauses. The spelling of filled pauses is not altered to reflect how the speaker pronounces the word (e.g., typing AH for a loud ”ah” or ummmm for a long ”um”). For English, this set includes ah, eh, er, uh, um. All filled pauses are indicated with a % sign preceding the word.

<table>
<thead>
<tr>
<th>English Filled Pauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>%ah</td>
</tr>
<tr>
<td>%eh</td>
</tr>
<tr>
<td>%er</td>
</tr>
<tr>
<td>%uh</td>
</tr>
<tr>
<td>%um</td>
</tr>
</tbody>
</table>

### [3.2.2.3] Partial words:

When a speaker breaks off in the middle of the word, annotators transcribe as much of the word as can be made out. A single dash - is used to indicate point at which word was broken off.

- ”yes, absolu- absolutely.”

### [3.2.2.4] Restarts:

Speaker restarts are indicated with double dash ”--”. Annotators use this convention for cases where a speaker stops short, cutting him/herself off before continuing with the utterance.

- ”i thought he -- i thought he was there”
- ”the thi- -- the thing we’re worried about is”

### [3.2.4.1] Hard-to-understand sections:

Sometimes an audio file will contain a section of speech that is difficult or impossible to understand. In these cases, annotators use double parentheses (( )) to mark the region of difficulty.
• Sometimes it is possible to take a guess about the speaker’s words. In these cases, annotators transcribe what they think they hear and surround the stretch of uncertain transcription with double parentheses:
  – "and she told me that ((i should just leave))"

• If an annotator is truly mystified and can’t at all make out what the speaker is saying, she/he uses empty double parentheses to mark the untranscribed region. Where possible, this untranscribed region gets its own timestamp, e.g.:
  – ""(( ))"

[3.2.4.2] Idiosyncratic words:

Occasionally a speaker will make up a new word on the spot. These are not the same as slang words; they’re words that are unique to the speaker in that conversation. If annotators encounter an idiosyncratic word, they should transcribe it to the best of their ability and mark it with an asterisk *. For instance:

• "do you dress like a *schlump yet"
• "why she said *drr i don’t know"

[3.2.4.5] Interjections:

The following standardized spellings are used to transcribe interjections. Interjections do not require any special symbol.

<table>
<thead>
<tr>
<th>English interjections</th>
</tr>
</thead>
<tbody>
<tr>
<td>ach huh-uh oh whew</td>
</tr>
<tr>
<td>duh hm okay whoops</td>
</tr>
<tr>
<td>eee jeepers oof woo-hoo</td>
</tr>
<tr>
<td>ew jeez ooh yay</td>
</tr>
<tr>
<td>ha mm uh-huh yeah</td>
</tr>
<tr>
<td>hee mhm uh-oh yep</td>
</tr>
<tr>
<td>huh nah whoa yup</td>
</tr>
</tbody>
</table>

The following conventions have been added to the LDC conventions. They are intended to extend and be compatible with the LDC conventions.

Self-speech:

When the speaker is clearly talking to him or her-self, and the speech is not directly intended for the listener, transcribe the "self-speech" in the tags "[self-speech] ? [/self-speech]".

• "[self-speech] the castle the castle [/self-speech] right so the castle is on castle loop"
• When it’s not clear whether speech is directed at the listener or not, the preference should be not to use the [self-speech] tags – i.e., When in doubt, leave them out. In general, hesitation words on their own should not be included in [self-speech] tags.

End-pointer errors:

It is possible that the beginning or end of the utterance has speech in which the speaker has been "interrupted" - i.e., the recording started a little after the speaker started, or just before the speaker ended. These interruptions are called "End-pointer errors" and should be noted in two ways.

• First, in addition to the transcription column, there are two other columns, "ep-error-start" and "ep-error-end." These columns are normally both blank. If there is an interruption caused by the end-pointer at the beginning or end of the utterance, note that by entering "true" in the appropriate column.

• Second, the word fragment you hear should be included in the transcription. For words cut off at the beginning of the utterance, put the dash before the word; for word fragments cut off at the end of the utterance, put the dash at the end of the word.

3.2 German Transcription Standards

In this section we provide the transcription conventions for German. First describing the underlying structure to be assumed, it is meant to provide convention on overlapping speech, orthography and punctuation as well as on more notational points stating what symbols to use in order to mark for instance truncation, false starts and mispronounced words. It also provides means to distinguish between different types of articulatory noises if needed. Our conventions generally follow German standards (Duden). When there are exceptions, we mention these explicitly.

3.2.1 Structure

Utterances

One utterance can be thought of as the spoken equivalent of a sentence. Normally, it should involve only one speaker.

Turns

One turn can consist of one or more utterances. A turn is defined as a segment with only one speaker. It may contain longer pauses, as long as the speaker stays the same.

Section

Several turns are grouped together as a section.

---

1 The LDC convention uses <side-speech> ? </side-speech> – however, this causes problems with ANVIL, so it has been changed here.
3.2.2 Overlapping Speech

The convention in the [AC97] is used here. Words of overlapping speech are marked with numbered square brackets with the number in parenthesis next to the right bracket. You can match up overlapping sections by finding all the bracketed text with the same index. Consider the two utterance units below. Here, “das neueste Album von (Nena) hören” and “wie wäre es mit (Somebody Somebody)” are marked as overlapping since they are both are bracketed with number 1 brackets.

T1 utt1: s: ich würde gerne [das neueste Album von (Nena) hören ] (1)
T2 utt2: w:[wie wäre es mit (Somebody Somebody) ] (1)

3.2.3 Orthography

Capitalization

Capitalization should follow the German standards (Duden). Exception: At the beginning of sentences, words are only capitalized if they would be capitalized when found in the middle of a sentence.

Numerals

Numerals are to be written out. Following the standards in the Duden for numerals turns out to be impractical for our purposes. As a convention, numerals between 0 and 99 are to be written in one word. Everything else is to be separated.

Examples:
fü nfundneunzig
ein hundert siebenundsiebzig
sechszehn
zehn Millionen fünfundsechzig tausend drei hundert zwölf

3.2.4 Punctuation

Transcriptions should contain no punctuation. No sentence-end punctuations (period, question mark, etc.) and no commas, quotes and so forth should appear in the transcription.

Exceptions are hyphens which should follow the German standards and should therefore be included in the transcription when required.

3.2.5 Symbols

Reduced Words

Reduced words should be spelled out to their proper form as specified by the German standards. For example in German the spoken form “Ich hab’ drei Alben.” should be transcribed: ich habe drei Alben
Contraction

The spelling of contractions should follow the German standards of the individual languages. If a word has been contracted and its spelling is allowed by those standards, the word should be spelled as specified there. A contracted form is a proper form of a word and hereby allowed (remember: reduced words are to be spelled out).

Example: complete form: spoken as: transcribed as:
in dem im im

Truncation

Truncation is a phenomenon where a speaker starts saying a word but stops before he is finished pronouncing it. This may happen when the speaker changes his mind about what to say, when thinking aloud, or when interrupted by another speaker (see: Overlapping Speech). Only the actually spoken part of the word should be transcribed. The end of the truncation should be marked with an equal sign (=) without inserting a space between the truncation and this sign.

Example:
ich verstehe= könntest du das nochmal sagen

Repeated Words and False Starts

Often in speech, a speaker will repeat words. If a word is repeated several times, the word should be written out as often as it occurred in the signal.

It occurs often that a sentence is started, but in the middle of the sentence the speaker changes his mind about what to say.

Speaker restarts are indicated with double equal sign (==). Use this convention for cases where the speaker stops short, cutting himself off before continuing with the utterance. False starts should be transcribed using truncation if necessary.

Examples:
füge aus == füge in die gerade erstellte Playlist
hinzü== hinzufügen

Mispronounced words

If a word is mispronounced (if for example a slip of tongue occurred), the correct spelling according to the German standards should be given with a plus sign (+) as a prefix hereby marking it as mispronounced.

Example:
ist dann +Interpret mit dem Namen (Billy Joel) dabei
Unintelligible Speech

Unintelligible speech is enclosed in double, round brackets: (( best guess )). If you can’t even make a guess, write double, round brackets (for technical reasons without a space in the middle): (())

Example:
keine Ahnung (( was gibt es noch ))

Individual letters (Spelling out)

A speaker may be purposely telling another one how to spell something. If the speaker is spelling in German, each letter should be capitalized and preceded by a swung dash (˜).

Example:
“Hallo” when spelled out should be transcribed in the following way: ˜H ˜A ˜L ˜L ˜O.

Sometimes if the word in question is from another language it might occur that the speaker spells it in the language this word comes from. Should this happen each letter should be capitalized and preceded by the hash sign (#). The language the foreign spelling is done in should be given using a tag which consists of the language name in angled brackets (〈language〉). This tag should be placed before the transcription of the actual word. For example the spelling of the english word “hello” should be transcribed in the following way:
〈eng> #H #E #L #L #O
(Note: Please compare also to acronyms II where not every letter is preceded by the hash sign but only the first one.)

Acronyms I

Acronyms of group I are pronounced as a single word. They are transcribed with the prefix @ and should be written in caps (without space between the letters). Acronyms of this group should be treated as normal words in the lexicon (with their pronunciation given there).

Example:
@NATO
@AIDS

Acronyms II

Acronyms of group II are pronounced as a sequence of individual letters (spelled out), e.g. ADAC. They should be written in all caps (without spaces) as well. In German acronyms of this group may also be pronounced as individual letters from a foreign language, e.g. FBI (pronounced the english way).

Treatment of the Acronym II group should follow the spelling standards for the individual letters with the exception that the corresponding prefix should only be mentioned once, not in front of every letter.

Example:
〈eng> #FBI
˜ADAC
Self-speech

When the speaker is clearly talking to himself and the speech is not directly intended for the listener, annotate it as follows: [self-speech] ... [self-speech].

Example:

[self-speech] (() (( spinnt er ))) [self-speech]

Idiosyncratic words

Occasionally a speaker will make up a new word on the spot. They are not the same as slang words since they are unique to the speaker in that conversation. If you encounter an idiosyncratic word, transcribe it to the best of your ability and mark it with an asterix (*).

Interjections

Interjections are considered as words in the lexicon since they carry meaning.

For German we distinguish the following ones:
mhm (affirmation)
m’m (negation)
okay
na ja

Non-lexemes

In addition to interjections (which are considered to be words), we also have a set of standardized spellings of hesitation sounds. Every such “non-word” is marked with the percent sign (%) in the transcripts.

%ach
%ah
%ahm
%äh
%ähm
%eh
%ehm
%hm
%hum
%öh
%öhm
%tja
%uh
%uhm
3.2.6 Articulatory Noises

If needed articulatory noises can be transcribed using the following tags:

{swallow}
{throat clearing}
{cough}
{laugh}

Articulatory noises, which cannot be categorized as one of the above cases should be transcribed as {noise}.

3.3 Spanish Transcription Standards

This section outlines the transcription standards to be used by the annotators of the Spanish MIMUS Corpus.

Our transcription standards follow the general orthographic conventions (spelling) for Spanish. Words that usually take initial capital letters in Spanish should take capital letters (such is the case of proper nouns —e.g., Carlos, Alberto, Paloma—, geographical names —e.g., América, España, Jaén—, surnames —e.g., Álvarez, Pantoja, Martínez—, etc.); otherwise, lower–case letters should be used. Spellings can be checked at the RAE Dictionary website.2

The sections below (adapted from the LDC “Spanish Hub 4 (Broadcast) Speech” conventions)[LDC99] contain a set of clearly defined symbols that should be used with items such as proper names, acronyms, partial and/or mispronounced words, interjections, and non–lexemes.

3.3.1 Spelling

Transcribers use established word segmentation and word spelling. When in doubt about the spelling of a word, annotators consult a standard reference, like the aforementioned RAE Dictionary online [RAE99].

3.3.2 Orthography

Capitalization

Transcribers follow the accepted standard way to capitalize words, including words at the beginning of a sentence, proper names, and so on.

Numerals

Annotators should write out numerals, they do not use digits.

\[ \text{e.g. } \text{n} \text{ abril de mil novecientos noventa y tres} \]
\[ \text{e.g. } \text{d} \text{o} \text{s millones quinientos mil d} \text{ó} \text{l} \text{a} \text{res} \]

2http://www.rae.es
Abbreviations
Annotators must write out all abbreviations (except those listed as examples in each language, if any).

Punctuation
The following punctuation marks should be used in the transcripts.

- **Periods** (‘.’) should be added at the end of declarative sentences.
- **Question Marks** (‘¿?’) should be added at the end of interrogative sentences.
- **Commas** (‘,’) are usually added to mark brief pauses within sentences (and as accepted in the standard orthography of Spanish).
- **Colons** (‘:’) are added to draw attention to the incoming part of discourse.
- **Semicolons** (‘;’) should be added to mark longer pauses than those identified by commas (and as accepted in the standard orthography of Spanish).
- **Ellipsis/suspension points** (‘…’) should be used to signal imprecise ends within declarative sentences.

Symbols

- **Acronyms I**: Those that are pronounced as a single word should be written in caps (no spaces) and preceded by an “@” symbol:
  
  e.g.  
  @OTAN

- **Acronyms II**: Those that are normally written as a single word but pronounced as a sequence of individual letters should be written in all caps (no spaces) and preceded by a “$” symbol:
  
  e.g.  
  $CNN

- **Individual Letters**: Those that are pronounced as such should be written in caps and preceded by a “#” symbol:
  
  e.g.  
  La compañía #R #J Raynolds

- **Proper Names**: Both proper names and place names should be marked with a “ˆ” symbol. Personal initials should be written in capital letters, be preceded by the “#” and must not have a period after them unless this marks the end of a sentence. If annotators encounter a proper name phrase, they should mark only those names that are true proper names on their own, that is, names that are not a non–name word, as in the third example below. An exception to this is some phrasal names where the components have lost the sense of being independent non-name words, as in the fourth example below.
• **Partial Words**: They are indicated with a dash (without any spacing between the dash and the word).

e.g.  la telev–

• **Interjections**: Standardized spellings for some of the most common interjections are used here.

<table>
<thead>
<tr>
<th>Some Spanish Interjections</th>
</tr>
</thead>
<tbody>
<tr>
<td>ajá                          ay</td>
</tr>
<tr>
<td>mhm (&quot;yes&quot;)                jo</td>
</tr>
<tr>
<td>hala                        vava</td>
</tr>
</tbody>
</table>

• **Hesitation sounds and filled pauses**: In addition to the interjections above, a set of standardized spellings for the most frequent hesitation sounds that speakers may make during conversation have been defined. These are usually referred to as “filled pauses”. Every filled pause in the transcript is marked with the “%” symbol.

<table>
<thead>
<tr>
<th>Some Spanish Filled Pauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>%ah                              %ey</td>
</tr>
<tr>
<td>%eh                              %pff</td>
</tr>
</tbody>
</table>

• **Idiosyncratic words**: Occasionally, a speaker may make up a new word on the spot. These are words that are unique to the speaker in that conversation. Should annotators encounter an idiosyncratic word, they will consult Spanish language references to transcribe it to the best of their ability. The should mark it with the “*” symbol before the word.

e.g.  *chocolatomanía

**Noises**

In order to account for sound phenomena such as distortion, coughs, breaths, unintelligible speech, words and phrases, etc., we make use of a set of specific brackets.

e.g.  {text} Sound made by the speaker.

<table>
<thead>
<tr>
<th>Use only the sounds listed below</th>
</tr>
</thead>
<tbody>
<tr>
<td>{laugh}</td>
</tr>
<tr>
<td>{cough}</td>
</tr>
<tr>
<td>{sneeze}</td>
</tr>
<tr>
<td>{breath}</td>
</tr>
<tr>
<td>{lipsmack}</td>
</tr>
</tbody>
</table>

Version: February 13, 2006 (Final version) Distribution: Public
Other conventions

The following conventions have been added to the LDC conventions (they are based on the English Transcription Standards proposed in [Wei04]).

- **Self-speech**: When the speaker is clearly talking to him or her-sel, and the speech is not directly intended for the listener, annotators should transcribe the “self-speech” in the tags “[self-speech] ? [\self-speech]”.

- **End-pointer errors**: It is possible that the beginning or end of the utterance has speech in which the speaker has been “interrupted” (i.e., the virtual assistant started to talk in the midst of the speaker’s turn). The word fragment that annotators hear should be included in the transcription. For words cut off at the beginning of the utterance, they should put the dash before the word; for word fragments cut off at the end of the utterance, they should put it at the end of the word.

  e.g.  *Sube la per- ; sí*
Chapter 4

Annotation Standards

In this chapter, we describe the standards for human annotation within the TALK project. Standards for the following annotation tasks are presented:

1. Task-completion annotation [UCAM]
2. Wizard understanding status [UCAM]
3. Grounding act annotation [UCAM]
4. Annotation standards for multimodal natural command language dialogues [USE]
5. Domain object names and database search results [USAAR]
6. Nativized pronunciation of foreign words [USAAR]
7. Syntactic features [USAAR]
8. Discourse-entity features [USAAR]
9. Clarification requests [USAAR]
10. Task-layer features [USAAR]

4.1 Task-completion annotation [UCAM]

4.1.1 Annotation Purpose

In a data collection not all dialogues are completed correctly and not always the user is able to collect all the information necessary to fully complete the task. If such data is to be used for machine learning, it is vital to know how successful a dialogue was.
4.1.2 Annotation Instructions

Each dialogue is annotated with an objective assessment of task completion. Task completion was determined after the end of the session by examining the user’s submitted map and task form - i.e., the content of the dialogues themselves was not considered. The same individual performed all task completion assessments. Task completion has been graded with respect to precision and recall and whether the dialogue was ended by the user or terminated by the experimenter.

Task Recall:

Indicates the amount of information collected with respect to the task without assessing its accuracy.

(3) - All information gathered
(2) - Minor pieces of information missing
(1) - Major pieces of information missing
(0) - Nothing attempted

Task Precision:

Indicates whether the information the user submitted on their map and task is accurate, without assessing what portion of the task it satisfied. Several guidelines that were followed are included in each category.

(3) - All information provided is correct. For example, establishment locations like hotels and bars must be on the correct "block", but may be on the opposite side of the street.
(2) - Most information is correct. For example, establishment locations may be on the "next" block.
(1) - Major inaccuracy. For example, any route description (such as a bus or tram) which goes via an incorrect street
na - Nothing attempted (i.e., Task Recall = 0)

Experimenter Terminated:

In some cases, the subject got stuck, and the experimenter would have to terminate the experiment.

“TRUE” indicates that the experimenter terminated the experiment.
“FALSE” indicates that the user ended the experiment.
4.2 Wizard Understanding Status [UCAM]

4.2.1 Annotation Purpose

In the simulated ASR dialogue recordings, the wizard is presented with corrupted utterances of the user. This annotation provides information on how much of what the user said was understood by the wizard. We based the definition of this annotation on work by Skantze [Ska03].

4.2.2 Annotation Instructions

Each wizard turn is annotated to indicate whether the intention expressed in the previous user turn was understood. Both the recognition result and the wizard’s actions are taken into account when assessing understanding status. If the meaning of the utterance was clearly obscured by the end-pointer, this should be considered in classification of this utterance. The following categories are used:

Full The full intention of the user was understood or inferred. For example, if the wizard seeks to confirm an intention, and that intention is correct, then it should be labelled "Full". The wizard’s response may include a little extra information than what the user requested.

Partial A portion of the user’s intention was understood or inferred, and nothing was misunderstood. Also, if the wizard provided a shotgun response i.e., lots of information which happens to include the requested information this should be annotated as "Partial".

Non No part of the user’s intention was understood. (Individual words may have been communicated by the recognizer, but the wizard could not form a plausible hypothesis based on them.)

Mis At least one aspect of the user’s intention was understood incorrectly. For example, if the wizard seeks to confirm an intention, and that intention isn’t correct, then it should be labelled "Mis". If there were no ASR errors yet the meaning was still misunderstood, that misunderstanding should still be annotated. Finally, if the wizard’s interpretation or reaction shows a misunderstanding, but then goes on to give an answer to the user’s question, that should still be labelled as "Mis". For example, What are the films on at the cinema? / No, there is only one cinema, and it’s playing X, Y, and Z should be labelled "Mis".

NA-Wizard Used when the wizard speaks first in a dialog, or when it cannot be inferred from the data whether the wizard understood the user’s intention.

4.3 Grounding Act Annotation [UCAM]

4.3.1 Annotation Purpose

To study the grounding process at a relatively low level it is necessary to annotate the grounding behaviour of both wizard and user.
4.3.2 Annotation Instructions

Each wizard and user turn is annotated to indicate what grounding behaviors were expressed. There are a cornucopia of schemes for annotating Dialog Acts in spoken conversation. We examined: schemes used in similar experiments by Skantze [Ska03]; Augmented C-STAR [DADH01]; DAMSL [SRC+00], HCRC Map Task Moves [TKIW98], Verbmobil [JKM+95], ATR [Nag92], and Traum [Tra94]. Traum’s Conversation Acts were closest to our needs, and broadly inspired our set. As we were interested in examining actions on a more granular level than Conversations Acts show, we created the following set of Grounding Tags.

Inform

*Description:* Statement or repetition of fact or desire.

*Example(s):*

- U: I’d like to go to the tower.
- W: There are three hotels that are near the main square.
- W: I don’t have any other restaurants.

*Notes:*

- ”No” answers to an offer are an Inform. E.g., ”Do you want to hear about the other restaurants?”
- ”No. Can you tell me about cheap hotels?” is Inform + Request.

Request

*Description:* Question or request (possibly repeated) which invites the other party to make a contribution.

*Example(s):*

- W: What do you want to do?
- U: Does bus one go there?
- U: How much does that cost?
- W: Can I help with anything else?

*Notes:*

- Asking for elaboration on a query, and providing an elaborated query, are both Requests. I.e., suppose agent-1 asks a question, and agent-2 understands the question but needs agent-1 to be more specific so asks for more information, then agent-1 provides a more elaborated query. In this example, all turns are annotated as Request.
- However, if agent-1’s response does not provide an elaborated query but rather provides the answer to agent-2’s question, then agent-1’s response is annotated as Inform.

Greeting/Farewell

*Description:* Conversation opening, closing, and small talk. The wizard’s opening may invite an initial response from the user.
Example(s):
  W: Oh, hi how can I help?
  U: Oh, hi there.
  U: That’s it. End Task.

Notes:
- “Hi” at the beginning of an utterance is annotated as GreetingFarewell
- (The first) “How can I help?” without a “hi” or “hello” is annotated as GreetingFarewell.
- Closing utterances like “That’s great”, “thank you very much” and “you’re welcome - would you like anything else?” and “No thanks” and “Do you have any other questions?” are all annotated as GreetingFarewell.

ExplAck

Description: Explicit show of (perceived!) understanding i.e., “I understand you.”

Example(s):
  W: Right.
  U: Ok.

Notes:
- ”Yes” which answers a yes/no question is not annotated as ExplAck but rather as Inform.
- ”Yes” at the beginning of an utterance which is not responding to a yes/no question is annotated as ExplAck.
- ”No” which is not responding to a yes/no question but which means ”I can’t do that” is not annotated as ExplAck.
  For example, ”Can you tell me about the bar?” / ”Yes, it’s upmarket and rather expensive” - the second turn is annotated as ExplAck + Inform;
  but ”Can you tell me where I can get a pizza in town?” / ”No, there isn’t a pizza restaurant in town” - the second turn is annotated as Inform.
- ”And” at the beginning of an utterance is not annotated as ExplAck.
- ”Thank you” and “you’re welcome” are annotated as ExplAck.

ReqAck

Description: Explicit request for communication of the other party’s understanding.

Example(s):
  W: Does that make sense?
  W: Does that answer your question?
  U: Did that come through ok?

Notes:
• ReqAck is used to annotate “are you there” queries – e.g., in the middle of a dialog, ”Hello?”
• ”Would you like me to repeat that?” is annotated as ReqAck.

UnsolicitedAffirm

*Description:* An unsolicited affirmation that the other party has understood the speaker’s past utterance correctly (i.e., not preceded by ReqAck).

*Example(s):*
  U: Yes, that’s right.
  U: Yes, I want to go to the castle.

RespondAffirm

*Description:* An explicit positive response to ReqAck.

*Example(s):*
  U: Yes, that’s right.
  U: Yes, I want to go to the castle.

RespondNegate

*Description:* An explicit negative response to ReqAck.

*Example(s):*
  U: No, that’s not right.
  U: No.

*Notes:* The sequence ”Would you like me to repeat that” / ”yes” is annotated as ReqAck / RespondNegate.

RejectOther

*Description:* Rejection of the other party’s interpretation of the speaker’s intention without a preceding ReqAck i.e., an uninitiated show that the other party hasn’t understood the speaker.

*Example(s):*
  U: No, that’s not right.
  U: You’ve got it wrong.

DisAck

*Description:* Explicit display of nonunderstanding i.e., show of absence of any contribution from other agent’s last turn.

*Example(s):*
  W: Uhh, I don’t understand.
  W: Nope, didn’t catch that.
Notes:

- "I didn’t understand what you said” and "I don’t follow what you mean” are both annotated as DisAck.
- "What do you mean?” is annotated as DisAck (and not as ReqRepeat).
- When "sorry” implies the intention "I don’t know what you mean/said”, it is annotated as DisAck. For example, "Sorry, can you repeat that?” is annotated as DisAck + ReqRepeat.

StateInterp

Description: A statement of belief or hypothesis in the desire or intention of the other agent.

Example(s):

W: So, you want prices of hotels near the main square.
W: ... prices of hotels near the main square.
W: I guess you want prices of hotels near the main square.

Notes:

- The StateInterp tag may refer to any earlier turn.
- The StateInterp tag may refer to one piece of information – it need not refer to the other party’s complete desire/intention.
- StateInterp is used when the statement of the other agent's desires/intentions is offset from the rest of the speaker’s turn. For example, "Ok, near the main square ... there is one hotel near the main square” is annotated as ExplAck + StateInterp + Inform, whereas "Ok, there is one hotel near the main square” is annotated as ExplAck + Inform.

ReqRepeat

Description: An invitation for the other agent to state their desire / intention again.

Example(s):

W: Can you say that again please.
U: Can you repeat please?

Notes:

- Re-asking the same question is not annotated as ReqRepeat. ReqRepeat is used when one speaker invites the other to say something ”again”, "a second time”, "another time”, etc.
- Requests to "rephrase” are annotated as ReqRepeat. For example, "Could you rephrase that?”
- ReqRepeat refers only to recent items. Requests to repeat more distant information from earlier in the dialog are annotated as Request.
HoldFloor

_description:_ A sound or request intended to keep the floor while not making a contribution.

_example(s):_

W: Uhh, so... just a sec.
U: Oh... wait...

_notes:_

- HoldFloor is used when there is a clear request to wait like “I’m looking” or “just a minute” etc.
- Turns which begin with dysfluencies are not annotated with HoldFloor.
- HoldFloor can be used in a sequence of grounding tags as well as to annotate a whole turn.
- Singing while looking for a piece of information is annotated as HoldFloor.

IncompleteUnknown

_description:_ A complete turn which has no interpretation within the above tags. For example, this tag can be used for extreme end-pointing errors.

_example(s):_

W: Uh, so... [end-pointer then interrupts wizard]
U: [noise] [end-pointer error]

_notes:_ Turns which include just dysfluency are annotated as IncompleteUnknown.

Each turn is annotated as a sequence of one or more of the tags above. Where possible, the surface order of the intentions is preserved. For example:

<table>
<thead>
<tr>
<th>Example Transcript</th>
<th>Grounding tags annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>“So, you want to go to the castle?”</td>
<td>StateInterp + ReqAck</td>
</tr>
<tr>
<td>“So, you want to go from the castle to the tower, let me see you should take bus number one.”</td>
<td>StateInterp + Inform</td>
</tr>
<tr>
<td>“No, I don’t understand - can you say again?”</td>
<td>DisAck + ReqRepeat</td>
</tr>
<tr>
<td>“It says ‘pizza street.’ I don’t think that’s right.”</td>
<td>StateInterp + DisAck Table</td>
</tr>
</tbody>
</table>

Finally, self-corrections are not annotated – i.e., only the corrected portion of a self-correction is annotated.

4.4 Annotation Standards for Multimodal Natural Command Language Dialogues (NCLDs) [USE]

The type of human–computer interaction (HCI) expected in the home scenario of the TALK Project may be described as a Natural Command Language Dialogue (NCLD), in accordance with the typology defined both in the Siridus and in the D’Homme EU Projects. In these dialogues,
The system needs to identify the task and the parameters from the user’s offering, in what the user will expect to be almost a one-shot environment. Follow-up dialogues will be expected to function mostly as disambiguation exercises with respect to the particular task domain of the system. Prime examples of such natural command language dialogues include speech interfaces to personal service devices such as video-recorders, cookers, telephones and personal computers [AQ00].

As far as the MIMUS (MultIModal, University of Seville) Corpus is concerned, two additional facts must be borne in mind:

1. As opposed to other unimodal corpora, the MIMUS Corpus derives from a type of HCI that includes a graphical mode (in addition to the speech modality already present in the D’Homme scenario);

2. Besides, the MIMUS Corpus is the result of an HCI where the person acting as the user in the first part of the experiment (1.A and 1.B) takes the role of the wizard during the second part (2): this means that there is an additional mode (in what concerns the user–system resulting conversation) that must be taken into account in the overall design of a multimodal interaction analysis.

### 4.4.1 Annotation Purpose

As referred in Chapter 2, the main goal of the multimodal WOZ experiments at the University of Seville is to record the interactions between a human user and the wizard. The singularity of these user-system interactions lies on the fact that they go a step beyond those obtained from existing spoken dialogue systems, for they exhibit new input and output modalities. The system thus deals with both graphical and spoken input, as well as with a combination of the two. It also reacts to the user’s moves by producing speech, displaying written messages or images, executing actions, or any combination of the former.

The main points to be drawn from the analysis of the MIMUS Corpus are stated in Manchón, Pérez and Amores [MPA05]:

Naïve subjects will provide reliable data about the first reaction of an untrained user before becoming familiar with the system. At the same time, as the subjects become more familiar with the system, we will learn about efficiency and learnability. The analysis will include, among other issues:

- possible obstacles or difficulties to communicate
- biases that prevent the interaction from being completely natural
- corpus of natural language in the home domain
- modality preference in relation to system familiarity
- modality preference in relation to task
- task completion time
- combination of modalities for one particular task
- inter-modality timing
- multimodal multitasking
It is crucial for the analysis proposed here that the overall outline of the user–system multimodal interaction be understood. It may be summarized as follows:

**Experiments 1.A and 1.B** (the subject is a naïve user, the wizard is a trained expert):  
- **User’s moves**: spontaneous speech + graphical input  
- **Wizard’s moves**: both pre–defined and spontaneous sentences + graphical output

**Experiment 2** (the subject adopts the wizard’s role):  
- **User’s moves**: pre–established commands + graphical input  
- **Wizard’s moves**: spontaneous speech + graphical output

The wizard is a knowledgeable system expert, capable of simulating the system functionality, as well as its constraints. The wizard follows a detailed script that shows not only what the user knows, but also more precise instructions as to how to proceed in each case. Special tools have been developed in order to enable the wizard to simulate the system functionality within a reasonable timespan. The wizard is out of sight but is able both to hear what the subject says and see their touch–screen. Although the user’s input is processed and logged by a speech recognition engine, the wizard pretends to understand everything (within a predefined set of guidelines), excepting a few artificially introduced recognition errors. When producing speech, the wizard uses synthetic speech.

EMMA (the W3C working draft on Extensible MultiModal Annotation markup language) [W3C05] distinguishes two properties for the annotation of input modality:

1. indicating the broader medium or channel (**medium**) and  
2. indicating the specific mode of communication used on that channel (**mode**).

The input medium is defined from the users’ perspective and indicates whether they use their voice (**acoustic**), touch (**tactile**)
\(^1\), or visual appearance/motion (**visual**)
\(^2\) as input.

<table>
<thead>
<tr>
<th>MEDIUM</th>
<th>DEVICE</th>
<th>MODE</th>
<th>RECORDING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoustic</td>
<td>Microphone</td>
<td>Speech</td>
<td>Audiofile</td>
</tr>
<tr>
<td>Tactile</td>
<td>Pen</td>
<td>Graphical User Interface (GUI)</td>
<td>Logging</td>
</tr>
<tr>
<td>Visual</td>
<td>Video Camera</td>
<td>Video</td>
<td>Movie</td>
</tr>
<tr>
<td></td>
<td>Webcam</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The information saved in execution time during the different parts of the experiment (that is, the logging) follows this standard. The goal for the information saved in post–execution time is mainly focused on the user’s behaviour at dialogue level.

ANVIL is the annotation tool used for the transcription process and the encoding of the elements recorded during the experiments. Resulting data will be eventually transformed into the TALK NXT format.

The ANVIL track “UserInput.spoken” will include the manual segmentation and transcription mentioned above. The track “UserInput.graphical” will be generated automatically from the information logged (in execution time) in the XML file “gui_in.xml”. Also automatically loaded are the tracks “GUIOutput.spoken” (from the log “speech_out.xml”) and “GUIOutput.graphical” (from the log “gui_out.xml”).

---

\(^1\) **Tactile** includes most hand–on input device types (pen, mouse, keyboard, and touch screen).  
\(^2\) **Visual** is used for camera input.
4.4.2 Annotation Levels and Annotation Instructions

For the MIMUS Corpus, the following levels of analysis are distinguished:

**Information–Level Annotation**

This section contains information given by the subjects in the questionnaires (before and after the first part of the experiment) and/or annotated by a human annotator after the recording of the experiment (such as the field for “type of disability”). Questionnaires are carried out in an interview fashion; a transcription stage follows in which the annotator listens to the interview, transcribes the answers to the questions and stores the information in XML format.

1. Specification of Recording
   - Date
   - Time
   - Location
   - Person Responsible

2. File Information
   - FolderName: InformantX_Surname1_Surname2
   - VideofileUser: Surname1_Surname2_videoY
   - VideofileWiz: Surname1_Surname2_video2A
   - WebcamfileUser: Surname1_Surname2_webcamZ
   - WebcamfileWiz: Surname1_Surname2_webcam2A

3. Information about the Subjects
   - User/WizardID: Surname1_Surname2
     - Gender
     - Age
     - Nationality
     - Type of Disability
   - Language Competence
     - Mother Tongue
     - Other Languages

4. Answers to Questionnaires
   - Survey Data Collection (before the experiments): Surname1_Surname2_Pre
     - Closed–ended questions
     - Open–ended questions
• Post–experimental Survey (twice, after experiments 1.A and 1.B): Surname1_Surname2_Post1
  Surname1_Surname2_Post2
  – Closed–ended questions
  – Open–ended questions

Task/Subtask–Level Annotation

Annotations within this level include information about the experiments’ specific design: the actual division into tasks and sub–tasks. This information is relevant insofar as it will provide an objective framework for fixing the limits of dialogues within the overall user–system conversation.

Dialogue–Level Annotation

At this level of analysis, annotators undertake two different tasks. In a first stage, speech transcription is done. Since the user’s spoken input is recorded during the experiments (and not logged directly), the annotator engages in a process of manual segmentation and transcription after the recordings. In a second stage, once the user–system conversation is built (by integrating both spoken and graphical turns), with speech and tactile modes showing together on the ANVIL annotations board, encoding of dialogue features start.

XML example code created for ANVIL:

```xml
<?xml version='1.0' encoding='ISO-8859-1'?> <annotation-spec> <head> ... </head> <body> ... <group name="UserInput"> <track-spec name="graphical" type="primary"> <attribute name="tactile" display="true"> </attribute> </track-spec> <track-spec name="spoken" type="primary"> <attribute name="speech" valuetype="String" display="true" /> </track-spec> </group> <group name="GUIOutput"> <track-spec name="graphical" type="primary"> <attribute name="tactile" display="true"> </attribute> </track-spec> <track-spec name="spoken" type="primary"> <attribute name="speech" valuetype="String" display="true" /> </track-spec> </group> ... </body> </annotation-spec>
```

Manual Transcription of the User’s Speech Mode  The transcription guidelines for annotating in Spanish are described in Chapter 3. Annotators follow the video recording and transcribe the user’s spoken input using ANVIL.
Dialogue Moves (DMs) Annotation  For the MIMUS Corpus, dialogue–level annotations concerning the user–system conversation will be assigned in terms of the aforementioned Natural Command Language Dialogues (NCLDs), as defined in [AQ01]. Since it is the broader concept of NCL that encapsulates the present framework of analysis, it seems natural to also employ Dialogue Moves (DMs) in annotating dialogue turns.

Another reason for choosing the NCL approach over Traum and Allen’s [TA94] Conversation Acts is that the former focuses on the internal aspects of dialogue, whereas the latter builds up to a level of common ground that is necessary for communication of beliefs, intentions, and obligations [TA94]. That is, a model built on the grounds of a NCL should be based more on what is said than what is in the minds of the participants when things are said. In other words, it should try to model external aspects of the dialogue rather than the participants’ internal state.

According to Traum and Allen [TA94], Grounding is the process of achieving mutual understanding between participants in a conversation. The “Grounding Acts” approach to discourse analysis assumes a model of dialogue in which a participant’s knowledge is characterised in terms of INFORMATION STATES which are subject to various kinds of updating mechanisms.” [MPT00]. For NCLDs, these updating mechanisms prompt a collaborative behaviour in which the participants are in some sense working together to reach the desired outcome [AQ02].

<table>
<thead>
<tr>
<th>Dialogue Moves (DMs) in NCL</th>
<th>SYSTEM (speech mode and/or GUI mode)</th>
<th>USER (speech mode and/or GUI mode)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Command–oriented DMs</td>
<td>askCommand</td>
<td>specifyCommand</td>
</tr>
<tr>
<td>Parameter–oriented DMs</td>
<td>askParameter</td>
<td>specifyParameter</td>
</tr>
<tr>
<td>Interaction–oriented DMs</td>
<td>askConfirmation</td>
<td>answerYN</td>
</tr>
<tr>
<td></td>
<td>askContinuation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>askRepeat</td>
<td>askHelp</td>
</tr>
<tr>
<td></td>
<td>answerHelp</td>
<td></td>
</tr>
<tr>
<td></td>
<td>errorRecovery</td>
<td></td>
</tr>
<tr>
<td></td>
<td>greet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>quit</td>
<td></td>
</tr>
</tbody>
</table>

Each one of these Dialogue Moves may be defined as follows:

1. Command–Oriented Dialogue Moves
   - **askCommand**: The system requests the user to specify a command or function to be performed.
   - **specifyCommand**: A specific command or function is selected.
   - **informExecution**: The system acknowledges the execution of the task.
2. Parameter–Oriented Dialogue Moves

  **askParameter:** The system asks for the value of a specific parameter.

  **specifyParameter:** The assignment of some value to one parameter.

3. Interaction–Oriented Dialogue Moves

  **askConfirmation:** Once a command has been completed, some situations will require an explicit and/or implicit confirmation.

  **answer YN:** The user replies yes/no.

  **askContinuation:** The system asks for the continuation of the dialogue.

  **askRepeat:** Any of the participants may request the other to repeat the last utterance, or even a specific parameter or command.

  **askHelp:** A petition for help (general, a specific command, or a specific parameter).

  **answerHelp:** The reply to an askHelp move.

  **errorRecovery:** For a situation in which the continuation of the dialogue is impossible.

  **greet:** The usual greeting operation.

  **quit:** The usual closing operation.

XML example code created for ANVIL:

```xml
<?xml version='1.0' encoding='ISO-8859-1'?>
<annotation-spec> 
<head>
  <valuetype-def>
    <valueset name="dialMovType">
      <value-el>askCommand</value-el>
      <value-el>specifyCommand</value-el>
      <value-el>informExecution</value-el>
      <value-el>askParameter</value-el>
      <value-el>specifyParameter</value-el>
      <value-el>askConfirmation</value-el>
      <value-el>answerYN</value-el>
      <value-el>askContinuation</value-el>
      <value-el>askRepeat</value-el>
      <value-el>askHelp</value-el>
      <value-el>answerHelp</value-el>
      <value-el>errorRecovery</value-el>
      <value-el>greet</value-el>
      <value-el>quit</value-el>
    </valueset>
  </valuetype-def>
</head>
</annotation-spec>
```
Sub–dialogue Type Annotation  Contrary to Information–Seeking Dialogues, where the overall task of the dialogue is predefined, NCLDs are based on the idea that the system does not know a priori which function the user may desire to perform. Given this, it is necessary that a number of sub–dialogue types be observed in the analysis. For instance, a task identification stage where the participants aim to form a plan of action would be associated with a deliberation dialogue type. As stated in [AQ02], An important aspect of NLCs is that they exhibit functional embeddings … that occur when the goal of a sub–dialogue shifts to another dialogue type. The following types of sub–dialogues are distinguished within a NCLD:

1. Deliberation dialogue
2. Action–oriented dialogue
3. Information–seeking dialogue
4. Negotiative dialogue

These sub–dialogue types may, in turn, be defined in terms of the functional embeddings that they exhibit.
<table>
<thead>
<tr>
<th>Sub-dialogue Type</th>
<th>Functional Embedding</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliberation Dialogue</td>
<td>Task identification</td>
<td>The participants aim to form a plan of action.</td>
</tr>
<tr>
<td>Action-oriented Dialogue</td>
<td>Command Execution</td>
<td>The user wishes to perform an action</td>
</tr>
<tr>
<td>Information-seeking Dialogue</td>
<td>Consultation</td>
<td>The user wishes to know the state of specific devices or consult the details of another user (phone number, e-mail, etc.)</td>
</tr>
<tr>
<td>Negotiative Dialogue</td>
<td>Clarification</td>
<td>Some parameter required for the successful completion of a command was misunderstood or missing</td>
</tr>
</tbody>
</table>

XML example code created for ANVIL:

```xml
<?xml version='1.0' encoding='ISO-8859-1'?>
<annotation-spec> <head>
  <valuetype-def>
    <valueset name="subdialogueType">
      <value-el color="#eeee00">
        Deliberation
      </value-el>
      <value-el color="#dd0000">
        Action-Oriented
      </value-el>
      <value-el color="#aa0044">
        Information-Seeking
      </value-el>
      <value-el color="#ee8800">
        Negotiative
      </value-el>
    </valueset>
  </valuetype-def>
</head>

Task Identification: The participants aim to form a plan of action.

Command Execution: The user wishes to perform an action.

Consultation: The user wishes to know the state of specific devices or consult the details of another user (phone number, e-mail, etc.).

Clarification: Some parameter required for the successful completion of a command was misunderstood or missing.
```
Gestures–Level Annotation  The ANVIL track “FacialExpressions” is defined according to a closed set of values for the attribute “gestureType”. These have been adapted from the SmartKom Project collection of multimodal data [SSG02]:

1. anger/irritation
2. pondering/reflecting
3. joy/gratification (being successful)
4. surprise
5. helplessness
6. neutral/anything else
7. face partly not visible

For the coding of facial expressions (ultimately aiming at evaluating the user’s behaviour in the interaction with the system), the video used in ANVIL (time–aligned with the video showing the progress of the experiment) will display the user’s face as captured during the execution of the experiment. Both recordings share a single annotations file saved in the *.anvil XML–based format.

These facial expressions convey the impression of certain emotions. Steininger, Schiel and Glesener [SSG02] define the aforementioned categories as follows:

1. **Anger/irritation**: The annotator perceives that *the user is in a negative mood, is visibly not content, is irritated, annoyed, exasperated, angry, disappointed or something similar*. The user may frown, his lips might be pressed together. The shaking of the head may occasionally be observed.

2. **Pondering/reflecting**: The annotator perceives that *the user is thinking hard*. He concentrates on the GUI, reads tasks, searches the display or listens attentively to phone calls or messages. Pondering/reflecting usually shows in eyebrows (knitted), forehead (frowning), lips and eyes.

3. **Joy/gratification**: The annotator perceives that *the user is in a positive mood, enjoys himself, is visibly content, amused or something similar*. The user might laugh or smile. Teeth are usually visible. The corners of the mouth are curved upward.
4. **Surprise**: The annotator perceives that the user is surprised, usually as a reaction to an external stimulus. The eyebrows may move upward and the head might move backwards.

5. **Helplessness**: The annotator perceives that the user is helpless, confused or interrogative. The user might look confused and seems to feel out of control. It usually shows in the upper part of the face (eyebrows, forehead).

6. **Neutral/anything else**: No emotional or cognitive state can be detected by the annotator. The user looks relaxed.

7. **Face partly not visible**: Sequences during which the face is partly occluded by the hands of the user.

The encoding of this track is done with regard to the subjective perception of the annotators. The annotator watches the video showing the user’s face and marks changes in the user’s facial expression.
XML example code created for ANVIL:

```xml
<?xml version='1.0' encoding='ISO-8859-1'?>
<annotation-spec> <head>
  <valuetype-def>
  ...
  </valuetype-def>
</head> <body> ...
<track-spec name="FacialExpressions" type="primary" >
  <attribute name="category" valuetype="gestureType" display="true">
    <value-el>anger/irritation</value-el>
    <value-el>pondering/reflecting</value-el>
    <value-el>joy/gratification</value-el>
    <value-el>surprise</value-el>
    <value-el>helplessness</value-el>
    <value-el>neutral</value-el>
    <value-el>face partly not visible</value-el>
  </attribute>
</track-spec> </body> </annotation-spec>
```

Resulting ANVIL Tracks

1. Track 1: Waveform
2. Track 2: WizardActions (clicks of the wizard)
3. Group:
   (a) Track 3: UserInput.graphical (clicks of the user)
   (b) Track 4: UserInput.spoken (manual segmentation and transcription of user’s speech: hand–annotated)
4. Group:
   (a) Track 5: GUIOutput.graphical (system’s graphical output)
   (b) Track 6: GUIOutput.spoken (endpointed TTS speech)
5. Track 7: DialogueMoves (Dialogue Moves:hand–annotated)
6. **Track 8: Subdialogues** (Types of Subdialogues: hand-annotated)

7. **Track 9: FacialExpressions** (user’s coded facial expressions: hand-annotated)

![Anvil Tracks](image)

**Figure 4.1: Anvil Tracks**
4.5 Domain object names and database search results [USAAR]

4.5.1 Annotation Purpose

In addition to the speech transcriptions described in Section 3.2 we annotated the following domain specific information directly in the transcripts: (i) all names of songs, albums and artists; (ii) utterances that convey the results of a database search. This makes it easier to find the domain object references and the results for various other purposes. It is also useful information in combination with the annotation of discourse entities described in Section 4.8.

4.5.2 Annotation Instructions

**Domain Object Names**

Each album title, artist name, or track title should be surrounded by parentheses “( )”. The parentheses should be separated from the name or title by a space e.g. ( Sunnday Bloody Sunday ).

**Database Search Results**

Those utterances of the wizard which communicate the result of a database search for the first time w.r.t. to a given search request should be annotated with the tag “\result”, as illustrated below:

1. U: Mp3 Player zeige mir das neueste Album von ( Nena )
   MP3 Player show me the latest album by ( Nena )
   W: ich suche jetzt nach diesem Album
      I am now looking for this album
      \result ich habe zwei Treffer aus dem Jahr zwei tausend und drei
      \result I have two hits from the year two thousand and three
   U: zeige mir beide Alben an
      show me both albums
   W: ich zeige dir jetzt eine Liste mit allen ( Nena ) Alben an
      I am shiwing you now a list of all ( Nena ) albums

Here are some additional examples of various ways database search results are communicated:

2. \result ich habe äußerst viele Treffer erhalten
   \result I found a great number of hits
3. \result ich zeige dir nun alle gespeicherten Playlists an
   \result I am now showing you all saved playlists
4. \result ich habe vier Treffer gefunden
   \result I found four hits
5. \result ich habe einen passenden Titel gefunden
   \result I found one result
6. \result leider konnte ich keine Treffer finden
   \result unfortunately I could not find any hits
4.6 Nativized Pronunciation of Foreign Words [USAAR]

USAAR is currently annotating the SAMMIE-1 and SAMMIE-2 corpora with phonetic transcriptions of non-native song, album, and artist names.

4.6.1 Annotation Purpose

As described in Deliverable 5.2 [BBG+05], one challenge for both speech recognition and synthesis in the MP3 domain is the frequent appearance of non-native (mostly English) song, album, and artist names, within German utterances. Annotation of actual pronunciations in the corpora has allowed us to better model pronunciation for the speech recognizer in the baseline system, as detailed in Deliverable 5.2 [BBG+05].

4.6.2 Annotation Instructions

To be able to cope with the challenge mentioned in section 4.6.1 that frequently foreign words are embedded in the German utterances in our corpora, we are annotating these non-native song, album, and artist names on a phonetic basis. For the most part, these foreign words are English. “The Beatles”, e.g., are a well-known group all over the world; so they do of course appear in a German wish list for an MP3 player as well. Also many German artists write songs in English; so it is indispensable for our German dialog system to be able to deal with English words. However, not only English artists or titles appear in our corpora, there are also French words, e.g., the composer “Chopin” or the song “Quand l’amour meurt”. In general, words from any language could appear.

Before going more into detail, we’ll first give you a brief overview of the structure of the annotation work: You first listen to the dialog between subject and MP3 player. If they talk about foreign song, artist, or album names, you transcribe those. To fit the phonetic transcription into the orthographic annotation (e.g., the text transcription), you simply put in the transcription word by word after each foreign utterance. We use ( ) for tagging song, artist, and album names and within those brackets // for tagging the phonetic transcription. You only need to transcribe foreign words. Separation of the phonemes within the // is simply done by putting space characters between the phonemes. This is not a necessity but makes the phonetic transcription much easier to read because some phonemes are displayed by more than one character. A completely annotated utterance could, for instance, look like this:

“Spiel mir das Lied ( One /ˈwʊn/ more /mɔːr/ time /ˈtaɪm/ von /ˈbɹɪtni/ Spears /s p ˈeə s/).”

A detailed description of the phonetic transcription standards is given in the following paragraphs. As a skilled German “IPA-reader”, you might have already recognized in the example above that in our corpora German speakers pronounce foreign words with a mixture of German and foreign phonemes; so

---

3For easier reading, this is transcribed with IPA symbols although we use SAMPA symbols as we will discuss below.

Furthermore you can see that we used // to tag the phonetic transcription although [] would be correct. Usually, // is used for phonemic and [ ] for phonetic labeling. This “misapplication” here is used to facilitate easier transcription.

4The /ɔ/ is only allowed as short vowel in German, not lengthened, and /i:/ is only allowed as long vowel /iː/, not as short vowel.
a German phoneme set would not be sufficient to transcribe the data correctly. Because of this, we expand the German phoneme set to be more global so that it contains all phonemes needed in our corpora. More detailed information about the phoneme set we are using is given in the paragraph “Phoneme Inventory” below.

**Transcription Level** There exist several standards for transcribing speech data, and unfortunately there are terminological discrepancies that can lead to misunderstandings. That is why we will give a short overview of different labeling standards and classify the transcription standard we are using.

A rough classification allows two common ways of transcribing data: labeling on a phonemic or canonical basis or on a phonetic basis. For annotation of words in the way they should be pronounced, phonemic transcription is used. This is the type of transcription found in dictionaries. In order to do this type of transcription, it is not necessary to listen to an actual production of those words; it is sufficient to read them. Therefore, this type of transcription is not adequate for annotating speech; it is better to use phonetic transcription.

With phonetic transcription, it’s possible to annotate phenomena like for example aspiration of plosives or reduction processes within particles like conjunctions or articles. This also covers e.g. phonological processes like voicing or devoicing, elisions, changing or reducing of vowel quality, co-articulation etc. To describe these phenomena that appear in speech, diacritics and suprasegmentals are used in this type of transcription.

To be more precise with what we mean by phonetic labeling, a more detailed description is needed. Barry and Fourcin [BF92] divide this level of labeling into narrow-phonetic and broad-phonetic labeling. The narrow-phonetic transcription is more precise. “…[I]t does differentiate as much as possible with respect to properties modifying the base sound (i.e. with diacritics indicating e.g. voicing/devoicing, nasality, rounding, spreading, etc.).” [BF92, p. 7] So the narrow-phonetic level of annotation is the most accurate for speech because every detail of articulation processes is put down on paper. The broad-phonetic level also is accurate in annotating speech but leaves out the diacritics that indicate modification of the base phoneme. “This is a level which only employs speech sound symbols that have a phonemic status (i.e. they distinguish words in English: e.g. /m/ vs. /n/: map vs. nap […] but uses them to indicate non-phonemic, i.e. continuous-speech phenomena such as the reduction and assimilation of ‘and’ […]”[BF92, p. 10].

So possible transcriptions of the English word “and” produced by a German speaker could be:

**phonemic** /æ n d/

**broad-phonetic** [ɛ n t]

**narrow-phonetic** [′ɛ ɔ t] 6

For our purpose, we need to perform the transcription on a phonetic rather than on a phonemic basis because we want to look at how German speakers actually pronounce foreign words. However, not every detail (like for example aspiration of plosives or tongue root position while producing vowels) is modeled...
by speech recognizers. That is why we do the transcription an a broad-phonetic rather than on a narrow-phonetic basis.

In our corpora, there are two aspects where you need to pay particular attention to: terminal devoicing and the non-native vowel quality of foreign words. The English word “big” for example is often pronounced \([b \ d k]\)\(^7\) by German speakers instead of /b d g/. An also fairly common characteristic of pronunciation of foreign words is the “nativisation” of vowel quality. As there is no /æ̞/ (open front vowel) in German, the vowel in “cat” (/k æ t/) for instance is often pronounced as a mid-open vowel, i.e., [k e t], because this vowel does exist in German and comes close to the /æ̞/. Another vowel that does not exist in German is the English /Æ̞/ like in “cut” (/k Æ̞ t/); here Germans often produce a neighbor vowel as well: [k a t]. However, in the corpora, some speakers may actually produce the sounds with the correct English phonemes, in which case the transcription should reflect the actual pronunciation. Those aspects of German pronunciation and others like, e.g., the realization of the English “th’” (/ð/ or /θ/) are the phenomena that especially interest us. That’s why we use the broad-phonetic transcription level - with the small addition of expanding the German phoneme set for our annotation by foreign phonemes that the speakers in our corpora use although not represented in their native phoneme set.

### Phoneme Inventory

As transcription symbols we are using the computer readable phonetic alphabet SAMPA (Speech Assessment Methods Phonetic Alphabet) and part of its extension X-SAMPA [Wel97]. One advantage of transcribing a database with SAMPA and X-SAMPA compared to IPA (International Phonetic Alphabet, [Ass96]) is that it is possible to write all phonemes, diacritics, and suprasegmentals in plain ASCII code. This makes it much easier to input transcripts and also to read them on computers where an IPA font is not installed. The SAMPA alphabet is already available in several languages (about 25, whereas most of them are European languages), expanding steadily. “The SAMPA transcription symbols have been developed by or in consultation with native speakers of every language to which they have been applied, but are standardized internationally.” [Wel97] If one wants to transcribe other languages or transcribe on a narrow-phonetic level, one can use X-SAMPA. This extension of SAMPA covers every phoneme, diacritic, and suprasegmental that can be transcribed with IPA.

For our annotation, we are using the German SAMPA phoneme set as general base. In Table 4.1 all German phonemes (written in SAMPA and IPA) are listed with examples.

<table>
<thead>
<tr>
<th>IPA Symbol</th>
<th>SAMPA Symbol</th>
<th>Example</th>
<th>Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p</td>
<td>Pein</td>
<td>pailn</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>Bein</td>
<td>baln</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>Teich</td>
<td>tayC</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>Deich</td>
<td>daiC</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>Kunst</td>
<td>kUnst</td>
</tr>
<tr>
<td>g</td>
<td>g</td>
<td>Gunst</td>
<td>gUnst</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>Verein</td>
<td>fE6”aIn</td>
</tr>
<tr>
<td>pf</td>
<td>pf</td>
<td>Pfahl</td>
<td>pfa:l</td>
</tr>
<tr>
<td>ts</td>
<td>ts</td>
<td>Zahl</td>
<td>tsa:l</td>
</tr>
<tr>
<td>tf</td>
<td>tS</td>
<td>deutsch</td>
<td>dOYtS</td>
</tr>
</tbody>
</table>

\(^7\)Note that we did not annotate the aspiration that would probably take place here for the final stop.
<table>
<thead>
<tr>
<th>dʒ</th>
<th>dZ</th>
<th>Dschungel &quot;dZUN=l</th>
</tr>
</thead>
<tbody>
<tr>
<td>f</td>
<td>f</td>
<td>fast</td>
</tr>
<tr>
<td>v</td>
<td>v</td>
<td>was</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>Tasse &quot;tas@</td>
</tr>
<tr>
<td>z</td>
<td>z</td>
<td>Hase &quot;ha:z@</td>
</tr>
<tr>
<td>ŋ</td>
<td>S</td>
<td>waschen &quot;vaS=n</td>
</tr>
<tr>
<td>ß</td>
<td>Z</td>
<td>Genie &quot;Ze&quot;ni:</td>
</tr>
<tr>
<td>ç</td>
<td>C</td>
<td>sicher &quot;ziC6</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>Jahr ja:6</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>Buch bu:x</td>
</tr>
<tr>
<td>h</td>
<td>h</td>
<td>Hand hant</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>mein maln</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>nein nalm</td>
</tr>
<tr>
<td>ñ</td>
<td>N</td>
<td>Ding dIN</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>Leim lalm</td>
</tr>
<tr>
<td>ð</td>
<td>R</td>
<td>Reim Ralm</td>
</tr>
</tbody>
</table>
### Vowels

<table>
<thead>
<tr>
<th>Symbol</th>
<th>phoneme</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I</td>
<td>Sitz</td>
</tr>
<tr>
<td>Ε</td>
<td>E</td>
<td>Gesetz</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>Satz</td>
</tr>
<tr>
<td>ι</td>
<td>O</td>
<td>Trotz</td>
</tr>
<tr>
<td>u</td>
<td>U</td>
<td>Schutz</td>
</tr>
<tr>
<td>ι</td>
<td>Y</td>
<td>hübsch</td>
</tr>
<tr>
<td>œe</td>
<td>9</td>
<td>plötzlich</td>
</tr>
<tr>
<td>i:</td>
<td>i:</td>
<td>Lied</td>
</tr>
<tr>
<td>e:</td>
<td>e:</td>
<td>Beet</td>
</tr>
<tr>
<td>e:</td>
<td>E:</td>
<td>spät</td>
</tr>
<tr>
<td>a:</td>
<td>a:</td>
<td>Tat</td>
</tr>
<tr>
<td>o:</td>
<td>o:</td>
<td>rot</td>
</tr>
<tr>
<td>u:</td>
<td>u:</td>
<td>Blut</td>
</tr>
<tr>
<td>y:</td>
<td>y:</td>
<td>süß</td>
</tr>
<tr>
<td>ø:</td>
<td>2:</td>
<td>blöd</td>
</tr>
<tr>
<td>ai</td>
<td>ai</td>
<td>Eis</td>
</tr>
<tr>
<td>aU</td>
<td>aU</td>
<td>Haus</td>
</tr>
<tr>
<td>OY</td>
<td>OY</td>
<td>Kreuz</td>
</tr>
<tr>
<td>e @</td>
<td>@</td>
<td>bitte</td>
</tr>
<tr>
<td>e</td>
<td>6</td>
<td>besser</td>
</tr>
<tr>
<td>i:e</td>
<td>i:e</td>
<td>Tier</td>
</tr>
<tr>
<td>e:</td>
<td>E:</td>
<td>Wirt</td>
</tr>
<tr>
<td>y:e</td>
<td>y:e</td>
<td>Tür</td>
</tr>
<tr>
<td>y:e</td>
<td>Y:e</td>
<td>Türke</td>
</tr>
<tr>
<td>e:e</td>
<td>e:e</td>
<td>schwer</td>
</tr>
<tr>
<td>e:e</td>
<td>E:e</td>
<td>Berg</td>
</tr>
<tr>
<td>e:e</td>
<td>E:e</td>
<td>Bär</td>
</tr>
<tr>
<td>φ:e</td>
<td>2:e</td>
<td>Föhr</td>
</tr>
<tr>
<td>œ:e</td>
<td>9:e</td>
<td>Wörter</td>
</tr>
<tr>
<td>a:e</td>
<td>a:e</td>
<td>Haar</td>
</tr>
<tr>
<td>a:e</td>
<td>a:e</td>
<td>hart</td>
</tr>
<tr>
<td>u:e</td>
<td>u:e</td>
<td>Kur</td>
</tr>
<tr>
<td>u:e</td>
<td>U:e</td>
<td>kurz</td>
</tr>
<tr>
<td>o:e</td>
<td>o:e</td>
<td>Ohr</td>
</tr>
<tr>
<td>œ:e</td>
<td>O:e</td>
<td>dort</td>
</tr>
</tbody>
</table>

Table 4.1: German SAMPA computer readable alphabet

Because it is impossible to transcribe foreign words phonetically only with the German phonemes, we are using additional phonemes that do not exist in German but that were used by our subjects. Therefore, we have added phonemes from the SAMPA set for other languages. These additional phonemes and corresponding examples are listed in Table 4.2. Sometimes, there are just small differences between a foreign
and a German phoneme; for example, the phoneme /ɔ/ exists in German, but it is only allowed/applicable as short vowel, not lengthened (/ɔ:/). Other phonemes differ completely from the German phoneme set such as /ʊ/ for example.

<table>
<thead>
<tr>
<th>IPA</th>
<th>SAMPA</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʊ</td>
<td>r</td>
<td>right</td>
</tr>
<tr>
<td>θ</td>
<td>T</td>
<td>thin</td>
</tr>
<tr>
<td>ð</td>
<td>D</td>
<td>this</td>
</tr>
<tr>
<td>w</td>
<td>w</td>
<td>walking</td>
</tr>
<tr>
<td>əʊ</td>
<td>@U</td>
<td>nose</td>
</tr>
<tr>
<td>ɛt</td>
<td>el</td>
<td>day</td>
</tr>
<tr>
<td>ʊə</td>
<td>U@</td>
<td>yours</td>
</tr>
<tr>
<td>ʊː</td>
<td>3:</td>
<td>world</td>
</tr>
<tr>
<td>ɔː</td>
<td>O:</td>
<td>cause</td>
</tr>
<tr>
<td>ɛ̞</td>
<td>e̞</td>
<td>Chopin</td>
</tr>
<tr>
<td>ɔ̞</td>
<td>o̞</td>
<td>bon</td>
</tr>
<tr>
<td>ɑː</td>
<td>A:</td>
<td>stars</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
<td>happy</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>into</td>
</tr>
<tr>
<td>Λ</td>
<td>V</td>
<td>cut</td>
</tr>
<tr>
<td>æ</td>
<td>{</td>
<td>pat</td>
</tr>
</tbody>
</table>

Table 4.2: Non-German IPA and SAMPA symbols used by our subjects

If, during transcription, you discover a phoneme that is not in the set, please consult with your supervisor to find the appropriate representation.

**Diacritics and Suprasegmentals**  Besides the annotation of the phonemes, you also need to transcribe several classes of diacritics and suprasegmentals. The only diacritics and suprasegmentals you should annotate are stress, syllabic, and mid-centralization, as listed in Table 4.3.

<table>
<thead>
<tr>
<th>IPA symbol</th>
<th>SAMPA symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʰi</td>
<td>&quot;i:</td>
<td>primary stress</td>
</tr>
<tr>
<td>ʰi</td>
<td>%i:</td>
<td>secondary stress</td>
</tr>
<tr>
<td>n</td>
<td>n=</td>
<td>syllabic consonant</td>
</tr>
<tr>
<td>ɛ</td>
<td>e̞</td>
<td>non-syllabic vowel</td>
</tr>
<tr>
<td>ɛ̞</td>
<td>e̞x</td>
<td>mid-centralized vowel</td>
</tr>
</tbody>
</table>

Table 4.3: IPA and SAMPA Symbols for suprasegmentals and diacritics used in our annotation
Usually, the diacritics primary and secondary stress are meant to be set in a single word that is longer than one syllable so that you know which syllable to stress. As our recordings are not meant to give advice on how to pronounce words correctly but how they actually were pronounced by this very speaker, not every word that is longer than one syllable needs to have a stress mark by default. We are only looking at compounds like multiword artist names, song or album titles. So “David Bowie” for example needs at least one primary stress mark in our transcription, but may have more. Secondary stress marks are optional. An example of a transcription from one of the corpora is:

( David /dIvId/ Bowie /bU/ )

There you can see that the accented syllable of the surname has been marked with the primary stress whereas the accented syllable of the first name only got a secondary stress.

After having given an example of the suprasegmental transcription of stress, we have a quick look at the diacritics transcription of the marked (non-)syllabicity, i.e., syllabicity with consonants and non-syllabicity with vowels. Common syllabifications of consonants are nasals and laterals in word-final position like for example “Mission /mIsmI/ Impossible /plbosIblpI/”, but there are much less frequently occuring examples in our database as well:

“Digital /dIts/”

Non-syllabic vowels only appear in syllables where already another vowel is the nucleus of the syllable, like for example in the misspoken artist name

( Freedom /fIdm/ Call /kOll/ ).

A last general note has to be made on the work of transcribing foreign words in our corpora: Because we need to transcribe all foreign words how they actually were pronounced by the subjects, you will actually need to transcribe many occurrences of the same word because different subjects (or even the subject himself) may pronounce the same words differently. There are two subjects, e.g., who wanted to listen to songs from Beyonce. Since they pronounced this name quite differently, both occurrences need to be transcribed: Beyonce /bIeO/ and Beyonce /bI0/.

---

8 As we have now described the SAMPA alphabet, all further examples will be transcribed in SAMPA like in our corpora.

9 German speakers sometimes have problems pronouncing the consonant cluster of a consonant and the alveolar approximant “r” (/s/).
4.7 Syntactic Features [USAAR]

4.7.1 Annotation Purpose

This annotation serves the objective of WP3, namely the generation of varied contextually appropriate system output. Its main purpose is to get a detailed description of the syntactic constructions used in the SAMMIE corpora, including grammatical functions and some semantic features. We are particularly focusing on word order related phenomena. This description will allow us to systematically investigate the use of a range of constructions in different contexts, and use the results to motivate sentence planning decisions (cf. also [Poe04], [Poe00] and [JW05]).

4.7.2 Annotation Instructions

The annotation task consists of:

- marking the elements to be annotated (= markables);
- assigning features (attribute values) to these markables.

Markables For our purposes, the markables of the syntax layer are clause or clause-like units:

1. main (i.e., matrix) clauses (also when it is elliptical as in example 9)

   (7) [ich habe einen Titel herausgesucht]
   I have a title chosen
   I've chosen a title

   (8) [ich möchte,] [dass Du Mozart spielst]
   I wish that you Mozart play
   I’d like you to play Mozart

   (9) [wird gesucht]
   is searched
   is being searched

2. relative clauses:

   (10) [ich habe vier Tracks gefunden] [in denen der Titel Yesterday vorkommt]
   I have four tracks found in which the title Yesterday occurs
   I’ve found four tracks with Yesterday in the title

3. that-clauses:

   (11) [ich möchte,] [dass Du Mozart spielst]
   I wish that you Mozart play
   I’d like you to play Mozart
4. all other subordinate clauses functioning as a free modifier (temporal, causal, etc.):

(13) [Bevor Du was sagen konntest] [habe ich einen Titel herausgesucht]  
before you something say could  I have a title chosen
I’ve chosen a title before you was able to say something

(14) [Weil es regnet] sind die Bäume grün  
because it rains  are the trees green
the trees are green because it is raining

5. elements of sentence coordination in which the VP is coordinated:

(15) [Maria singt] [und tanzt]  
Mary sings  and dances
Mary is singing and dancing

(16) [Hans wollte schwimmen] [und ins Kino gehen]  
John wanted swim  and to the cinema go
John wanted to swim and to go to the cinema

However, this does not apply to NP coordination (e.g., [Maria und Hans gehen ins Kino]).

6. short, verbless utterances:

(17) [ja]  
yes
yes

(18) [richtig]  
exactly
right

(19) [bitte]  
please
please

(20) [ja bitte]  
yes please
yes please

(21) [gerne]  
voluntarily
You’re welcome!
(22)  [aber gerne]  
    but willingly  
    You’re welcome!

Material marked as self-speech is not to be annotated.

Features For our purposes, we are interested in the following features:

1. type;  
2. elliptical;  
3. gFunction;  
4. isFronted;  
5. polarity;  
6. verbed;  
7. diathesis;  
8. hasSplitNP;  
9. vorfeldContent;  
10. hasExtraposition;  
11. hasScrambledMF;

Using the features above, we describe clauses in a traditional way, capturing the clause type, grammatical function, diathesis, etc. Some of the features capture the information about the Topological Field Model. The Topological Field Model (cf. [Höh83]) is a linguistically motivated theory-neutral description of the macro-structure of both simple clauses and complex sentences. The basic model assumes a clause division into: **Vorfeld** (pre-field), **Linke Klammer** (left bracket), **Mittelfeld** (middle field), **Rechte Klammer** (right bracket), and **Nachfeld** (post-field) (cf. Figure 4.2). For certain fields, there are restrictions on the number and type of constituents. In the **Vorfeld**, for instance, German grammar rules restrict the number of constituents to at most one. Not all the fields have to be occupied. In complex sentences, the model is applied to each clause individually, i.e. in paratactically conjoined clauses iteratively, while in hypotactically conjoined clauses recursively. 

From the point of view of the Topological Field Model, the clause macrostructure can be characterized as follows: (i) in verb-initial and verb-second clauses, the finite verb occupies the **Linke Klammer**; (ii) in verb-final clauses, the finite verb occupies the **Rechte Klammer**. Furthermore, the **Vorfeld** of main declarative clauses can be occupied either by one argument of the finite verb or by an adjunct\(^{10}\)

**General note:** In order to keep a uniform annotation scheme and to catch all possible phenomena not covered by the defined values, some of the features will get a value called other. The feature value none is assigned whenever the feature under scrutiny does not apply to a specific annotation unit.

\(^{10}\)In case of adjuncts of the same semantic type, a cluster of adjuncts may also be allowed in the **Vorfeld**.
### Figure 4.2: Examples of topological analyses of German sentence types

<table>
<thead>
<tr>
<th>Vorfeld</th>
<th>Linke Klammer</th>
<th>Mittelfeld</th>
<th>Rechte Klammer</th>
<th>Nachfeld</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kannst</td>
<td>du Mozart (oder Bach)</td>
<td>spielen?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spiele</td>
<td>Mozart!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Die Kinder</td>
<td>spielen</td>
<td>Mozart.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wer</td>
<td>spielt</td>
<td>Mozart?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wenn</td>
<td>du</td>
<td>willst.…</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wenn du willst,</td>
<td>kannst</td>
<td>Du Mozart</td>
<td>spielen.</td>
<td></td>
</tr>
<tr>
<td>Maria</td>
<td>spielt</td>
<td>ein Stück.</td>
<td>das von Mozart komponiert wurde.</td>
<td></td>
</tr>
<tr>
<td>… das</td>
<td>von Mozart</td>
<td>komponiert wurde.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich</td>
<td>glaube,</td>
<td>daß Maria Klavier spielen kann.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>… daß</td>
<td>Maria Klavier</td>
<td>spielen kann.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**type**

Type characterizes the global syntactic pattern, according to traditional grammar. For main clauses, the values also denote the mood of the clause.

1. **declarative**

   (23) [Maria singt]
   
   Mary sings
   
   Mary is singing

2. **interrogativeWH**

   (24) [Was ist das?]
   
   what is that
   
   What’s that?

3. **interrogativeYNgen**

   (25) [Magst Du Mozart?]
   
   like you Mozart
   
   Do you like Mozart?

4. **interrogativeYNimp**

   (26) [Kannst Du mir den Namen des Künstlers sagen?]
   
   can you me the name of the artist tell
   
   Can you tell me the name of the artist?
(27)  [Magst Du Mozart oder lieber Bach?]
like you Mozart or rather Bach
Do you like Mozart or rather Bach?

6. imperative

(28)  [Zeige die Liste an!]
Show the list
Show the list!

7. imperativeINF

(29)  [Liste anzeigen]
list show
Show the list!

8. exclamative

(30)  [Exzellen!] 
excellent
Excellent!

9. relative

(31)  [das Lied [das ich mag] ist nicht in dieser Liste]
the song that I like is not in this list
the song I like is not in this list

10. that-clause

(32)  [dass er annotiert] [ist schön]
that he annotates is nice
it is nice that he is annotating

11. temporal

(33)  [als ich noch Corpora annotierte] [war ich in Saarbrücken]
when I still corpora annotated was I in Saarbrücken
It was in Saarbrücken when I was still annotating corpora

12. conditional

(34)  [wenn ich annotiere] [mußt Du Dich um das Lexikon kümmern]
if I annotate must you yourself of the lexicon care
if I annotate you have to take care of the lexicon
13. causal

(35) [ich habe Kopfschmerzen] [weil ich annotiere]
I have headaches because I annotate
since I’m annotating I have a headache

14. concessional

(36) [ich habe keine Kopfschmerzen] [obwohl ich annotiere]
I have no headaches although I annotate
although I annotate

15. consecutive

(37) [es is so kalt] [dass die Fenster zufrieren]
it is so cold that the windows freeze up
it is so cold that the windows are frozen

16. final

(38) [ich bereite das Programm vor] [damit ich annotieren kann]
I prepare the program with it I annotate can
I’m preparing the program to be able to annotate

17. modal

(39) [indem ich anotiere] [mache ich die Welt glücklich]
in it I annotate make I the world happy
by annotating, I make the world happy

18. local

(40) [es ist mir gleich] [wo ich anotiere]
it is me equal where I annotate
It doesn’t matter where I annotate

19. directional

(41) [es ist mir gleich] [wohin ich gehe]
it is me equal where I go
It doesn’t matter where I go

20. infinitiveEXT
(42) [ich muss mich beeilen] [um zu annotieren]
I must myself rush in order to annotate
I have to rush in order to annotate

21. shortUtterance

(43) [ja]
yes
yes

22. other: if none of the above is applicable

**elliptical**

This feature stores the information whether a markable is a genuine elliptical form, non-genuine elliptical form, i.e., a clause which is part of a bigger sentence, or not elliptical. By **genuine ellipsis**, we mean clause units that are fragmentary, or include gaps. **Non-genuine ellipses** are defined as clauses of complex sentences, or as verbal phrases in a coordination of verbal complexes. Hence, they are syntactically incomplete, because they are part of a larger construction. All constructions which are syntactically complete clause units are defined as **non-elliptical**.

1. ellipsisGEN:

(44) Subj: [ich möchte eine Liste erstellen]
Subj: I wish a list make
Subj: I’d like to make a list

MP3: [wie groß?]
MP3: how big

MP3: how big?

(45) Subj: [wie heißt das Album?]
Subj: how is called the album
Subj: what is the name of the album?

MP3: [Greatest Hits]
MP3: Greatest Hits

MP3: Greatest Hits

2. ellipsisNONGEN:

(46) [Hans wollte] [dass ich ihn am Klavier begleite]
John wanted that I him on the piano accompany
John wanted me to accompany him on the piano
(47)  \textit{[Hans wollte schwimmen]} \textit{[und ins Kino gehen]}
John wanted to swim and to go to the cinema
John wanted to swim and to go to the cinema

3. nonEllipsis:

(48)  \textit{[Hans wollte schwimmen]} \textit{[und ins Kino gehen]}
John wanted to swim and to go to the cinema
John wanted to swim and to go to the cinema

(49)  \textit{[Hans und Maria sangen]}
John and Mary sang
John and Mary were singing

(50)  \textit{[die Kinder sangen ein Lied]}
the children sang a song
the children were singing a song

(51)  \textit{[ja]}
yes
yes

(52)  \textit{[richtig]}
exactly
right

gFunction
This feature stores the information of the grammatical function of a clause within a matrix clause. The values of this feature are the following:

1. subject

(53)  \textit{[dass Maria annotiert,]} \textit{[stört mich]}
that Mary annotates disturbs me
I don’t like that Mary annotates

2. object

(54)  \textit{[dass Maria annotiert,]} \textit{[mag ich]}
that Mary annotates like I
I like that Mary annotates

3. sComp
(55) [ich weiß] [dass Maria annotiert]
I know that Mary annotates
I know that Mary annotates

4. other

(56) [während Maria annotierte] [Hans schlief]
while Mary annotated John slept
John slept while Maria was annotating

isFronted

This feature stores the information whether a unit occurs in the Vorfeld of another one or not, and applies only to embedded clause units. Non-elliptical units, genuine elliptical units and matrix clauses are to be assigned the value none.

1. true

(57) [dass Maria annotiert,] [stört mich]
that Mary annotates disturbs me
I don’t like that Mary annotates

2. false

(58) [es stört mich,] [dass Maria annotiert]
it disturbs me that Mary annotates
I don’t like that Mary annotates

3. none

(59) [ich möchte eine Liste mit drei Liedern erstellen]
I wish a list with three songs make
I’d like to make a list with three songs

(60) [es stört mich,] [dass Maria annotiert]
it disturbs me that Mary annotates
I don’t like that Mary annotates

(61) [dass Maria annotiert,] [stört mich]
that Mary annotates disturbs me
I don’t like that Mary annotates

(62) [Ja]
Yes
Yes

(63) [Yesterday von den Beatles,]
Yesterday by the Beatles
Yesterday by the Beatles.
polarity

This feature stores the information about the occurrence of negation particles in the unit. All non-elliptical and non-genuine elliptical clauses and clause-like units containing negative particles are marked as negative as well as short negative answers (nein), while non-elliptical and non-genuine elliptical clauses and clause-like units without a negative particle as well as short positive answers (ja, jawohl) are marked as positive. Genuine elliptical units are to be assigned the value negative if if contains an explicit negation particle and none otherwise.

1. positive

(64) [Maria annotiert]
Mary annotates
Mary is annotating

(65) [Ja]
Yes
Yes

(66) [Yesterday von den Beatles]
Yesterday by the Beatles
Yesterday by the Beatles

2. negative

(67) [Maria annotiert nicht]
Mary does not annotate
Mary is not annotating

(68) [Nein]
No
No

(69) [nicht Yesterday von den Beatles]
not Yesterday by the Beatles
not Yesterday by the Beatles

3. none

(70) [Yesterday von den Beatles]
Yesterday by the Beatles
Yesterday von den Beatles

(71) [bitte]
please
please

(72) [einen Moment]
a moment
just a moment
(73) [gerne]  
    willingly  
    You’re welcome!

verbed

If the unit contains a verbal element the value this feature gets the value true otherwise false.

1. true

(74) [Hans und Maria sangen]  
    John and Mary sang  
    John and Mary were singing

(75) [wird gesucht]  
    is searched  
    searching

(76) [gefunden]  
    found  
    found

2. false

(77) [dieses Lied]  
    this song  
    this song

(78) [Yesterday von den Beatles]  
    Yesterday by the Beatles  
    Yesterday by the Beatles

(79) [ok]  
    ok  
    ok

diathesis

This feature stores the information about the genus verbi of the unit. Only non-elliptical units that contain transitive verbs are to be assigned either the value active, passive or medial, all other units are to be assigned the value none.

1. active

(80) [Maria annotiert das Corpus]  
    Mary annotates the corpus  
    Mary is annotating the corpus
2. passive

(81) [das Corpus wird von Maria annotiert]
the corpus is by Mary annotated
the corpus is being annotated by Mary

3. medial

(82) [das Corpus läßt sich leicht annotieren]
the corpus lets itself easily annotate
it is easy to annotate the corpus

4. none

(83) [Ja]
Yes
Yes

(84) [Yesterday von den Beatles]
Yesterday by the Beatles
Yesterday by the Beatles

(85) [verstehe]
understand
I understand

(86) [wird gesucht]
is searched
Searching

(87) [einen Moment]
a moment
just a moment

(88) [Hans schläft]
John sleeps
John is sleeping

**hasSplitNP**

This feature stores the information whether a unit contains split NPs or not.

1. true

(89) [mit drei Liedern möchte ich eine Liste erstellen]
with three songs wish I a list make
I’d like to make a list with three songs
2. false

(90) [ich möchte eine Liste mit drei Liedern erstellen]
I wish a list with three songs make
I’d like to make a list with three songs

vorfeldContent

This feature stores the information about the content of the Vorfeld in V2 units, i.e., which grammatical function does the element occupying the Vorfeld have. Other than V2 clauses get the value none.

1. subj

(91) [Maria annotiert den Satz]
Mary_{nom} annotates the clause_{acc}
Mary is annotating the clause

2. dObj

(92) [den Satz annotiert Maria]
the clause_{acc} annotates Maria_{nom}
Mary is annotating the clause

3. iObj

(93) [Peter annotiert Maria den Satz]
Peter_{dat} annotates Mary_{nom} the clause_{acc}
Mary annotates the clause for Peter

4. pObj

(94) [für Peter annotiert Maria den Satz]
for Peter_{acc} annotates Mary_{nom} the clause_{acc}
Mary annotates the corpus for Peter

5. sComp

(95) [dass Maria das Corpus annotiert] [ist schön]
that Mary_{nom} the corpus_{acc} annotates is nice
it is nice that Mary is annotating the corpus

6. advTime
(96) [gestern hat Maria den Satz annotiert]
yesterday has Mary\textsubscript{nom} the clause\textsubscript{acc} annotated
Yesterday, Mary annotated the clause

(97) [als es regnete] [hatte Maria den Satz annotiert]
when it rained had Mary\textsubscript{nom} the clause\textsubscript{acc} annotated
when raining, Mary was annotating the clause

7. advLoc

(98) [im Pool annotiert Maria den Satz]
in the pool annotates Mary\textsubscript{nom} the clause\textsubscript{acc}
Mary is annotating the corpus in the pool

8. advDir

(99) [in die Schule geht Maria jetzt]
in the school goes Mary now
Mary is going to school now

9. advManner

(100) [gewissenhaft annotiert Maria den Satz]
conscientiously annotates Mary\textsubscript{nom} the clause\textsubscript{acc}
Mary is annotating the corpus conscientiously

10. advCause

(101) [suchtbedingt annotiert Maria das Corpus]
addictively annotates Mary\textsubscript{nom} the corpus\textsubscript{acc}
due to her addiction Mary is annotating the corpus

11. advInstr

(102) [mit MMax annotiert Maria das Corpus]
with MMax annotates Mary\textsubscript{nom} the corpus\textsubscript{acc}
Mary is using MMax to annotate the corpus

12. none

(103) [dass Maria das Corpus annotiert] [ist schön]
that Mary\textsubscript{nom} the corpus\textsubscript{acc} annotates is nice
it is nice that Mary is annotating the corpus

13. other: if none of above applies
hasExtraposition

This feature stores the information whether a unit which is verb-second and has a non-empty right bracket (Rechte Klammer) contains extraposed elements or not as well as the type of the extraposed element; all other clause units be assigned the value none.

1. PP

(104) [ich möchte eine Liste erstellen mit drei Liedern]
I wish a list make with three songs
I’d like to make a list with three songs

2. predNom

(105) [sie soll heissen Favoriten]
she shall be called Favoriten
its name shall be Favoriten

3. relCl

(106) [ich möchte eine Liste erstellen,] [die drei Lieder enthalten soll]
I wish a list make which three songs contain shall
I’d like to make a list with three songs

4. other

(107) [ich möchte eine Liste erstellen jetzt]
I like a list make now
I’d like to make a list now

5. none

(108) [dass Maria das Corpus annotiert] [ist schön]
that Mary the corpus annotates is nice
it is nice that Mary is annotating the corpus

(109) [Ich weiß] [dass Maria das Corpus annotiert]
I know that Mary the corpus annotates
I know that Mary is annotating the corpus

(110) [Ich habe gewußt] [dass Maria das Corpus annotiert]
I have known that Mary the corpus annotates
I knew that Mary is annotating the corpus

(111) [ich möchte eine Liste, [die drei Lieder enthalten soll,] erstellen]
I wish a list which three songs contain shall make
I’d like to make a list with three songs
hasScrambledMF

This feature stores the information whether a clause unit – be it elliptical or not, be it verbed or not – exhibits non-default positioning of elements or not. Short answers are assigned the value *none*.

1. true

   (112)  [ich möchte mit drei Liedern eine Liste erstellen]
   I wish with three songs a list make
   I’d like to make a list with three songs

   (113)  [von den Beatles Yesterday]
   by the Beatles Yesterday
   Yesterday by the Beatles

2. false

   (114)  [ich möchte eine Liste mit drei Liedern erstellen]
   I wish a list with three songs make
   I’d like to make a list with three songs

   (115)  [Yesterday von den Beatles]
   Yesterday by the Beatles
   Yesterday by the Beatles

3. none

   (116)  [Singe!]
   sing
   Sing!

   (117)  [dieses Lied]
   this song
   this song

   (118)  [ok]
   ok
   ok
4.8 Discourse-Entity Features [USAAR]

4.8.1 Annotation Purpose

In order to investigate systematic reference phenomena in our domain, so that we can incorporate the findings in the natural language generation module of the SAMMIE system, and thus improve the quality of the system output, we annotate expressions that introduce discourse entities. We assign them the properties of both the surface expressions (their form) and the discourse entities (their content and relationships between them).

In formulating our annotation scheme we adhered to the recommendations of the Text Encoding Initiative\textsuperscript{11} and the Discourse Resource Initiative. We were also inspired by existing reference annotation schemes: MUC-6 and MUC-7 (MUC Coreference Specification), DRAMA (cf. [Pas96]), the DRI guidelines (cf. [CDRW97]) the MATE project (cf. [PBR99])\textsuperscript{12}, the GNOME project (cf. [Poe04])\textsuperscript{13}, the scheme used by the NLP group at EML-Research (cf. [MS01])\textsuperscript{14} and our own earlier efforts in the MULI project (cf. [BBHS+04]) and the DIALOG project (cf. [WVT+04]).

4.8.2 Annotation Instructions

Your task is to annotate reference. You will mark expressions that evoke entities in discourse, and you will assign them features that reflect the properties of the expressions themselves (e.g., what linguistic form they have, for example definite NP vs. pronoun) and features that reflect the properties of the underlying discourse entity (e.g., the type, for example, a song vs. an artist). You will also determine whether a discourse entity evoked by a linguistic expression is being mentioned for the first time or has been mentioned before, and in the latter case you will identify the previous mention.

We call an expression that evokes a discourse entity \textit{markable}, because your task is to mark them in the text of the dialogue. The second part of the task is to determine the features of each markable by assigning values to the listed attributes.

In the following, we first explain which expressions are markable, and then how to determine the features.

\textbf{Markables}

A markable is to be created for every linguistic expression that evokes a \textit{discourse entity} (or: discourse referent in the sense used in Discourse Representation Theory and alike). Intuitively, a discourse entity is “something that can be talked about”. It can be something concrete, such as the discourse entity evoked by the occurrence of the NP \textit{the chair} or something abstract, such as the discourse entity evoked by \textit{our freedom}. A discourse entity may correspond to a concrete or abstract “object” as well as an “eventuality” (event, state, process), e.g., \textit{swimming}.

Thus, the linguistic expressions that are markable are noun phrases and NP surrogates (i.e., NP-like expressions which occur in free variation (are replaceable) with prototypical NPs. This includes also pronouns, nominalizations, gerunds and participles (when they are replaceable by an NP), as well as possessives and

\textsuperscript{11}The TEI website \url{http://www.tei-c.org/}
\textsuperscript{12}The MATE project website \url{http://mate.mip.ou.dk}
\textsuperscript{13}The GNOME project website \url{http://www.hcrc.ed.ac.uk/~gnome/}
\textsuperscript{14}The MMAX project website \url{http://mmax.eml-research.de}
pronominal adverbs (e.g., *daraus*). Below we give examples of markables, and point out some difficult cases.

**Prototypical NPs** A prototypical NP has a noun as its head. It may start with an article or determiner, and it may contain pre- or post-modifiers. The referent of a prototypical NP functions as an argument of a predicate (even though sometimes the predicate may not be explicit, e.g., isolated NPs in headings or short responses)

(119) [I] saw [a/the/this/his/every/some/no chair], [the metal chair], [the orchestra chair].
(120) Did [you] see [a chair]?
(121) [I] didn’t see [a chair].
(122) [Which chair] did [you] see?
(123) [What] did [you] see?

Note that a markable may be contained within another markable: [the chair in [the front row]], [das Lied von [den Beatles]], [the chair (that) [she] always sits in]. See below for instructions how to handle conjoined NPs and apposition.

**Names and Other Named Entities** Proper nouns, such as names of persons, places, songs, albums etc. and other named entities, such as dates, times, currency amounts, and percentages are all markables.

(124) [Herbert Grönemeyer]
(125) [President Clinton] was re-elected.
(126) [der [Heidelberger] Hauptbahnhof]
(127) [Hey Jude]
(128) [(the) year 2002]
(129) [die Neunziger Jahre]
(130) [die neunziger Jahre]
(131) [die Neunziger]
(132) [die 90er]
(133) [zwangig Minuten]

In addition to named entities, other identifiers that are, in the opinion of the annotator, clearly not decomposable should be treated as atomic as well:

(134) [the large strange-looking building, which is [Widener Library]]
(135) okay then [I]’ll take [engine E two] … so uh [the plan] is to take [engine E two]
Conjoined NPs  In NP conjunctions, the entire conjoined NP as well as each of the individual conjuncts are markables:

(136)  [Edna Fribble and Sam Morton] addressed [the meeting] yesterday. [Ms. Fribble] discussed [coreference], and [Mr. Morton] discussed [unnamed entities].


Gerundive clauses, infinitives and present participles  Gerundive clauses, infinitives and present participles are markable, if and only if they occur in free variation with prototypical NPs (i.e., not as part of a verbal group)

(138)  Sorry, [displaying all songs] it not possible.

(139)  [They] had been accused [of ignoring [the environment]].

(140)  [He] is prone [to semi-catatonic trances induced [by the playing [of the violin]].

(141)  [slowing [of the economy]], [program trading], [excessive spending]

Nominalizations  Nominalizations are also markable, if and only of they occur in free variation with prototypical NPs


(143)  [Das Singen] ist [[des Informatik]es] Lust].

Single words functioning as noun phrases  This includes various types of pronouns and nominal expressions that function as NPs.

- Personal pronouns

  (144)  [I] don’t believe [you].

  (145)  [She] doesn’t believe [him].

  (146)  [He] shot [himself] [with [his] revolver].

- Demonstrative pronouns

  (147)  [The rate, which was [6 percent]], was higher than [that offered [by the other bank]].


  (149)  [Das] mache [ich] gerade.

  (150)  [Dieses] gefällt [mir].
• Indefinite Pronouns

(151) [The nurse] got [a raise] and [[his friend] wants [one].
(152) [The student] heard about [the seminars] and attended [some].
(153) If [anyone] thought [of the John Harvey], . . .
(154) . . . and [whatever else it was that [she] carried]. [Which] was [poison gas].

• Reflexive pronouns

(155) [He] shot [himself] [with [his] revolver].

• Numeral/Quantifier

(156) [The first woman] had [one son] and [the second] had [three].

• Non-restrictive relative pronouns when they do not modify a head (i.e., in headless relative clauses): who, which

Why not to mark the whole clause???

(157) [The aborigine] lives [on the cruelest land] [I] have ever seen.[Which] does not mean that [it] is ugly.
(158) . . . and [whatever else it was that [she] carried.] [Which] was [poison gas].

Nominal modifiers

• Possessives

Both the possessive modifier (pronoun or genitive NP) and the entire NP are markables. The head noun of the NP by itself is not markable.

(159) “There is [no business reason] [for [my] departure]”, [he] added.
(160) [[its] chairperson]
(161) [He] shot [himself] [with [his] revolver].
(162) [[Peter’s] book]
(164) . . . [die Schriften [Winkelmanns]]

• Nominal pre- and post-modifiers

The pre- and post-modifier(s) as well as the entire NP are markables. The head noun of the NP by itself is not markable.
The price of aluminum siding has steadily increased, as the market for aluminum reacts to the strike in Chile.

He was accused of money laundering and drug trafficking. However, the trade in drugs...

Ocean Drilling & Exploration Co. will sell its contract drilling business. Ocean Drilling said it will offer [(15% to 20%) of the contract drilling business] through an initial [(public offering)] in the near future.

The rate, which was [6 percent], was higher than that offered [by the other bank].

Nominal groups with no or ersatz head noun

- Titles, Epithets and Idioms
  
  (169) [The old so-and-so] finally bought [a new car].
  
  (170) [I] need [a new whatcha-ma-call-it].

- Adjectives as heads
  
  (171) [I] prefer [the largest].
  
  (172) [The poorest [of the seven]] gave up [a brilliant law practice] to enter [[Washington’s] Cabinet].

- Quantifiers as heads
  
  (173) [A few people] found [their] way [to the destination] but [a great many] did not understand [the directions].
  
  (174) [(Its] dangerous effects] have been downgraded [to the public] by some who believe [national security] requires [further testing]].

- Wh-pronoun clauses, e.g., who, whom, whoever, what, whatever
  
  (175) [Whoever rises [to the occasion]] walks [a treacherous path [to leadership]].
  
  (176) [Who won] is not revealed, but [[Winslow’s] daughter Eleanor] says [they] got [up to 1,212 words].
  
  (177) [the John Harvey] would likely be blown up with [[her] own ammo] and [whatever else it was that [she] carried]].

Predicate Nominals  NPs within predicate nominals are markables.

(178) [Bill Clinton] is [the President [of the United States]].
(179) [Mediation] is [a viable alternative [to bankruptcy]].
(180) [Farm-debt mediation] is [one [of the [Farm Belt’s] success stories.]]
(181) [Herr Müller] ist [ein junger Maler].
(182) [Gerda] wird [Mutter]
(183) [Schröder] blieb [Bundeskanzler].
Appositions  Appositives, i.e., expressions as apposition, are not to be annotated as separate markables. There is to be one markable spanning the units within the apposition.

(184) [The Beatles song *Yesterday*] is very nice.
(185) [*Yesterday, the Beatles song,*] is [one [of [my] favourite]].

But note that “the Beatles” itself is also a markable, which is embedded within the markable spanning the whole apposition. So, the full marking will be:

(186) [The *[Beatles] song Yesterday*] is very nice.
(187) [*Yesterday the *[Beatles] song*] is [one [of [my] favourite]].

Non-referential NPs and Pronouns  Non-referential pronouns, such as the subject in presentational constructions (there-insertion), pleonastic or expletive pronouns are markables.

(188) [It] is raining.
(189) [It] is important that *[he]* comes.
(190) And [*there*] was [light].
(191) [Es] gibt [zwanzig Lieder [mit diesem Namen]] [in der Datenbank].

Features

Each markable should be annotated with the following features:

1. deType;
2. typeMention;
3. properNameMention;
4. modification;
5. npForm;
6. quantNP;
7. gFunction;
8. informationStatus.
**deType**

For markables which refer to objects in the MP3 domain, this feature stores the information about the Domain Object type of the corresponding discourse entity. The following **domain object types** are considered: song, album, playlist, artist, name, genre, length, and year. The types participant, event, other, and none do not count as domain objects. Expressions referring to the participants of the dialogue will get the value participant.

Markables corresponding to sets of objects are assigned the value according to the type of the elements they contain. For instance, *einige Songs* or *alle Lieder von den Beatles* will have the type song, *drei Alben* the type album, etc.

Pronouns and other anaphorically referring expressions which do not convey a type on the surface will be assigned a type according to the resolved reference. The same applies to notions like *Version* or *Titel* when referring to songs, albums, or whatever entities, as well as to pronominal adverbs like *darin*, *darauf*, etc.

(192) **MP3**: *[Ich] habe [drei Alben]album gefunden.*

MP3: I have three albums found
MP3: I’ve found three albums.

**MP3**: *[Sie]album sind [von den 70ern].*

MP3: They are from the 70ies
MP3: They are from the 70ies.

(193) **MP3**: *[Welches Lied]song willst [Du] [von den Beatles]?*

MP3: which song wish you from the Beatles
MP3: Which Beatles song do you want?

**Subj**: *[Ein beliebiges]song.*

Subj: An arbitrary

Subj: An arbitrary one.

(194) **MP3**: *[Ich] habe [ein Album]album gefunden.*

MP3: I have one album found
MP3: I’ve found one album.

**MP3**: *[Darauf]event gibt [es] [15 Lieder].*

MP3: Thereon gives it 15 songs
MP3: It contains 15 songs.

Discourse deixis is deictic reference to a portion of a discourse relative to the speaker’s current “location” in the discourse. Discourse deictic expressions will get the value event.

---

15We assume that all sets referred to in the SAMMIE corpora are homogenous.
(195) Subj: Kannst [Du] [mir] [die Songs [von den Beatles]] zeigen?
Subj: Can you show me the songs by the Beatles?

MP3: [Das]even mache ich.
MP3: This I do

MP3: I’ll do this

Markables that are neither domain objects nor of type participant, nor event are assigned the type other. Nonreferential NPs, e.g., expletive es as in Es regnet will be assigned the type none.

The entire list of possible values for the attribute type is:

1. song

(196) [ich] möchte [dieses Lied]
I wish this song
I’d like this song

(197) [ich] möchte [Yesterday [von den Beatles]]
I wish Yesterday by the Beatles
I’d like Yesterday by the Beatles

(198) [ich] habe [drei Songs] gefunden
I have three songs found
I’d found three songs

(199) [ich] habe [viele Titel [diesen Namens]] gefunden
I have many titles of this name found
I’d found many songs with this name

2. album

(200) [ich] möchte [das neueste Album [von Björk]]
I wish the latest album by Björk
I’d like Björk’s latest album

(201) [ich] möchte [eine Version des Albums The Very Best [der Gruppe REM]]
I wish a version of the album The Very Best of the group REM
I’d like a version of the album The Very Best of the group REM

(202) [ich] habe [viele Titel [mit dem Namen Hits for you]] gefunden
I have many titles with the name Hits for you found
I’d found many songs with the name Hits for you

3. playlist
(203) [ich] füge [das] [zu [deiner Playlist]]
I add this to your playlist
I add this to your playlist

(204) [welche Playlist] ist [das]? 
which playlist is this
which playlist is this?

4. artist 

(205) [ich] möchte [dieses Lied [von Nena]]
I wish this song by Nena
I'd like this song by Nena

(206) [ich] möchte [Yesterday [von den Beatles]]
I wish Yesterday by the Beatles
I'd like Yesterday by the Beatles

(207) [ich] möchte [dieses Lied [von der Sängerin Nena]]
I wish this song by the singer Nena
I'd like this song by the singer Nena

(208) [ich] möchte [Yesterday [von der Gruppe The Beatles]]
I wish Yesterday by the band The Beatles
I'd like Yesterday by the Beatles

5. name 

(209) [ich] möchte [dieses Lied [von der Sängerin [namens Nena]]]
I wish this song by the singer named Nena
I'd like this song by Nena

(210) [ich] möchte [Yesterday [von der Band [mit dem Namen The Beatles]]]
I wish Yesterday by the band with the name the Beatles
I'd like Yesterday by the band called the Beatles

(211) [ich] möchte [das Lied mit dem Titel Yesterday]
I wish the song with the title Yesterday
I'd like the song called

(212) [ich] möchte [das Album mit dem Namen Weihnachten]
I wish the album with the name Weihnachten
I'd like the album called Weihnachten

6. genre 

(213) Kannst [du] [mir] [Lieder [aus dem Genre Elektronik] anzeigen can you me songs from the genre Elektronik show
Can you show me songs from the genre Elektronik
7. length

(214) [dieses Lied] dauert [5 Minuten].
this song lasts 5 minutes
this song takes 5 minutes

(215) [welche Gesamttspielauer] hat [das Album]?
which total length has the album
which is the total length of the album?

8. year

(216) [ich] möchte [eine Version [von 1963]]
I wish a version from 1963
I’d like a version from 1963

9. participant

(217) [ich] möchte [eine Version [von 1963]]
I wish a version from 1963
I’d like a version from 1963

(218) [was] möchtest [du]?
what wish you
what do you like?

(219) Möchtest [Sie] [[das Album] oder [das Lied]]?
wish you the album or the song
Do you want the album or the song?

(220) [MP3-Player], spiele [das]!
MP3 player play this
MP3 player, play this!

10. event

(221) Kann [ich] sonst noch [irgendwas] tun?
can I else more something do
Can I do something else?

(222) [Das] habe [ich] gemacht.
this have I done
I’ve done that

11. other
(223) [ich] möchte [eine Mistgabel]
  I wish a dung fork
  I’d like dung fork

(224) [ich] habe [keine Zeit]
  I have no time
  I have no time

(225) [ich] habe [das Lied] [in der Datenbank] nicht gefunden
  I have the song in the database not found
  I haven’t found the song in the database

12. none

(226) [Es] regnet.
  it rains
  It's raining

(227) [Es] gibt [kein Lied] [in der Datenbank].
  It gives no song in the database
  There is no song in the database.

**typeMention**

If the markable is a Domain Object and contains an explicit mention of one of the discourse entity types listed under deType the value of this attribute is **true**, otherwise **false**.

1. **true** for: song, album, playlist, artist, name, genre, length and year

(228) [ich] möchte [dieses Lied]
  I wish this song
  I’d like this song

(229) [ich] möchte [den Song Yesterday [von den Beatles]]
  I wish the song Yesterday by the Beatles
  I’d like the song Yesterday by the Beatles

(230) [das Lied [aus dem Jahr 2002]]
  the song from the year 2002
  the song from 2002

2. **false** for: participant, event, other and none

(231) [ich] möchte [dieses]
  I wish this
  I’d like this
(232) [ich] möchte [Yesterday [von den Beatles]]
    I wish Yesterday by the Beatles
    I’d like Yesterday by the Beatles

(233) [ich] möchte [das [von 1963]]
    I wish that from 1963
    I’d like that from 1963

(234) [ich] möchte [eine Version [von 1963]]
    I wish a version from 1963
    I’d like a version from 1963

(235) [das Lied] aus [2002]
    the song from 2002
    the song from 2002

(236) [Es] regnet.
    it rains
    It’s raining

(237) [Es] gibt [kein Lied] [in der Datenbank].
    It gives no song in the database
    There is no song in the database.

properNameMention

If the markable is a Domain Object and contains an explicit mention of its proper name the value of this attribute is true, otherwise false. The value “false” does not mean that a markable that does not contain any mention of a proper name (cf., for instance, examples 240 versus 242).

1. true

(238) [ich] möchte [das Lied Yesterday [von den Beatles]]
    I wish the song Yesterday by the Beatles
    I’d like the song Yesterday by the Beatles

(239) [ich] möchte [Yesterday].
    I wish Yesterday
    I’d like Yesterday.

(240) [ich] möchte [das Lied Yesterday [von den Beatles]]
    I wish the song Yesterday by the Beatles
    I’d like the song Yesterday by the Beatles

2. false

(241) [Ich] möchte [dieses Lied].
    I wish this song
    I’d like this song.
(242) [Ich] möchte [den Song [von den Beatles]].  
I wish the song by the Beatles  
I'd like the song by the Beatles.

(243) [Ich] möchte [eine Version [von 1963]].  
I wish a version from 1963  
I'd like a version from 1963.

**modification**

This feature stores the information whether a markable is a complex one in the sense that it contains other information than the type of the discourse entity and/or its proper name. When there is some additional information, we distinguish whether it is domain-specific information (MP3 domain) or general information. As domain-specific information we consider anly information corresponding to one of the domain-specific types listed above under deType.

Discourse entities that are not of one of the domain object types will be assigned the value none.

1. none

(244) [ich] möchte [das Lied Yesterday]  
I wish the song Yesterday  
I'd like the song Yesterday

(245) [ich] möchte [dieses Lied]  
I wish this song  
I'd like this song

(246) [ich] möchte [Yesterday]  
I wish Yesterday  
I'd like Yesterday

(247) [ich] möchte [dies]  
I wish this  
I'd like this

(248) [ich] möchte [den Song [von den Beatles] [aus diesem Album]]  
I wish the song by the Beatles from this album  
I'd like the song by the Beatles from this album

(249) [ich] möchte [den Song [von den Beatles] [aus diesem Album]]  
I wish the song by the Beatles from this album  
I'd like the song by the Beatles from this album

2. domain – the markable of a domain-specific type contains additional information, besides type and/or name, of the domain-specific type: song, album, playlist, artist, name, genre, length and year
(250) [ich] möchte [das Lied Yesterday [von den Beatles]]
I wish the song Yesterday by the Beatles
I’d like the song Yesterday by the Beatles

(251) [ich] möchte [das Lied von den Beatles]
I wish the song by the Beatles
I’d like the song by the Beatles

(252) [ich] möchte [das Lied [aus diesem Album]]
I wish the song from this album
I’d like the song from this album

(253) [ich] möchte [den Song [von den Beatles] [aus diesem Album]]
I wish the song by the Beatles from this album
I’d like the song by the Beatles from this album

(254) [ich] möchte [eine Version [von 1963]]
I wish a version from 1963
I’d like a version from 1963

3. nondomain – the markable of a domain-specific type contains additional information, besides type and/or name, but it’s none of the domain-specific type (song, album, playlist, artist, name, genre, length and year)

(255) [ich] möchte [das Lied Yesterday, das [Du] [mir] vorhin gezeigt hast]
I wish the song Yesterday which you me before shown have
I’d like the song Yesterday you shown be right before

(256) [ich] möchte [die neueste Version]
I wish the latest version
I’d like the latest version

(257) [ich] möchte [drei Lieder]
I wish three songs
I’d like three songs

4. both – the markable of a domain-specific type contains additional information – both domain-specific and general --, besides its own type and/or name.

(258) [ich] möchte
I wish
[das Lied Yesterday [von den Beatles], das [Du] [mir] vorhin gezeigt hast]
the song Yesterday by the Beatles which you me before shown have
I’d like the song Yesterday by the Beatles you shown be right before

(259) [ich] möchte [die neueste Version [aus dem Album Loulou]]
I wish the latest version from the album Loulou
I’d like the latest version from the album Loulou
npForm

This feature stores the information about the surface form of a markable – pronominal or not –, and about the presence of determiners if it is non-pronominal.

Note: When multiple values apply, choose the more specific one. For example, in the case of definite NPs, we differentiate between defNP as in *das Lied* and thisNP as in *dieses Lied*. You should assign the value thisNP to *dieses Lied*. This applies to expletives and other specialized cases.

1. persPron

(260)  [Ich] habe [es] gefunden
I have it found
I’ve found it

2. persPronExpl

(261)  [es] regnet
it rains
it is raining

3. possPron

(262)  [ich] füge [das] [zu [deiner] Playlist]
I add this to your playlist
I add this to your playlist

4. reflPron

(263)  [Das] läßt [sich] machen
this lets itself make
this is possible

5. demPron

(264)  [Ich] habe [dieses] gefunden
I have this found
I’ve found this

(265)  [Ich] habe [das] gefunden
I have this found
I’ve found this

6. bareNP
(266) [Ich] habe [viele Lieder] gefunden
   I have many songs found
   I’ve found many songs

7. defNP

(267) [Ich] habe [das Lied Yesterday] gefunden
   I have the song Yesterday found
   I’ve found the song Yesterday

(268) [die vielen Lieder, die [ich] gefunden habe,] kann [ich] nicht anzeigen
the many songs that I found have can I not show
I can not show the many song I’ve found

8. thisNP

(269) [Ich] habe [dieses Lied] gefunden
   I have this song found
   I’ve found this song

9. indefNP

(270) [Ich] habe [ein Lied mit dem Namen Yesterday] gefunden
   I have a song with the name Yesterday found
   I’ve found a song named Yesterday

10. pronAdv

(271) [Ich] habe [darin] [ein Lied [mit dem Namen Yesterday]] gefunden
   I have in it a song with the name Yesterday found
   I’ve found there a song named Yesterday

11. coordNP

(272) [Ich] habe [ein Lied] und [ein Album] gefunden
   I have a song and an album found
   I’ve found a song and an album

12. headlessNP

(273) [Ich] will [die zweite]
   I will the second
   I’d like the second one

13. other: gerundium and alike
quantNP

This feature stores the information whether a markable contains a quantifier or not.

1. true

(274)  [ich] habe [drei Lieder] gefunden
I have three songs found
I’ve found three songs

(275)  [ich] habe [ein Lied] gefunden
I have a song found
I’ve found a song

(276)  [ich] habe nur [ein Lied] gefunden
I have only a song found
I’ve found only a song

(277)  [ich] habe [Alben] aber [keine Lieder] gefunden
I have albums but no songs found
I’ve found albums but no songs

2. false

(278)  [ich] habe [die Lieder] gefunden
I have the songs found
I’ve found the songs

(279)  [ich] habe [Yesterday [von den Beatles]] und [Yesterday [von Nena]] gefunden
I have Yesterday by the Beatles and Yesterday by Nena found
I’ve found Yesterday by the Beatles and Yesterday by Nena

(280)  [ich] habe [Yesterday [von den Beatles]] und [Yesterday [von Nena]] gefunden
I have Yesterday by the Beatles and Yesterday by Nena found
I’ve found Yesterday by the Beatles and Yesterday by Nena

gFunction

This feature stores the information about the grammatical function of a markable. At the clause level, we differentiate between arguments (subject, dirObject, indirObject, and prepObject) and adjuncts (vpModifier). Modifiers of nominal heads get the value npModifier.

Note: In cases in which it is difficult to decide between a modifier and an argument, you should assign the value vpModifier.

The values of this feature are the following:

1. subject
(281) [Ich] habe [das Lied Yesterday] gefunden
   I have the song Yesterday found
   I've found the song Yesterday

(282) [das Lied Yesterday] wurde gefunden
   the song Yesterday is found
   the song Yesterday has been found

2. dirObject

(283) Ich habe das Lied Yesterday gefunden
   I have the song Yesterday found
   I've found the song Yesterday

3. indirObject

(284) Ich spiele dir das Lied Yesterday
   I play you the song Yesterday
   I'll play the song Yesterday for you

4. prepObject

(285) ich füge das zu deiner Playlist
   I add this to your playlist
   I’m adding this to your playlist

5. vpModifier

(286) ich habe das in dem Album Loulou gefunden
   I have this in the album Loulou found
   I’ve found this in the album Loulou

6. npModifier

(287) ich möchte den Song Yesterday von den Beatles
   I wish the song Yesterday by the Beatles
   I’d like the song Yesterday by the Beatles

(288) ich möchte eine Version von 1963
   I wish a version from 1963
   I’d like a version from 1963
informationStatus

This feature stores the information about the relationship of a discourse entity to other discourse entities in the context. A discourse entity might have been mentioned in the preceding dialogue – hence being discourse old –, or not – discourse new –, or can be inferred given some mentioned discourse entity.

1. old: a discourse entity is old if it has been mentioned in the preceding dialogue, or if it is one of the personal, reflexive or possessive forms of the first and second person. Moreover, addressing expressions like MP3-Player are to be annotated as discourse old, too. Also markables expressing generic agents are to be annotated as discourse old: man (lately also frau), wir, du, jeder, die Menschen, etc.

(289) Subj: [MP3-Player] [ich] möchte [eine Liste] erstellen [mit drei Liedern]
Subj: MP3 player I wish a list make with three songs
Subj: MP3 player I’d like to make a list with three songs

MP3: how shall then the playlist be called that you make wish

MP3: how to call the playlist you like to make

(290) [Man] will [Rock] hören
one wants rock hear
one wants to hear rock music

(291) [Ich] habe [mich] [dafür] entschieden
I have myself for this decided

I’ve decided for this

2. mediated/inferrable: a discourse entity is mediated if it has itself not been mentioned earlier in the dialogue, but it is already familiar to the hearer because of a bridging inference given another discourse entity or general knowledge.

(292) Subj: Spiele [etwas]!
Subj: Play something
Subj: Play something!

MP3: Kannst [Du] [mir] wenigstens [die Musikrichtung] nennen?
MP3: Can you me at least the music style say

MP3: Can you tell me at least which music style?

3. new: a discourse entity is new if it occurs in the discourse for the first time, hence does not fall in neither of the categories above.
(293)  Subj: [ich] möchte [eine Liste] erstellen [mit drei Liedern]
Subj: I wish a list make with three songs

Subj: I’d like to make a list with three songs

MP3: how shall then the playlist be called that you make wish

MP3: how to call the playlist you like to make
4.9 Clarification Requests [USAAR]

4.9.1 Annotation Purpose

Good clarification strategies in dialogue systems help to ensure and maintain mutual understanding and thus play a crucial role in robust conversational interaction. To get a better understanding of the use, cause and effects of multimodal clarification strategies employed by human wizards, we are annotating Clarification Requests (CRs) in the SAMMIE-2 corpus.

We define a clarification sub-dialogue to be a triple of the utterance being the antecedent, the CR itself, and the reply given by the addressee. The CR and the reply get annotated with additional attributes.

4.9.2 Annotation Instructions

Overall Procedure

For annotating Clarification Requests (CRs) the following steps are required:

1. First, you should identify CRs (isCR).
2. Once you decided that an utterance is a CR you annotate the source and severity for that CR (CRsource, CRseverity).
3. After that you are searching for the antecedent, i.e which utterance cause the understanding problem, mark it, and assign a antecedent type (antType).
4. Then you search for the reply given to that CR. You mark the reply and specify the reply type (replyType).

In sum, a CR object is a triple of three related utterances; one utterance being the CR itself, an antecedent (what caused the CR) and a reply to that CR. As an annotation aid decision trees are provided following the legend as explained in Figure 4.3.
Identify CRs (isCR)

A clarification request (CR) signals a problem with understanding a previous utterance and requests repair of that problem. The grounding status of the problematic utterance is pending.

CRs are attempts to resolve a problem with understanding by 'looking back' to the problematic utterance and 'look forward' to elicit an answer that would resolve this problem. CRs look forward in the dialogue in terms that they impose the obligation on the other dialogue participant to provide a reply. Unlike the situation where the speaker explicitly acknowledges what he/she understood, when asking a CR the speaker is only satisfied if the other dialogue participant provides some feedback. Therefore what we mark to be the CR reply (as discussed in Section 4.9.2) is important for CR identification as well.

CRs look backwards to the previous discourse in terms that they signal some lack of understanding a previous utterance (what we call the antecedent). Unlike normal questions, CRs are not concerned with acquiring new propositional information, but rather enquire about aspects of previous utterances. It is most important that unlike normal questions they do not make requests about the state of the world, but about aspects of the previous discourse which are presupposed by the other dialogue participant. Therefore what we mark to be the CR antecedent (as discussed in Section 4.9.2) is important for CR identification as well.

CRs can address several different types of understanding problems, from acoustic understanding to reference resolution. They can also take many different surface forms. In the following examples we will see that he identification of an utterance as CR depends on contextual information.

In examples 4.1 and 4.2, the dialogue participant B requests information about a part of the previous discourse, but in each case the function is different. Only in example 4.1 the contribution of B can be considered a CR because the knowledge of a unique referent is presupposed by speaker A (using the definite article the). In A4.2 the speaker does not presuppose that there is a unique referent, but it's rather requesting for new propositional information.

Example 4.3 is also a CR, but in contrast to example 4.2, example 4.3 only indirectly imposes the obligation to provide some more constraints. The same holds for example 4.4 where the request is for a new/revised user goal though, rather than some more constraint.

For utterances where the CR initiator makes a suggestion, the antecedent determines whether the utterance is a CR or whether the CR initiator introduces new material. In examples 4.5 and 4.6 speaker B is asking a CR, requesting clarification concerning the interpretation of the respective verb phrase. In example 4.7 he is requesting to clarify a missing verb phrase. Note, that in examples 4.5, 4.6, and 4.7 speaker B already requested an action which speaker B tries to clarify by suggesting an interpretation. Whereas in example 4.8 the suggestions to play the song a second time comes from speaker B and therefore should not be labelled as CR.

To help to resolve these problems, we provide three tests: The forward looking, the backward looking and the propositional content test (see figure 4.4).

(4.1)  **A1:** Please play the song Layla.

          **B1:** The title exist 127 times. From which album?

(4.2)  **A2:** Is there a song called Layla?

          **B2:** The title exist 127 times. From which album?
Figure 4.4: Identify CRs

Forward looking test: Does the utterance impose an obligation on the other dialogue participant to provide an answer?

no → noCR → no

Backward looking test: Does the utterance signal a lack of understanding?

no → noCR → no

yes → yes

Propositional content test: Does the speaker of the problematic utterance assume that the utterance provides sufficient information to be understood?

no → noCR → no

yes → yes

CR
(4.3) A3: Please play the song Layla.
   B3: The title exist 127 times.

(4.4) A4: Please play the song Layla.
   B4: Sorry, I couldn’t find the song in the data base.

(4.5) A5: Take the song Layla.
   B5: What should I do with it?

(4.6) A6: I’d like to hear ‘Layla’.
   B6: Should I add it to the playlist?

(4.7) A7: Ah, playlist . . .
   B7: Should I build a new playlist?

(4.8) A8: I’d like to hear ‘Layla’.
   B8: OK. [music]
   B8: Would you like to hear it a second time?
      (no CR!)

<table>
<thead>
<tr>
<th>Value</th>
<th>Value description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td>Utterance labeled as clarification request</td>
</tr>
<tr>
<td>noCR</td>
<td>Utterance not labeled as clarification request</td>
</tr>
</tbody>
</table>

**Classify CRs**

**CR Problem Source (CRsource)** The problem source of the clarification request describes the type of understanding problem which caused the need to clarify.

From a theoretical perspective, the model of grounding (e.g. [Cla96]) describes several levels of understanding involved in communication, each of which is a possible locus for communication problems. These levels are building up on each other, including low level communication, such as establishing contact, up to higher level understanding such as agreeing upon a common goal. CRs perform negative feedback actions at one or more levels, indicating lack of ability to perceive (or understand etc.) the message.

The problem source of a CR cannot as easily be seen from its surface form and requires some interpretation. We use the reply of the CR addressee, and the reaction of the CR initiator to that reply as a guide for the interpretation that was chosen by the dialogue participants. Hence what we annotated as problem source could more properly be called “mutually agreed upon interpretation of the CR”, and that is not necessarily what the CR initiator might initially have had in mind.

To facilitate annotation we propose a three step iterative process: the self-monitoring test, followed by the understanding of the addressee test, and the satisfaction of the initiator test (see table 4.4).
The CR initiator has doubts whether the addressee still pays attention. The reply to that CR helps to establish contact (as in example 5).

The CR initiator has a problem with the 'acoustic' understanding, i.e. in the transcribed user utterance words are deleted in such a way that the wizard has problems understanding it. The reply to that CR helps the acoustic understanding (as in example 6).

The CR initiator has a problem with resolving the reference in the database. From the content problematic utterance it can be assumed that the CR addressee had a specific object in mind, but the CR initiator cannot find a (unique) referent in the database. The reply to that CR helps to find a (unique) referent (as in example 7).

The CR initiator has a problem with identifying the goal of the other dialogue participant. The CR aims to elicit further attributes that refine the user’s goal. It cannot be assumed that the CR addressee had a specific (unique) object in mind. The reply to that CR helps to refine the user's goal (as in example 8).

In this case it cannot be seen from the surface form of the problematic utterance whether the CR addressee had a specific goal in mind or not. The CR might enquire about refining the goal or about uniquely identify the referent. The reply can serve both, goal refinement or referent refinement (as in example 9).

The CR initiator has problems in understanding the intention behind the utterance, i.e. he does not know what the other dialogue participant expects him to do. The reply to that CR helps to clarify the intuition of the speaker (as in example 10).

Figure 4.5: Annotate CR Source

Version: February 13, 2006 (Final version) Distribution: Public
Self-monitoring test: Get an idea what kind of understanding problem the CR initiator was dealing with by just reading the CR.

Understanding of the addressee test: Examine the response(s) of the other dialogue participant to determine which reading was chosen by the addressee.

Satisfaction of the initiator test: Examine the reaction of the CR initiator to that reply. If the initiator does not seem to be satisfied with the answer, he might reformulate the CR. The source which is finally indicated and answered is to be the mutual agreed interpretation of the clarification.

Table 4.4: Tests to identify the problem source of a CR

(4.9) **A**: Hello? Are you still there?
      **B**: Yes.

(4.10) **A**: Delete the song from (NOISE).
       **B**: From where?
       **A**: From the current playlist.

(4.11) **A**: Please play Nevermind.
       **B**: I found 127 songs with that name.
       **A**: The one by Nirvana.

(4.12) **A**: I’d like to hear a song by Madonna.
       **B**: There are 127 songs by Madonna in the data base. Do you know the title?
       **A**: No. Show me a selection.

(4.13) **A**: I’d like to hear this Madonna song.
       **B**: There are 127 songs by Madonna in the data base. Do you know the title?
       **A**: No. Show me a selection.

(4.14) **A**: Select the first three songs.
       **B**: What should I do with these songs?
       **A**: Add them to the playlist.
CR Problem Severity (CRseverity) CR problem severity describes which type of feedback the CR-initiator requests from the other dialogue participant in terms of how many hypothesis are present in the CR.

The type of feedback that is requested from the addressee is also called the severity because it indicates how severe the problem is in terms of how many hypotheses are available. If the CR initiator has no hypothesis he might ask for repetition or further elaboration as in example 4.15 on the acoustic level or in example 4.16 on the reference level. Having a vague hypothesis he might want the other dialogue participant to confirm that hypothesis as in example 4.17. If the CR initiator has too many hypotheses he might ask for disambiguation by presenting his (several) hypothesis as in example 4.18 (graphical disambiguation) or 4.19 (disambiguation by naming the choices).

For some layers some severities are very unlikely. For example you hardly will find the request to disambiguate on the contact layer.

Note, that for assigning severity the source layer already needs to be identified. It might even happen that depending on the problem source the same utterance gets annotated having a different severity. In example 4.20 and 4.21 both CRs by B request to resolve an understanding problem caused by too many matches in the data base. In example 4.20 the problem source is an understanding problem on the reference level (as we can assume that A had an specific object in mind). Presenting several hypothesis on the reference level, the CR in 4.20 should get annotated as disamb. For the CR in 4.21 we have a goal problem (as we cannot assume to clarify an specific object). The CR here is indirectly asking for further elaboration to narrow down the goal of A and should therefore be annotated as repet/elab.

To facilitate annotation we propose two tests: the source test and the hypotheses test (see table 4.5).

| Source test: Understanding of the addressee test | Which problem source layer is addressed? 
Examine the response(s) of the other dialogue participant to determine which reading was chosen by the addressee. |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypotheses test:</td>
<td>How many hypotheses on that layer are (graphically or verbally) presented in the CR?</td>
</tr>
</tbody>
</table>

Table 4.5: Tests to identify the severity of a CR

(4.15) A: Excuse me? What song?

(4.16) A: I found 50 matches in the database. Do you know the artist?

(4.17) A: ’Layla’ by Eric Clapton?

(4.18) A: I found 50 matches in the database. \([displays list]\)

(4.19) A: ’Layla’ by Eric Clapton or by Derek and the Dominos?

(4.20) A: Please play Nevermind.

    B: I found 127 songs with that name.

    A: The one by Nirvana.
Figure 4.6: Annotate CR Severity

The CR initiator asks for repetition or elaboration. No hypothesis about the understanding on the understanding level indicated is presented (as in example 10 and 11).

The CR initiator presents a hypothesis in the CR and asks to confirm this hypothesis. Asking for confirmation indicates that the CR initiator has an unsure hypothesis about the understanding (as in example 12).

The CR initiator asks to disambiguate several hypotheses. Asking for disambiguation means that the CR initiator presents more than one hypothesis in the CR (as in example 13 and 14).

A: I'd like to hear a song by Madonna.
B: There are 127 songs by Madonna in the data base.
A: No. Show me a selection.

Identify the antecedent

The antecedent is defined as a pointer to the previous problematic utterance the CR initiator wants to clarify.

**CR Antecedent Type** antType

The feature antType describes the grammatical category the CR requests to clarify.
The extent of the CR can be a whole sentence (S) (as in example 4.22), a part of a sentence being a phrase type of a syntactic category (PP, NP, VP), or a word being a lexical category. For lexical categories mark the part-of-speech (ADJ, ADV, CONJ, DET, INTERJ, PRON, PREP, NN, V, other). In example 4.23, B clarifies the determiner (DET).

If the CR is a NP or PP phrase type it might be further specified, listing the type of object contained in the phrase. In example 4.24 and 4.25, the phrase type is an NP one time clarifying a song object (4.24) and in (4.25) clarifying a playlist object. In example 4.26 the object type is also playlist, but the phrase type is PP. In example 4.27 the phrase type is VP.

Note, that the antecedent type is influenced, but not necessarily determined by the problem source indicated in the CR.

(4.22) A: Play Layla.
    B: Sorry, could you please repeat?

(4.23) A: List all songs.
    B: What do you mean with 'all'?

    B: I found ten different versions of Layla.

(4.25) A: Show me the playlist.
    B: Which one?

(4.26) A: Please delete the song Layla from a playlist.
    B: The song Layla exists on two playlists, 'Classics' and 'Best of Eric'.

(4.27) A: I like to hear Layla.
    B: Should I play it to you?

Identify and classify the reply

Identify the CR reply (replyType) Reply points to the utterance containing the reply (or answer) given to the CR. Reply points to the utterance containing the reply (or answer) given to the CR.

CR Reply Type (replyType) CR reply classifies the reply to the CR in terms of information provided.

CR reply classifies the reply to the CR, and not the CR itself. The classification is done in terms of what kind of information is provided in the CR. For example if it is a short confirmation in terms of a yes/no answer, or if it provides additional information by elaboration. It describes the information gain with respect to the clarification itself.

Note, that the reply type is influenced, but not necessarily determined by the severity indicated in the CR. As shown in example 4.28 these answers can also be multimodal. Actions in both modalities should be considered when annotating (if available). Furthermore, the reply should be annotated by it’s maximal information content, i.e. in example 4.28 the second sentence of the reply provides additional information and therefore this utterance should be annotated as “answer additional” (ans-add).
Figure 4.7: Annotate antecedent type

antType
The feature “antType” describes the grammatical category the CR requests to clarify.

Sentence (S)
As in example (?22)

Phrase

Word

PP
As in example (26)

NP
As in example (24,25)

VP
As in example (27)

other

Type of object contained in the phrase

artist
album
song
playlist

Other
As in example (23)

ADJ
ADV
CONJ
DET
INTERJ
PRON
PREP
NN
V

others
(4.28) A: Does this list contain the song?
B: Yes. It’s number four. [/clicks on the fourth item]

(4.29) A: Should I add these songs to the playlist?
B: Yes. Would be great.

(4.30) B: Please, create a playlist with the name “Eric”.
A: Should I create a new playlist?
B: Yes, create a playlist with the name “Eric”.

(4.31) A: Please mark the first two songs.
B: Sorry?
A: Please mark “Layla” and “Lola”.

(4.32) A: This song exists 127 times.
B: Search for the one by Eric Clapton

(4.33) A: This song exists 127 times.
B: I don’t care. Just play any.

B: I’m sorry. I can’t find that in the data base.
A: OK. Then just any version.

(4.35) A: Please select the first four songs and put them on a playlist.
B: I’m sorry?
A: Please select the first four songs by Eric Clapton.

(4.36) A: I found 127 titles with that name.
B: It’s number four on the list.

(4.37) A: I found 127 titles with that name.
B: Show me the first 10.

(4.38) A: I am searching for the title ’Venus’.
B: Sorry, I cannot find a title with this name.
A: Please show me all titles of the album ’Walkie Talkie’.

(4.39) A: Should I add the song to the playlist?
B: No, delete it.
Figure 4.8: Annotate CR Reply Type (please also see next page)
(4.40) **A1:** Please play "Secret".

**B1:** Sorry, there are many hits in the data base with that name.

**A2:** Please show me a selection.

**B2:** OK. [shows screen output]

**General Remarks**

Please note that a reply to a CR can be the antecedent for another CR. In example 4.41 A1 is the antecedent for B1. A2 is the answer to B1 as well as the antecedent for B2.

(4.41) **A1:** Please select four titles.

**B1:** Sorry, I did not get that.

**A2:** Please select the first four titles on the list.

**B2:** And what would you like to do with those items?
4.10 Task-Layer Features [USAAR]

4.10.1 Annotation Purpose

The conversations between wizard and subject in our study are task-directed. Each subject had to fulfil 4 tasks, 2 with operating the driving simulator and 2 without. For evaluation we are interested in quality and quantity of the dialogues with respect to the task following an approach by [WP01]. Therefore the task boundaries get marked with the topic ID (we will refer to this as “section”). Sections, typically but not necessarily correspond 1-1 to tasks because tasks can be resumed. In that case we mark the two attempts with the same task ID. The task layer contains a second unit, sub-tasks which can be associated with sections via time stamps.

4.10.2 Annotation Instructions

Overall Procedure

For annotating task layer features the following steps are required:

1. First, for every session you are annotating check the session ID and look up the associated task ID in the appendix 4.10.2.

2. Read the respective task descriptions in appendix 4.10.2.

3. Read through the whole session, mark the task boundaries and assign the respective task ID to the first dialogue act concerned with that task.

4. The part you marked with a task ID we will refer to as “sections”. Sections, typically but not necessarily correspond 1-1 to tasks because tasks can be resumed. In that case you mark the two attempts with the same task ID.

5. Annotate whether the user was operating the driving simulator during that task (isDriving).

6. For each section, mark the subsection boundaries according to the type of a subtask (subtaskType).

7. For each section, annotate the (numberOfConstraints).

8. Then you assign a label for (taskTermination).

9. Finally, evaluate the (actualTaskCompletion) for each section.

islDriving

Is driving describes whether the user was operating the driving simulator while solving that task. Values are true and false.

<table>
<thead>
<tr>
<th>Value</th>
<th>Value description</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>The user was operating the driving simulator while solving the task.</td>
</tr>
<tr>
<td>false</td>
<td>The user was <em>not</em> operating the driving simulator while solving the task.</td>
</tr>
</tbody>
</table>
**subtaskType**

**Subtask type describes steps towards the (user’s) goal which is solved by both dialogue participants.** The subtask type labels the region of a subsection. in the DATE scheme (Walker and Passonneau, 2001). Subtasks are the subsequent number of actions required to accomplish a goal. Subtasks are joint actions and it doesn’t matter which one of dialogue participant starts a subtask. A subtask begins with the first dialogue act addressing some type of sub-task defined in the value list. It ends with the last dialogue act concerned with solving that subtask before the dialogue participant change to a different subtask. Subtasks regions can overlap. Subtasks can involve spoken actions as well as displaying and selecting items on the screen.

<table>
<thead>
<tr>
<th>Value</th>
<th>Value description</th>
</tr>
</thead>
<tbody>
<tr>
<td>search-song</td>
<td>Dialogue actions concerned with searching for one or more song(s) in the database or in the playlists, like disambiguating the title, defining the artist, etc.</td>
</tr>
<tr>
<td>delete-song-from-playlist</td>
<td>Dialogue actions concerned with deleting a song from a playlist, e.g. specifying the name of the playlist.</td>
</tr>
<tr>
<td>add-song-to-playlist</td>
<td>Dialogue actions concerned with adding a song to a playlist.</td>
</tr>
<tr>
<td>play</td>
<td>Dialogue actions concerned with playing a song, an album, or a playlist.</td>
</tr>
<tr>
<td>mark</td>
<td>Dialogue actions concerned with marking an item on the display.</td>
</tr>
<tr>
<td>length</td>
<td>Dialogue actions concerned with enquiring about the length of a song, an album, or a playlist.</td>
</tr>
<tr>
<td>search-album</td>
<td>Dialogue actions concerned with searching for an album in the database or in the playlists, like disambiguating the title, defining the artist, etc.</td>
</tr>
<tr>
<td>build-playlist</td>
<td>Dialogue actions concerned with building and naming a playlist.</td>
</tr>
<tr>
<td>aside</td>
<td>Dialogue actions which are not concerned with the actual task and are <em>not</em> directed from wizard to user or from user to wizard. Those are for example self speech of the user or the wizard, utterances by the experimentator, or interaction between wizard and experimentator.</td>
</tr>
<tr>
<td>not-about-task</td>
<td>Dialogue actions which are not concerned with the actual task and are directed from wizard to user or from user to wizard. Those are for example greetings, or other social or situational actions.</td>
</tr>
<tr>
<td>other</td>
<td>None of the values listed above do apply.</td>
</tr>
</tbody>
</table>

**taskTermination**

**This feature describes how the task terminated.** In contrast to actualTaskCompletion, this feature describes not the task achievement but the circumstances how a task was terminated.
<table>
<thead>
<tr>
<th>Value</th>
<th>Value description</th>
</tr>
</thead>
<tbody>
<tr>
<td>no-problems</td>
<td>The user finished the task without any problems. Both the user believes that she/he fulfilled the goal.</td>
</tr>
<tr>
<td>gave-up</td>
<td>The user abruptly abandoned the task and did not resume it at a later point.</td>
</tr>
<tr>
<td>interrupted</td>
<td>Task was interrupted by the experimentator. This can be due to time constraints or technical problems.</td>
</tr>
<tr>
<td>other</td>
<td>None of the values listed above do apply.</td>
</tr>
</tbody>
</table>

**numberOfConstraints**

The **number of task constraints** describe how many constraints the user needs to meet to complete the task. The number and type of task constraints are listed in appendix B.

**actualTaskCompletion**

Actual task completion describes whether the task was completed according to the task descriptions.

<table>
<thead>
<tr>
<th>Value</th>
<th>Value description</th>
</tr>
</thead>
<tbody>
<tr>
<td>exact</td>
<td>A task was completed exactly if all the constraints in the task definition are met.</td>
</tr>
<tr>
<td>{1...6}</td>
<td>Any scenario completion indicates that some constraints were not met. The argument {1...6} indicates the number of task constraints which were met. Note that the maximum number of task constraints which cannot be met is equal to numberOfConstraints (although the annotation tool will not reflect that fact). Not meeting constraints can be due to the inattentiveness of the user or to a misunderstanding by the wizard and happens during the dialogue. Incomplete task completion can also be due to early termination of the dialogue.</td>
</tr>
<tr>
<td>other</td>
<td>None of the values listed above do apply.</td>
</tr>
</tbody>
</table>
### Sessions and associated Task IDs

<table>
<thead>
<tr>
<th>Session ID</th>
<th>Task IDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>S.01</td>
<td>E.01 E.02 E.03 E.04</td>
</tr>
<tr>
<td>S.02</td>
<td>E.05 E.06 E.07 E.08</td>
</tr>
<tr>
<td>S.03</td>
<td>E.09 E.10 E.11 E.12</td>
</tr>
<tr>
<td>S.04</td>
<td>E.13 E.14 E.15 E.16</td>
</tr>
<tr>
<td>S.05</td>
<td>E.03 E.04 E.21 E.22</td>
</tr>
<tr>
<td>S.06</td>
<td>E.05 E.06 E.07 E.08</td>
</tr>
<tr>
<td>S.07</td>
<td>E.09 E.10 E.11 E.12</td>
</tr>
<tr>
<td>S.08</td>
<td>E.13 E.14 E.15 E.16</td>
</tr>
<tr>
<td>S.09</td>
<td>E.17 E.18 E.19 E.20</td>
</tr>
<tr>
<td>S.10</td>
<td>E.21 E.22 E.23 E.24</td>
</tr>
<tr>
<td>S.11</td>
<td>E.01 E.02 E.03 E.04</td>
</tr>
<tr>
<td>S.12</td>
<td>E.17 E.18 E.19 E.20</td>
</tr>
<tr>
<td>S.13</td>
<td>E.21 E.22 E.23 E.24</td>
</tr>
<tr>
<td>S.14</td>
<td>E.25 E.26 E.27 E.28</td>
</tr>
<tr>
<td>S.15</td>
<td>E.29 E.30 E.31 E.32</td>
</tr>
<tr>
<td>S.16</td>
<td>E.01 E.02 E.03 E.04</td>
</tr>
<tr>
<td>S.17</td>
<td>E.05 E.06 E.07 E.08</td>
</tr>
<tr>
<td>S.18</td>
<td>E.25 E.26 E.27 E.28</td>
</tr>
<tr>
<td>S.19</td>
<td>E.29 E.30 E.31 E.32</td>
</tr>
<tr>
<td>S.20</td>
<td>E.33 E.34 E.35 E.36</td>
</tr>
<tr>
<td>S.21</td>
<td>E.37 E.38 E.39 E.40</td>
</tr>
</tbody>
</table>

### Task Descriptions

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>Constraints</th>
<th>Constr. #</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.01</td>
<td>Eine Freundin hat Liebeskummer. Du willst sie bei ihrer Trauerarbeit unterstüßen und stellst ihr eine Playlist mit 3 herzzerreissenden Liebesschnulzen zusammen. Du willst auf jeden Fall “Forget him” dabei haben! Speichere die Playlist unter dem Namen “Trostpflaster” ab.</td>
<td>3 songs, song=&quot;Forget Him&quot;, built playlist, playlist=&quot;Trostpflaster&quot;</td>
<td>4</td>
</tr>
<tr>
<td>E.02</td>
<td>Löschke “Love me tender” von einer bereits im System enthaltenen Playlist.</td>
<td>song=&quot;Love Me Tender&quot;, delete from playlist</td>
<td>2</td>
</tr>
<tr>
<td>E.03</td>
<td>Erstelle eine Playlist von Madonna die ca. 10 Minuten dauert. Und speichere die Playlist unter einem beliebigen Namen.</td>
<td>song artist=&quot;Madonna&quot;, build playlist, playlist length= 10min</td>
<td>3</td>
</tr>
<tr>
<td>E.04</td>
<td>Lasse “Layla” abspielen.</td>
<td>song=&quot;Layla&quot;, play song</td>
<td>2</td>
</tr>
<tr>
<td>ID</td>
<td>Description</td>
<td>Constraints</td>
<td>Constr. #</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>E_05</td>
<td>Erstelle eine Playlist von Eric Clapton die ca. 10 Minuten dauert und speichere sie unter einem beliebigen Namen ab.</td>
<td>song artist=&quot;Eric Clapton&quot;, build playlist, playlist length=&quot;10min&quot;</td>
<td>3</td>
</tr>
<tr>
<td>E_06</td>
<td>Lasse &quot;Nevermind&quot; abspielen.</td>
<td>song=&quot;Nevermind&quot; or Album =&quot;Nevermind&quot;, play album or play song</td>
<td>2</td>
</tr>
<tr>
<td>E_07</td>
<td>Füge &quot;Tonight&quot; zu einer bereits im System enthaltenen Playlist hinzu.</td>
<td>song=&quot;Tonight&quot;, add to playlist</td>
<td>2</td>
</tr>
<tr>
<td>E_08</td>
<td>Es ist mal wieder Weihnachten. Erstelle eine Playlist mit 3 Weihnachtsliedern. Du willst auf jeden Fall &quot;Alle Jahre wieder&quot; dabei haben! Speichere die Playlist unter dem Namen &quot;Schöne Bescherung&quot; ab.</td>
<td>3 songs, song = &quot;Alle Jahre wieder&quot;, build playlist, playlist name=&quot;Schöne Bescherung&quot;</td>
<td>3</td>
</tr>
<tr>
<td>E_09</td>
<td>Du willst Deine &quot;wilde Jugend&quot; als Punk wieder aufleben lassen. Erstelle eine Playlist die ca. 10 Minuten dauert und nur Lieder von Punkgruppen enthält. Du willst auf jeden Fall &quot;Green Day&quot; dabei haben! Speichere die Playlist unter dem Namen &quot;Null Bock&quot; ab.</td>
<td>artist or song or album =&quot;Green Day&quot;, build playlist, playlist name=&quot;Null Bock&quot;, playlist length=&quot;10min&quot;</td>
<td>4</td>
</tr>
<tr>
<td>E_10</td>
<td>Füge &quot;I feel fine&quot; zu einer bereits im System enthaltenen Playlist hinzu.</td>
<td>song=&quot;I feel fine&quot;, add to playlist</td>
<td>2</td>
</tr>
<tr>
<td>E_11</td>
<td>Lösche &quot;Smile&quot; von einer bereits im System enthaltenen Playlist.</td>
<td>song=&quot;Smile&quot;, delete song from playlist</td>
<td>2</td>
</tr>
<tr>
<td>E_12</td>
<td>Erstelle eine Playlist von Robbie Williams die 3 Lieder enthält und speichere sie unter einen beliebigen Namen ab.</td>
<td>3 songs, song artist=&quot;Robbie Williams&quot;, build playlist</td>
<td>3</td>
</tr>
<tr>
<td>E_13</td>
<td>Dein Papa ist ein echter Alt-Hippie. Du willst ihm eine Freude bereiten und stellst ihm eine Playlist mit fetzigen Oldies aus den 70ern zusammen. Du willst auf jeden Fall &quot;All Along the Watchtower&quot; dabei haben! Speichere die Playlist unter dem Namen &quot;Flower Power&quot; ab.</td>
<td>song=&quot;All Along the Watchtower&quot;, build playlist, playlist name =&quot;Flower Power&quot;</td>
<td>3</td>
</tr>
<tr>
<td>E_14</td>
<td>Lösche &quot;Wish you were here&quot; von einer bereits im System enthaltenen Playlist.</td>
<td>song=&quot;Wish You Were Here&quot;, delete song from playlist</td>
<td>2</td>
</tr>
<tr>
<td>E_15</td>
<td>Erstelle eine Playlist von den Beatles die ca. 10 Minuten dauert. Und speichere die Playlist unter einem beliebigen Namen ab.</td>
<td>song artist=&quot;Beatles&quot;, build playlist, playlist length=10min</td>
<td>3</td>
</tr>
<tr>
<td>E_16</td>
<td>Lasse &quot;Naked&quot; abspielen.</td>
<td>song=&quot;Naked&quot;, play song</td>
<td>2</td>
</tr>
<tr>
<td>E_17</td>
<td>Erstelle eine Playlist von Michael Jackson die ca. 10 Minuten dauert. Und speichere sie unter einem beliebigen Namen ab.</td>
<td>song artist= &quot;Michael Jackson&quot;, build playlist, playlist length=10min</td>
<td>3</td>
</tr>
<tr>
<td>E_18</td>
<td>Lasse &quot;Believe&quot; abspielen.</td>
<td>song = &quot;Believe&quot;, play song</td>
<td>2</td>
</tr>
<tr>
<td>E_19</td>
<td>Füge &quot;White Christmas&quot; zu einer bereits im System enthaltenen Playlist hinzu.</td>
<td>song= &quot;White Christmas&quot;, add song to playlist</td>
<td>2</td>
</tr>
<tr>
<td>ID</td>
<td>Description</td>
<td>Constraints</td>
<td>Constr. #</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 songs, song = ”Enter Sandman”, build playlist, playlist name= ”Schwere Jungs”</td>
<td>4</td>
</tr>
<tr>
<td>E_21</td>
<td>Du willst dir eine Playlist zusammenstellen, die Lieder der Legenden der Rockgeschichte enthält und ca. 10 Minuten dauert. Du willst auf jeden Fall “Light my Fire” dabei haben! Speichere die Playlist unter dem Namen “Meine Helden” ab.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>song=”Light my Fire”, build playlist, playlist length= 10min, playlist name = ”Meine Helden”</td>
<td>4</td>
</tr>
<tr>
<td>E_22</td>
<td>Füge “Hey Joe” zu einer bereits im System enthaltenen Playlist hinzu.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>song=”Hey Joe”, add song to playlist</td>
<td>2</td>
</tr>
<tr>
<td>E_23</td>
<td>Lösche “Yesterday” von einer bereits im System enthaltenen Playlist.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>song=”Yesterday”, delete song from playlist</td>
<td>2</td>
</tr>
<tr>
<td>E_24</td>
<td>Erstelle eine Playlist von Queen die 3 Lieder enthält und speichere sie unter einem beliebigen Namen ab.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 songs, song artist = ”Queen”, build playlist</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 songs, song album=”Holiday”, song artist= ”Candy”, song title = ”Sunshine” or ”Summertime”, 4 different song artists, build playlist</td>
<td>6</td>
</tr>
<tr>
<td>E_26</td>
<td>Lösche “Millenium Hits” von einer im System bereits vorhandenen Playlist.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>song=”Millenium Hits”, delete song from playlist</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>song= “I got a Woman”, live version, play song</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 songs, song artist=”Elvis Presley”, different albums for all songs</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 songs, song artist=”The Doors”, different albums for all songs</td>
<td>3</td>
</tr>
<tr>
<td>E_30</td>
<td>Lasse das neuste Album von Bjork abspielen.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>album artist=”Bjork”, play album</td>
<td>2</td>
</tr>
<tr>
<td>E_31</td>
<td>Füge ”Can’t buy me love” in der live Version zu einer im System bereits enthaltenen Playlist hinzu.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>song=”Can’t buy me love”, live version, add song to playlist</td>
<td>2</td>
</tr>
<tr>
<td>E_32</td>
<td>Erstelle eine Playlist mit 3 Liedern für eine Party. Stelle sicher, dass alle Titel von verschiedenen Interpreten aus dem Genre “Electronic” stammen. Es soll auf jeden Fall das Lied “Alone in Kyoto” von “Air” dabei sein.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 songs, genre=electronic, 3 different artist, build playlist, song= “Alone in Kyoto”</td>
<td>5</td>
</tr>
<tr>
<td>ID</td>
<td>Description</td>
<td>Constraints</td>
<td>Constr. #</td>
</tr>
<tr>
<td>----</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>E_33</td>
<td>Lasse das neuste Album von U2 abspielen.</td>
<td>album artist= “U2”, play album</td>
<td>2</td>
</tr>
<tr>
<td>E_34</td>
<td>Erstelle eine Playlist mit Deinen 3 Lieblingsliedern. Achte darauf, dass die Spielzeit unter 12 Minuten liegt.</td>
<td>3 songs, build playlist, playlist length ≤ 12min</td>
<td>3</td>
</tr>
<tr>
<td>E_35</td>
<td>Lösche “Sunrise” von einer bereits im System enthaltenen Playlist.</td>
<td>song= &quot;Sunrise&quot;, delete song from playlist</td>
<td>2</td>
</tr>
<tr>
<td>E_36</td>
<td>Erstelle eine Playlist mit 3 Liedern, welche die folgenden Kriterien erfüllen: Alle Lieder sollen aus dem Genre Pop stammen; Keines der Lieder soll über 5 Minuten lang sein; Du willst auf jeden Fall das letzte Lied des Albums “The very Best” der Gruppe R.E.M. dabei haben.</td>
<td>3 songs, genre= Pop, song length ≤ 5min, song name= “The very Best”, song artist=REM, build playlist</td>
<td>6</td>
</tr>
<tr>
<td>E_38</td>
<td>Erstelle eine Playlist mit 3 Liedern, welche die folgenden Kriterien erfüllen: Alle Lieder sollen aus dem Genre Rock stammen; Keines der Lieder soll über 5 Minuten lang sein; Du willst auf jeden Fall das dritte Lied vom Album “The Wall” der Gruppe Pink Floyd” dabei haben</td>
<td>3 songs, genre= Rock, song length ≤ 5min, song = “The Wall”, song artist= “Pink Floyd”, build playlist</td>
<td>6</td>
</tr>
<tr>
<td>E_39</td>
<td>Lasse den Soundtrack zum Film “Forrest Gump” abspielen.</td>
<td>album= Soundtrack to ”Forrest Gump”, play album</td>
<td>2</td>
</tr>
<tr>
<td>E_40</td>
<td>Erstelle eine Playlist mit deinen 3 Lieblingsliedern. Achte darauf, dass die Spielzeit unter 12 Minuten liegt.</td>
<td>3 songs, build playlist, playlist length ≤ 12min</td>
<td>3</td>
</tr>
</tbody>
</table>
Bibliography


Ldc transcript conventions for spanish hub 4 (broadcast) speech, 1999.


