Draft DISC Grids and Life Cycles

The DISC Consortium

Contents

1. Introduction

The documentation below shows the state of development, at the end of Year 1 of DISC, of the DISC grids and life cycle models. The grids and life cycle models are presented at two different levels of abstraction. At the highest level are the general DISC grid and the general DISC life cycle model. At the level below are the detailed contributions to the DISC grid and life cycle entries that have been found necessary to describe the DISC exemplar systems and components from the perspective of looking at a particular aspect of SLDSs, such as speech recognition, speech generation, natural language understanding and generation, dialogue management, human factors, and system integration.

The grids and life cycles presented are in no way final. They will be consolidated following the DISC Year 1 Workshop, tested on new exemplar systems and components, discussed with SLDS developers, and finally packaged in a way which will optimise their usability to developers. For these purposes, the present collection of grids and life cycles provides a starting point. Some observations on the material as it stands are:

- that the distribution between information in the high-level grid and in the specific grids needs to be reconsidered for each aspect;

- that a sharper focus on specifically industrial development and evaluation issues is needed in the life cycles; and

- that work on careful specification of evaluation criteria in the specific life cycles should have priority from now on. This is in conformance with the focus on best practice in Year 2 of DISC.

Niels Ole Bernsen and Laila Dybkjær (Eds.).

2. Draft High-Level DISC Grid

The DISC Consortium

System/component identifier	
Description. Research prototype/product, system/component? What does it do? URL:	
System performance	
Cooperativity	How has design of co-operative system utterances been done?
Initiative	Domain communication: who has the initiative? If mixed, how is the distribution of user and system initiative? Meta-communication: can both the system and the user initiate meta-communication?
Influencing users	Walk-up-and-use system or ...? Explicit and/or implicit user instructions provided? Is there a system's introduction? Is it optional? How is it de-selected? In what ways is the introduction intended to influence users - wrt. vocabulary, grammar, utterance length, style etc.?
Real-time	Does the system respond in real time? Are there any exact measurements?
Transaction success	How defined? Measured? How? Results? Was the measurement procedure appropriate? [Note that transaction success may be translation success (in spoken translation systems)]. What percentage should be targeted for realistic applications?
General evaluation	Has any ISO standards or other well-known methods been used? Which? How? Which evaluation methods have been applied to the system/component? (Results to be described in the life cycle.)
Speech input	
Nature	Continuous? Spontaneous? Speaker-adaptive? Etc. If speaker-adaptive, how much training is required? Language(s)? Noisy environment allowed?
Device(s)	Describe the input device(s) (telephone, tape recorder, other).
Phone server	Has a phone server been deployed? Please describe it.
Acoustic models	Describe the acoustic models used by the recogniser. Which approach was used?
Search	Describe the type of search used by the recogniser.
Vocabulary	Size? What are the single entries (lemmas or inflected forms)?

	Things other than words (multi-words, formulae)?
	Special treatment of any entries?
	Specify the nature/proportion of active vocabulary at any one time.
Barge-in	Does the system listen when it speaks or is DTMF interruption possible?
Word hypotheses	Describe recogniser output.
	Recogniser score values used?
Grammar	Describe the grammar in the speech recogniser, if any?
	Any labelling? What format used?
	How to get info out of the lexicon (partial queries?)?
	Lexical classification used?
	Specify the nature/proportion of active grammar at any one time.
Prosody	Does the system process input prosody?
	What does the prosody module recognise?
	What is prosodic information being used for?
Speech output	
Device(s)	Describe output device(s).
Language(s)	Which language(s) does the system speak?
Input	What is the form of the input? Basic representation?
	Any additional notations?
Lexicon	Possible entries: What are the types of things found in the lexicon (e.g. lemmas or inflected forms)?
	Treatment of entries: What are the methods for handling different types of entries?
	Size: How big is the lexicon?
	Tagging: Are entries tagged in any way?
Sound generation technique	Characterise generation approach: canned and or concatenation (Basic units? How much data needed?), analysis-synthesis, synthesis by rule.
	Units: rule-based versus libraries?
	Combination of units?
Prosody	Is prosody included?
	What prosodic control is used (prosodic phrasing, stress, accent)?

Pronunciation description units	What are they?
	How are they combined?
Flexibility	Are different voice characters (e.g. male/female) allowed for?
	Different speaking styles?
	Different speaking rates?
	Emotion?
Miscellaneous	Disambiguation of homographs?
	Use of articulatory models?
	Time versus frequency?
User utterances	
Lexicon	Size?
	For each type of entry element, how many instances are there in the dictionary?
	Describe the lexical semantics.
	Coverage and macrostructure: What are the single word entries (e.g. lemmas or inflected forms)? Other linguistic objects (multi-word lexemes (idioms, etc.), multi-word conversational formulae, etc.)? If none how are such phenomena dealt with (e.g. in the grammar via exception lists, etc.)? Special treatment of domain-specific or otherwise "technical" items (e.g. proper names, domain-specific items, etc.)?
	Information types and microstructure: Grammatical information: For which descriptive level (or types of phenomena) is there any lexical information available (e.g. category information, syntactic valency (subcategorisation) information, semantic information (e.g. predicate-argument structures, etc.))?
	Information relevant for the interaction with other components (e.g. pointers to dialogue acts, to the domain model, etc.)?
	Representation and organisation: What format is used (e.g. feature structures, database records, data types of programming languages, other)?
	Which intermediate steps are necessary to access information from the lexicon? Is it possible to retrieve lexical information through partial queries (in which case probably a set of answers would be given)?
	Is there a system of lexical classification in use (e.g. class/instance model for each level; types; paradigms; etc.)?

	How is the organisation done?
Grammar	Describe the grammar.
	Phenomena covered: Which fragment is covered by the grammatical specification (e.g. types of questions, sentence types, word/constituent order types, types of valency constructions, etc.)?
	Is the fragment defined according to a corpus? According to a test suite?
	Specific phenomena (hard at syntactic and semantic level): Is there a treatment of co-ordination, comparatives, ellipsis, hard-to-analyse spoken language phenomena, in the syntax?
	What kind of syntactic representations are produced: are there several layers of representation (e.g. c-structures, functional structures)?
	Phenomena covered: Is the semantics fragment isomorphic wrt. the syntax fragment?
	Are elliptical constructions dealt with?
	Is there a component for anaphora resolution?
	Semantic representation: Which type of semantic representation is used (e.g. frame-like specialised structures, predicate-argument-structures, UDRS or MRS)?
	Ambiguity representation: How are ambiguities dealt with (e.g. through packed representations, through underspecification, via heuristic choice)?
	Grammar model/approach/theory: Which (type of) specification language is used (e.g. APSG, PATR-style grammar, HPSG, LFG, etc.)?
Parsing	Describe the syntactic/semantic parsing, if any. What parsing strategy is used (e.g. chart parser, island parser, top-down, bottom-up, etc.)?
	Describe the construction of semantic representation. How robust? Measured? How? Results? Was the measurement procedure appropriate? Is the approach satisfactory/sufficient (e.g. wrt. possible domain inferences)?
Style	Terse? Free? Possible utterance length?
	Is the approach satisfactory/ sufficient (e.g. wrt. the load it imposes on recogniser and grammar, or the restrictions it imposes on the users' utterances)?
	Does meta-communication require a special style? If so, is the approach satisfactory/sufficient (e.g. wrt. the load it imposes

	on the users' utterances)?
Semantics	Describe the semantics.
Discourse, context	Describe the use of discourse and context information.
System utterances	
Generation	How generated (e.g. as part of dialogue management, an independent generation module)?
Lexicon	Size? Describe the lexicon. Does it contain words which the system cannot understand?
	Coverage and macrostructure: What are the single word entries (e.g. lemmas or inflected forms)? Other linguistic objects (multi-word lexemes (idioms, etc.), multi-word conversational formulae, etc.)? If none how are such phenomena dealt with (e.g. in the grammar via exception lists, etc.)?
	Special treatment of domain-specific or otherwise "technical" items (e.g. proper names, domain-specific items, etc.)?
	Information types and microstructure: Grammatical information: For which descriptive level (or types of phenomena) is there any lexical information available (e.g. category information, syntactic valency (subcategorisation) information, semantic information (e.g. predicate-argument structures, etc.))?
	Information relevant for the interaction with other components (e.g. pointers to dialogue acts, to the domain model, etc.)?
	Representation and organisation: What format is used (e.g. feature structures, database records, data types of programming languages, other)?
	Which intermediate steps are necessary to access information from the lexicon? Is it possible to retrieve lexical information through partial queries (in which case probably a set of answers would be given)?
	Is there a system of lexical classification in use (e.g. class/instance model for each level; types; paradigms; etc.)?
	How is the organisation done?
Grammar	Describe the grammar.
	Does the system use grammatical constructs which it cannot itself understand?
	Source representations: At which level of abstraction are the input representations situated?

	Is the task of the generator morpho-syntactic/syntactic/"tactic" in nature? Or is there a set of templates for answers?
	How is the potentially large number of alternative system utterances constrained, such that exactly one kind of result sentence (or utterance) is produced (e.g. via phrasal lexicon, answer templates, stylistic heuristics, constraints on the generation grammar (exclusion of certain alternatives), use of information from the user questions to select among the system's reply alternatives, etc.)?
Semantics	Describe the semantics.
Style	Characterise the style. Is it different from what the system itself can understand?
Processing	Which type of generator is used? Does it make use of separate components, e.g. a morphology component? Is the generator robust? Incremental? What does it take as input and what does it produce as output?
Discourse, context	Describe the use of discourse and context information.
Multimodal aspects	
Device(s)	Describe the non-speech input/output device(s).
	Describe their purposes.
None	The system is unimodal (speech-only).
Non-speech input	In addition to speech, which input modalities does the system accept and which information do they convey?
Non-speech output	In addition to speech, which output modalities does the system accept and which information do they convey?
Role(s)	Describe the role of the additional modalities one by one.
Attentional state	
Focus, prior	Prior system focus: is there prior system focus (i.e. what the system has in active memory before the user's next utterance)? How is it modelled? What does it cover?
	Which messages (how precise?) can be sent to recogniser and parser?
	Does the system make use of linguistic markers, such as topic shift markers which are relatively easy to implement?
	Are statistical methods being used?
Sub-task id.	Sub-task identification: does the system do sub-task identification or topic spotting? How?
	Influence on prior focus?

	Relation to speech acts?
	Are linguistic data being used to identify sub-tasks (specific prosodic information, information about contextually determined word order, etc.)?
	Are statistical methods being used?
Expectations	Are predictions being used? How are they modelled?
	Relations to prior system focus?
Intentional structure	
Task(s)	Describe in detail the task(s) handled by the system.
	Are the tasks, as handled by the system, naturally circumscribed or are there borderline tasks or sub-tasks which might have been handled but have not been?
	Describe the task structure.
Task complexity	Is the task well-structured (such as having a prototypical order in which information is being exchanged) or ill-structured?
	How many pieces of basic information can the system ask for/provide?
	How deep is the task structure (i.e., how many levels if considered a hierarchy)?
Communication	Domain communication: is it system directed, user directed or mixed initiative?
	Free or bound order of main tasks?
	Is the communication unconstrained, as evidenced from such metrics as very long sentences, free topic order, any number of topics per sentence - or is it constrained in any way (describe)?
	Is discontinuous user input 'input gaps' handled? How?
	Describe the contextual inferences performed.
	Are indirect speech acts handled? How?
	Are incomplete user requests handled. How?
	Are incomplete answers to system questions handled? How?
	Describe type(s) of system (domain) feedback?
	System-initiated meta-communication: Does the system initiate repair (including repetition) dialogues with users? When? How?
	Does the system initiate clarification dialogues with users? When? How?

	User-initiated meta-communication: is it unconstrained or is it constrained in any way (describe)?
	Can the user initiate repair (including repetition) dialogues with the system? When? How?
	Can the user initiate clarification dialogues with the system? When? How?
	Problems: Does the system have specific problems as a result of how it communicates about the domain and about the communication? Characterise the problems.
	Does the system use other forms of communication than domain and meta-communication? Which?
	For what purpose(s) (e.g. dialogue introduction and closure)?
Interaction level	Describe the levels involved in the system initiated meta-communication, if any.
	Is graceful degradation used?
	Does the system have specific problems as a result of its level(s) of communication? What have the developers done to analyse the problems? Was their approach appropriate?
Implementation of dialogue management	How is dialogue management implemented.
	Which of the following dialogue management approach(es) some of which are partially overlapping) have been implemented in the dialogue manager?
	Dialogue Grammar: Give a description of the dialogue parsing process, accounting for at least the following points: What type of dialogue parser has been chosen (a task dependent parser, a topic-dependent parser, etc.)? Give a brief description of the parsing process, including a specification of parsing constraints (contextual constraints, etc.) How does the dialogue parser account for the correspondence between intentional and linguistic structure? Does parsing automatically give rise to a representation of the linguistic structure? How does the dialogue parser account for updates of the corresponding attentional states?
	Plan-based Approach: Please describe.
	Topical Approach: Please describe.
	Conversational Games Theory: Please describe.
	Collaborative Approach: Please describe.
	Other: Please describe.
	Which characteristic, approach-dependent options have been realised in the implementation? For instance, in case of a

	Dialogue Grammar, the grammar may be a generic one or (one reduced to) a domain-dependent subgrammar, the grammar may be presented in a graphical form or implemented in a Finite State Machine, etc.
	Does the system comprise a hybrid approach, i.e. also making use of additional statistical methods?
	Give a specification of the set of dialogue functions for which statistical methods are used.
	How is the dialogue's goal structure modelled in terms of the chosen dialogue management approach? Although task-oriented dialogues underlie a goal structure, these dialogues are not always modelled in terms of the goals to perform these tasks.
	In case of a multiple approach, are all the containing single approaches uniformly applied to both the higher and lower intentional levels in the dialogue, or are they applied to just a specific part of this structure?
	How many intentional levels are distinguished?
	Is every underlying discourse segment purpose represented separately in terms of the chosen approach?
	As for a multiple approach, how do the containing approaches differ with respect to the phenomenon of an intentional continuum?
	Does the system define a hierarchical order for the fine-graded intentional levels it distinguishes?
	If only larger intentional segments are distinguished by the system, which are these and are they hierarchically structured?
	How is the subdivision in intentional units motivated?
	What motivation is given for the assumed basic dialogue units, specific classes of which constitute the higher-order levels?
	How are basic dialogue units determined?
	What is said about the relation between dialogue phases and the classes of basic dialogue units constituting these phases?
	Which typology is defined for basic dialogue units?
	What may be parametrised, i.e. how are the intentional and linguistic structures indicated?
	Which kinds of model are there (e.g. task structure, turn-taking structure)?
	Is the control model separated from the rest of the system? Is it solely based on semantic information so that it is language

	independent?
	Are dialogue patterns used?
Linguistic structure	
Speech (dialogue) acts	Does the system identify speech (or dialogue) acts in the users' input? How?
	How many speech acts does the system recognise?
	What is the information being used for?
	Can indirect speech acts be handled?
	Is there any difference between the system's use of speech acts and its ability to do topic spotting (sub-task identification)?
Discourse particles	Does the system identify discourse particles in the users' input? How?
	How many discourse particles does the system recognise?
	What is the information being used for?
Co-reference	Does the system do co-reference resolution? How?
	Does it cover all co-references which can be expected in the input?
	How does the system behave with respect to different linguistic realisations of the same referential object (pronouns, anaphoric descriptions, etc.)?
	How does co-reference relate to the dialogue's attentional state and intentional and linguistic structure?
Ellipses	Does the system do any particular processing of ellipses (instead of just working with the uttered words)? How?
Segmentation	Does the system do any user turn segmentation? How? What is the information being used for?
Interaction history	
Linguistic	Does the system maintain a record of the surface language of the users' utterances?
	How represented?
	What is the record being used for?
	Does the system represent the segmentation structure of the preceding dialogue, and what is the relation with the incremental representation of the corresponding goal structure?
	Does the system keep a history of co-referential expressions?
Topic	Does the system maintain a record of the order in which topics have been addressed through the interaction?

	How represented?
	What is the record being used for?
Task	Does the system maintain a record of the task-relevant information which has been exchanged?
	How represented?
	What is the record being used for?
Performance	Does the system maintain a record of the user's performance during interaction?
	How represented?
	What is the record being used for?
	Can the system adjust its communication according to the performance model during interaction with the user?
Domain model	
Data	Describe the domain data the system uses during interaction.
	Is the data artificially limited or fully realistic?
	How are they represented?
Rules	Describe the rules operating on the domain data, such as completions, temporal inferences or constraints.
	What are they used for?
	Are the rules placed in a separate system module?
	If not, which module(s) are they part of and why?
User model	
Goals	Describe, from the system's point of view, the user's goal during interaction.
Beliefs	Describe what the system does to handle the user's beliefs during interaction, e.g. does it store information during the interaction which it can draw upon later in the same interaction, such as e.g. a customer number which is given by the user while starting a reservation dialogue and which is needed again if the user wants to make a second reservation later during the same call? In this case the system may just check with the user that it already has the right customer number instead of asking for it again.
Preferences	Describe what the system does to handle the user's preferences during interaction.
	Does the system have access to knowledge about user preferences from earlier dialogues or does it only have the information which it can obtain during the ongoing interaction? (Knowledge from previous dialogues could e.g.

	be a user preference for a certain flight departure. If no such knowledge is available the system will have to ask for user preferences during the dialogue such as whether the user has a preference for discount fares and their associated departure times.)
User group	Does the system assume any distinctions among user groups, such as between domain novices and experts, novices and experts in using the system, other?
	Any modelling of social settings?
Cognition	Has anything been done to take into account the specific cognitive characteristics of users, such as task load, limited memory, natural "response packages" or limited attention span?
	Are such characteristics not being considered relevant to the interaction?
	If not, is this justified, or is it possible to characterise specific problems the system has because, e.g., cognitive load issues were not considered?
Component architecture and function	
Generic architecture	Does the component have a domain-independent, generic architecture?
	Indicate the level of genericity.
	Specify the input and output characteristics and representations of the component.
	Describe the interaction with and interfaces to other system components.
Sub-components	Describe the number of sub-components and their functionality.
	Describe their representations.
Flow	Describe the process flow among the sub-components. If possible provide a diagram.
	Which information is exchanged between the individual sub-components?
	Describe the data flow from input to output for the component, preferably illustrated by an example.
Function	How does the component contribute to the overall system requirements? What is the function of the component (e.g. a dialogue component may have a mediating function, as is usually the case with translation systems in which the system does not really act as a dialogue participant)?

System architecture	
Platform	Describe the hardware and operating system platform. Is the platform adequate according to today's standards?
Tools and methods	Describe the tools and methods used.
Generic	Describe the generic software architecture (e.g. multi-agent, object-oriented). Is the generic software architecture adequate according to today's standards?
No. components	Describe the number of system components (cf. the architecture diagram).
Flow	Describe the process flow among the system components (cf. the architecture diagram). What are the components of the system? Which information is exchanged between the individual components? Describe the data flow from input to output for the system, preferably illustrated by an example.
Processing times	Describe the average percentages of processing time spent on the different processing tasks the system has.
System integration	
System resource utilisation	Describe CPU usage. Describe memory usage. Is it possible to configure any of the modules as standalone modules? Client/server?
Shared information resources	Which modules share which information resources?
Interactions	Which system modules interact? How?
Data passing	For each system module describe what input it receives, from where it is received and in what form it is received. For each system module describe what output it passes on, to which module it is passed on and in what form.

System/component architecture

If possible, provide a diagram outlining the architecture.

System dialogue(s)

For each aspect analysis, DISC should provide one or more complete and annotated/commented example dialogues with the system being analysed. This will contribute to providing a concrete "feel" for the system in question. The example dialogues should be selected by the analysers, not by the developers.

System/component screen shot(s)

If relevant, provide one or more screen shots.

3. Draft Specific DISC Grids

The DISC Consortium

3.1 Speech recognition grid

Telephone interface and switching components (as appropriate)

Signal processing modules

Base phone sets

Lexica (dictionaries)

- Coverage and macrostructure

    What are the single entries (lemmas or inflected forms)?

    Things other than words (multi-words, formulae)?

    Size: How many?

    Phenomenological: Special treatment of any entries?

- Information types and microstructure

    Grammatical information: Any labelling?

    Relations to other components: Any connections noted?

- Representation and organisation:

    Formalism and format: What format used?

    Access: How to get info out of the lexicon (partial queries?)?

    Organisation: Lexical classification used?

Language models

Acoustic models

Search organisation and control

Relationships to other components for speech recognition

3.2 Speech generation grid

Input to component: What is the form of input, basic representation, any additional notations, for example prosodic markers?

Output from component: Speech in the telephone network, speech only or in combination with for example graphics?

Lexicon:

- Possible entries: what are the types of things found in the lexicon? For example morphemes, words, words with transcriptions, symbols.

- Treatment of entries: what are the methods for handling different types of entries?

- The size: how big is the lexicon?

- Are the entries tagged in any way, for example part-of-speech?

Resources, synthesis method rule-based vs. libraries? Combination of units?

Resources, synthesis parts: What are the units for the synthesis, e.g. phonemes morphemes words?

Rule-based: If so, what is controlled by rules. Special treatment of or special rules for any entries?

Pronunciation description units: what are they, how are they combined?

Disambiguation of homographs

Prosodic control

Resources, different voices

Multilingual, which languages?

Flexibility: Different speaking styles, speaking rates, emotion?

Evaluation: Any performed test, which measures have been used, results, is the method appropriate, satisfactory, sufficient?

Sound generation technique: Canned and/or concatenation (basic units, how much data needed?) or formant synthesis, analysis-synthesis, synthesis by rule, time versus frequency domain?

Use of articulatory models

Utterance generation

Interaction with other resources

Relationships to other components for speech generation

3.3 Natural language understanding and generation grid

The following is an incomplete draft list of high-level questions to be used to analyse the natural language understanding and generation components of SLDSs and the resources used by these components. The list still contains very high-level questions, which need to be further broken down into a real grid format (partly allowing for "yes/no"-features on) the different possible options).

1. Grid questions

1.1. Overview

The following structure is proposed for the organisation of the grid questions for NL resources; it follows the model given in figure 2.16 of the book by [Bernsen et al. to appear].
• User utterances
  - Lexicon;
  - Grammar;
  - Semantics;
  - Discourse, Context.
• System utterances)
  - Lexicon;
  - Grammar;
  - Semantics;
  - Discourse, Context.

This structure can accommodate both, systems where analysis and generation use the same resources, and systems where there are different resources (in this case, there is a need to have subparts on analysis lexicons vs. generation lexicons, etc.).

1.2 Resources

1. Lexicon (both analysis and generation, possibly to be refined for each sub-area):
(a) Coverage and macrostructure
•   Formal entry definition: for single word entries, which types of word forms can be entries? (e.g. any inflected form; only lemmas?[1])
•   Contents-wise entry definition: besides single word entries, are there other linguistic objects in the lexicon (e.g. multiword lexemes (idioms, etc.), multiword conversational formulas, etc.)? If

---

[1]   If a lemma lexicon is used, there must be some kind of mophological component, used along with the lexicon.

no, how are such phenomena dealt with (e.g. in the grammar, via exception lists, etc.)?

- Size: For each type of entry element (e.g. single words vs. multi-words), how many instances are there in the dictionary?

- Phenomenological level: is there a specific treatment of domain-specific or otherwise "technical" items (e.g. proper names, domain-specific items, etc.)?

(b) Information types and microstructure:

- Grammar-related information: For which descriptive level (or: types of phenomena) is there any lexical information available (e.g. category information, syntactic valency (subcategorization) information, semantic information (e.g. predicate-argument structures, etc.))?

- Information relevant for the interaction with other components (e.g. pointers to dialogue acts, to the domain model, etc.)

(c) Representation and organisation:

- Representation formalism and format of entries: which format is used (e.g. feature structures, database records, data types of a programming language, other...)?

- Representation and access[2]: which intermediate steps are necessary to access information from the lexicon? Is it possible to retrieve lexical information through partial queries (in which case probably a set of answers would b e given)?

- Lexical organisation: Is there a system of lexical classification[3] in use (e.g. class/instance model for each level; types; paradigms; etc.)? How is the organisation done?

2. Grammars

(a) Analysis - syntactic level

- Phenomena covered: which fragment is covered by the grammatical specification (e.g. types of questions, sentence types, word/constituent order types, types of valency constructions, etc.)? Is the fragment defined according to a corpus? According to a test suite[4]?

- Specific phenomena (hard at syntactic and semantic level): Is there a treatment of co-ordination, comparatives, ellipsis, hard to analyse spoken language phenomena, in the syntax?

- What kind of syntactic representations are produced: are there several layers of representations (e.g. c-structures, functional structures, etc.?)

(b) Analysis - semantic and pragmatic level

- Phenomena covered: is the semantics fragment isomorphic wrt. the syntax fragment? Are elliptical constructions dealt with? Is there a component for anaphora resolution?

- Semantic representation: which type of semantic representation is used (e.g. frame-like

---

[2] This question could also be moved to the "interactions"-block.

[3] This question must be seen in connection with the problem of lexical updates: if there is a classification of lexical phenomena according to common linguistic properties, this may help to find prominent examples from where to copy descriptions in case of updates. In addition, consistent class-wise updates are eased.

[4] A related life-cycle question has to do with tools and methods used for evaluating the grammatical (and lexical) coverage of an SLDS.

specialised structures, predicate-argument-structures, UDRS or MRS etc.)?

- Ambiguity representation: how are ambiguities dealt with (e.g. through) packed representations, through underspecification, via heuristic choice, etc.)?

(c) Generation - Syntax

- Representations used to generate from: at which level of abstraction are the input representations situated? Is the task of the generator morphosyntactic/syntactic/"tactic" in nature? Or is there a set of "templates" for answers?

- Choice: How is the potentially large number of alternative system utterances constrained, such that exactly one kind of result sentence (or utterance) is produced (e.g. via phrasal lexicon, answer templates, stylistic heuristics, constraints on the generation grammar (exclusion of certain alternatives), use of information from the user questions to select among the system's reply alternatives, etc.)?

1.3 Processing

1. Understanding

(a) Grammar model/approach/theory: which (type of) specification language is used (e.g. Augmented PSG, PA TR-style grammar, HPSG, LFG, etc.)?

(b) Parsing strategy: which kind of parsing strategy is used (e.g. chart parser, island parser, top-down, bottom-up, etc.)?

2. Generation

(a) Which type of generator is used? Does it make use of separate components, e.g. a morphology component?

(b) Properties of the generator: is the generator robust/incremental?

2. Relationships, interactions

Some of the questions in this section may later be moved to the chapter of the DISC WP-1-report which deals with System Architecture. The questions below fall into two parts: (1) those concerning potential relationships between the understanding and the generation side of the NL component, and (2) those concerning relationships with other components of the overall SLDS.

2.1 Sharing of resources within the NL component: Understanding vs. Generation

1. Analysis vs. Generation within the NL component

(a) Use of resources: do analysis and generation share lexical or grammatical resources? Which ones? What is the relationship between the analysable fragment and the generatable fragment (these may definitely be asymmetric, as more utterances may have to be analysable than need

to be generated)?

(b) Use of intermediate representations: do analysis and generation use the same type of representations? Can lexical and grammatical choice in generation use representations of analysed user input?

2.2 Interaction of NL resources with resources from other components

(a) Resources: Is there any interaction between any of the following; if yes, how is it implemented?

- SR lexicon <-> NL analysis lexicon;

- NL analysis lexicon <-> domain model;

- NL analysis lexicon <-> keywords for dialogue act identification;

- Domain and dialogue act description <-> generation lexicon.

(b) Representations: In which way, through which types of mappings, interfaces, etc. is it possible, for a given component, to use the output representations of its "predecessor" as an input? E.g.:

- Use of prosody to determine correct grammatical analysis;

- Recognizer output <-> grammatical analysis;

- Syntactic and semantic analysis results <-> dialogue management;

- Semantic analysis results <-> interaction history; <-> settings for generation of replies, at dialogue plan level.

3.4 Dialogue management grid

1. Dialogue Component Tasks

- What are the main tasks of the Dialogue Component?

  Example

  As for the Verbmobil translation system, its main tasks are 1) constraining the search space of other system components[5] 2) providing contextual information to other system components in order to realize adequate translations[6] 3) following the discourse in case that no translation is asked for (shallow processing) 4) dialogue control for clarification meta-dialogues.

- Does the dialogue component only have a mediating function, as is usually the case with translation systems in which the system does not really act as a dialogue participant?

  Example

  In contrast to, e.g., the Sundial system, the Verbmobil translation system only fulfils a mediating function in human-human spoken dialogues, except in cases of induced clarification dialogues in which the system functions as a real dialogue participant, implying the fulfilment of a dialogue control function.

- To which system requirements do the tasks of the dialogue component contribute (robustness, efficiency, etc.), and what does this contribution consist of?

  Example

  As for the Verbmobil system, the task of constraining the search space of other system components contributes to the overall system requirement of efficient (real-time) dialogue processing, whereas the task of following the discourse in case that no translation is asked for contributes to the requirement of providing translations on demand.

2. Dialogue Component Architecture

    a   *Architecture specification*

- What are the submodules of the dialogue management component of the system, what are their functions, and how do they interact (see also below)?

  Provide, if possible, a figure of the architecture of the dialogue management

---

[5] For instance, contextual information in the form of top-down dialogue act predictions are used to constrain the search space of the Speech Recognition component by constraining the set of words which are likely to occur in the next utterance. The same type of predictions are also used to constrain the search space of both the Syntactic Analysis component (narrowing down the set of applicable grammar rules to a specific subgrammar) and the Keyword Spotting component (delimit its search space to the most probable keywords).

[6] In this respect contextual information is used for, e.g., anaphora resolution during semantic evaluation, for the disambiguation of translation equivalents during transfer and for controlling lexical variation in the generation of target language expressions.

component, thereby indicating the interactions between the submodules.

b.   *Generic architecture*

- Does the Dialogue Management component of the system have a generic architecture? Give an answer to this question insofar it is relevant to the grid analysis, thereby also indicating the level of genericity (language-independent architecture, domain-independent architecture, etc.).

  Example

  The dialogue management component of the Daimler-Benz dialogue system (e.g. Heisterkamp et al. 1992) is based on the Sundial system (e.g. Bilange 1992) which has a generic, both language- and domain-independent architecture. Dialogue is modelled dynamically in terms of belief and intention states of the system rather than in terms of the notions of speaker intentions and (domain-dependent) speech (dialogue) acts.

c.   *Interaction with other system components*

- Specify the input/output characteristics of the dialogue component.[7]

3.1   The Implemented Approach

a.   *Kind of descriptive approach*

Which of the following dialogue management approach(es) (some of which are partially overlapping) have been implemented in the dialogue system, this according to the descriptions given below?

Note that although the approaches differ wrt. to how they primarily account for both dialogue structure and the way the development of this structure is controlled, they may make use of techniques provided by other approaches, such as dialogue grammar and plan-based techniques.

- *Dialogue Grammar*

  The dialogue grammar approach (among others, Polanyi & Scha 1984) accounts for dialogue structure (dialogue management) in terms of a dialogue grammar. Similar to sentence grammar rules defining the set of well-formed sentences, dialogue grammar rules give rise to sequential and hierarchical constraints on acceptable dialogues. In the dialogue systems, the dialogue grammar rules may function as a *bottom-up parser* and/or a *top-down prediction mechanism,* the latter being responsible for predicting possible next sequences on the basis of an input sequence.[8] It is not an exception that the nodes in the syntactic tree are enriched with semantic features. An example of a spoken language dialogue

---

[7]   Although the interaction between system components is relevant for the grid description of these components, it probably can be better accounted for in the framework of a description of system integration.

[8]   The Verbmobil system which follows a multiple approach in dialogue management uses a (finite state machine variant of a) dialogue grammar just for controlling the dialogue flow, namely the identification of inconsistencies of an incoming dialogue act with what is predicted on the basis of the given input. Dialogue parsing, on the other hand, is realised by a plan-based parser.

system which accounts for dialogue structure in terms of a dialogue grammar is the Waxholm system (see also below).

- *Plan-based Approach*

  The plan-based approach (e.g. Allen & Perrault 1981, Grosz & Sidner 1986) accounts for dialogue structure primarily in terms of the underlying dialogue segment purposes, which are modelled as *speakers' intentions* that have to be recognised in the dialogue understanding process. Dialogue segments consist of (classified sets of) dialogue acts which are planned by the speaker to achieve his/her goals. Many spoken language dialogue systems account for dialogue structure in terms of the plan-based approach, including the Verbmobil translation system.

- *Conversational Games Theory*

  Conversational games theory (e.g. Kowtko, Isard & Doherty 1991), though not reducible to the other approaches, often makes use of dialogue grammar and/or plan-based techniques to account for dialogue structure in terms of so-called *conversational games*. The hierarchical structure of the dialogue reflects a recursive application of conversational games, each of which consists of a set of general and game-specific *moves*. An example of a dialogue system using this approach is BT's MailSec (Williams 1996), a spoken language dialogue system giving oral access to one's e-mail by telephone.

- *Topical Approach*

  While a plan-based approach accounts for dialogue structure in terms of the underlying speakers' intentions, a topical approach (e.g. Grimes 1975, Groenendijk & Stokhof 1993, Johnson-Laird 1983, Joshi & Weinstein 1981, Van Kuppevelt 1991) explicates dialogue structure in terms of the assumed sequential or hierarchical structure of topics defined for its segments. In the Waxholm system, e.g., higher-order dialogue structure is modelled in terms of a topical approach.

- *Collaborative Approach*

  In contrast to plan-based approaches, collaborative approaches (e.g. Allen 1991, Grosz & Kraus 1993, Grosz & Sidner 1990, Sidner 1992) provide an account of dialogue structure in terms of *joint intentions* rather than speakers' intentions. Dialogue is modelled as a joint activity committing the agents to understand one another, thereby motivating the occurrence of confirmations, clarifications, etc. Collaborative approaches use techniques from the plan-based and dialogue grammar approaches. Examples in our exemplar list which imply a collaborative approach are the Danish Dialogue System for flight ticket reservation (e.g. Bernsen, Dybkjaer & Dybkjaer 1995) and the Daimler-Benz dialogue system (e.g. Heisterkamp, McGlashan & Youd 1992).

- *Other approaches*

b. *Use of single versus multiple approach*

- Does the system comprise a multiple approach?

- If the system uses a multiple approach, what are the reasons for assuming this multiple approach, what are the functions of the single approaches and what is their prominence (equally prominent, one subservient to the other, etc.)?

Examples of a multiple approach

Verbmobil makes use of both of a plan-based approach and a Finite State Machine (see point *b.* below for a description of these approaches). The Waxholm system, which makes use of a multiple approach too, uses both a topical approach and a discourse grammar. However, the two systems differ as to whether the same approach is used for all structural levels or not (see below, sect. 3.2 A, point c (i)).

c. *Implementation characteristics*

- Which characteristic approach-dependent options have been realised in the implementation?

Example of implementational aspects

In case of a dialogue grammar, we may deal with e.g. the following implementational aspects. Genericity: the dialogue grammar may be a *generic grammar* or (one reduced to) a *domain-dependent subgrammar*. Presentation: the grammar may be presented e.g. in a *graphical form*. Implementation: the grammar may be implemented e.g. in a *Finite State Machine*. Etc.

d. *Use of hybrid identification and prediction methods*

- Does the system comprise a hybrid approach, also making use of additional statistical methods?

- Give a specification of the set of dialogue functions for which statistical methods are used (see also below), making explicit whether the method applies to all structural levels in the dialogue or just to a specific kind?

Examples

Besides a Dialogue Planner and a Finite State Machine, the statistical submodule of the Verbmobil dialogue component makes use of additional statistical methods to predict the next low-level, basic dialogue unit which is a dialogue act. In Waxholm, on the other hand, statistical knowledge is used for predicting dialogue units which are considered to be of the highest structural level, called topic units.

3.2   Dialogue Structure

A.   Intentional Structure

Given the task-oriented character of the dialogues under consideration, we take the intentional structure of a dialogue as the (mostly hierarchical) structure of dialogue segment purposes to perform a (sub)task.

a. *Task characteristics*

(i) *Task specification*

Which specific tasks are performed by the system (making a reservation, showing a time table, fixing a meeting date, etc.)?

(ii) *Task structure*

Are all tasks well-structured, i.e. do all tasks have a stereotypical structure explicating both the amount and order of information updates needed to complete the task (see, in particular, Bernsen et al. 1997 on this point)?

b. *Communication types*

(i) *Domain communication*

Specify the relevant characteristics of the task-oriented domain communication, in particular whether and/or how the system accounts for the following phenomena (some of which in fact may give rise to meta-communication):

- System directed, user directed or mixed initiative communication
- Free or bound order of main tasks
- Discontinuous user input ('input gaps')[9]
- Contextual inferences of different types[10]
- Indirect speech acts[11]
- Incomplete, underspecified user requests
- Incomplete answers to system questions
- Other

(ii) *Meta-communication*

Does the system allow for both system- and user-initiated meta-communication, and in what way?

How does the system behave with respect to different types of meta-communication, such as:

- Repairs (repair types, repair success, etc.)
- Clarifications (location-restricted/-unrestricted, system-/user-initiated, etc.)
- Repetitions
- Reasons ('nucleus reasons' vs. 'satellite reasons')[12]

---

[9] Discontinuous user input involving large gaps between given and new system input may be very problematic in case that system performance is dependent on predictions derived on the basis of just recent system input.

[10] Different types of contextual inferences may require a variety of inference rules, e.g. those used in Verbmobil to account for the fact that a counterproposal (for a date, etc.) implies the rejection of a previous proposal, that a new proposal (for a date) implies the acceptance of a (not incompatible, less specific) previous proposal and that a change of dialogue phase implies the acceptance of a previous proposal (for a date).

[11] A problem for most spoken language dialogue systems is the determination of indirect speech (dialogue) acts. Because indirect speech acts cannot be detected on the basis of surface cues, keyword spotting is useless in these cases.

- Deliberations (deliberation types)
- Feedback (feedback Realization types)
- Other

(iii) *Other communication types*

- Dialogue introduction and dialogue closure phases (domain-independent communication)
- Side structure information induced by digressions (related to, but out of the domain communication)

c.   *Modelling of intentional structure*

How is the dialogue's intentional structure modelled in terms of the chosen dialogue management approach?

Although task-oriented dialogues have an underlying goal structure, these dialogues are not always modelled in terms of the goals to perform these tasks (see above).

(i)   *Uniform versus partial application*

In case of a multiple approach, are all the single approaches uniformly applied to both the higher and lower intentional levels in the dialogue, or are they applied to just a specific part of this structure?

Examples

In Verbmobil both the plan-based approach and the Finite State Machine capture all intentional levels. The same holds for the Daimler-Benz spoken language dialogue system which modelled intentional structure in terms of common goals to perform a (sub)task.

The Waxholm system, on the other hand, does not apply its two approaches to all levels: while a topical approach is applied to higher intentional structure (using statistical information, the higher level task communication is parsed into a sequence of topic units), topic-internal structure is accounted for in terms of a dialogue grammar which is presented into a graphical form.

(ii) *Intentional levels*

How many intentional levels are distinguished?

a.   *Intentional continuum*

Is every underlying discourse segment purpose represented separately in terms of the chosen approach?

- As for a multiple approach, how do the containing approaches differ with respect to the phenomenon of an intentional continuum?

- Does the system define a hierarchical order for the fine-graded intentional levels it distinguishes?

---

[12] As opposed to satellite reasons, reasons occurring in a nucleus dialogue segment must be considered to belong to domain communication because of the property of providing an intentional value.

Note that the phenomenon of an intentional continuum characteristic for human-human dialogues is usually not fully accounted for in dialogue systems.

b.   *Representation of larger intentional units*

If only larger intentional segments are distinguished by the system, which are these and are they hierarchically structured?

Example

The Verbmobil system distinguishes the following hierarchy of intentional units, assigning a corresponding structure to the dialogue act sequences constituting a dialogue:

- highest level of *individual dialogues*

- lower level of *dialogue phases*

  (the main phases are: Introductory Phase, Negotiation Phase and Closing Phase)

- level of *dialogue turns*

- lowest level of *dialogue acts*

Relevant questions wrt. the assumption of larger intentional units are the following:

- How is the subdivision in intentional units motivated?

- What motivation is given for the assumed basic dialogue units, specific classes of which constitute the higher-order levels?

- How are basic dialogue units determined?

- What is said about the relation between dialogue phases and the classes of basic dialogue units constituting these phases?

- Which typology is defined for basic dialogue units? Indicate whether the set of basic dialogue units is an open set, whether the proposed typology contains domain-dependent basic units, etc.

d.   *Dialogue parsing*

Give a description of the dialogue parsing process, accounting for at least the following points:

- What type of dialogue parser has been chosen (a task dependent parser, a topic-dependent parser, etc.)?

- What kind of parsing strategy has been followed (top-down or bottom-up strategy, complete or partial parsing strategies, etc.)?

- Give a brief description of the parsing process, including a specification of parsing constraints (syntactic constraints, semantic constraints, prosodic constraints, contextual constraints, etc.).

- How does the dialogue parser account for the isomorphism between intentional and linguistic structure? Does parsing automatically give rise to a representation

of the linguistic structure?

- How does the dialogue parser account for updates of the corresponding attentional states, either directly or indirectly?

B.   Attentional state

The term attentional state is used to refer to the *current dialogue information state* and comprises the current (domain- or meta-communication) subtasks which are not yet closed off, i.e. both the most recent (sub)task currently dealt with and the set of superimposing (sub)tasks to which this subtask contributes. The attentional state is dynamic, representing the (remaining) set of objects, properties or relations in *focus of attention* at each point in the development of the discourse. If following Sidner (1979) and Grosz & Sidner (1986), a distinction can be made between *local focus* ('focus prior') and *global focus*, referring to the distinction between what is brought into focus of attention by the current subtask addressed by, e.g., the latest question and that which, at the same time, is still in focus of attention as the result of previous superimposing subtasks.

a.   *Modelling of attentional state*

(i)   *Local focus*

- How does the system model local focus?

- In what specific way does the modelling of local focus give rise to expectations for the next user input?

- How do focus values selected by answers to local questions give rise to updates on the global level (see also above)?

- How is local focus determined in case of other speech acts than questions, and how is it determined in case of extended answers to questions?

- Of what types of linguistic data is made use to identify local focus (specific prosodic information, information about contextually determined word order, etc.)?

Example of local focus

In the Waxholm system, e.g., local focus is that which is in focus of attention at the current stage arrived at in the topic-internal graph structure. Although this is not made explicit by the developers, local focus is the contextually provided set of alternative realisations of the lower-order task associated with the current point in the graph structure. From this set of options one is chosen as the result of which a further stage in the graph is reached. Every further stage in the graph structure can be seen as bringing us closer to the realization of the higher-order task associated with the current topic.

(ii)   *Global focus*

- Does the system model global focus? If so, how does it model global focus in relation to local focus?

- If the system makes use of expectations induced by global focus, what is

their nature and impact?

- Does the system make use of linguistic markers of global focus phenomena, e.g. topic shift markers which are relatively easy to implement?

Example of global focus

In the Waxholm system global focus is determined by the current topic (e.g. a time-table topic, a show-map topic, an out-of-the-domain topic), comprising the contextually provided set of alternative realisations (e.g. the set of maps in case of a show-map topic) of the higher-order task associated with this topic.

b. *Additional expectations*

- What other (statistical or other) methods are implemented to compute, in particular, the immediate local expectations which arise from a current system/user question?

- What is the function of a multiple approach in this respect?

c. *Co-operativity*[13]

Task realization in dialogue is considered to be part of its dynamic, attentional structure. How does the system account for a co-operative performance of these tasks, e.g. with respect to the following phenomena?

- Underinformativeness (e.g. the system's exhaustiveness strategies)

- Overinformativeness (e.g. the system's refinement strategies in case that the requested task is too broad)

- Truthfulness of provided information (the system's information checking mechanisms)

- Brief and orderly presentation of requested information (the system's presentation mechanisms)

If possible, indicate whether and how co-operative action initiation as part of these strategies is goal determined.

C. Linguistic Structure

The linguistic structure of a dialogue constitutes its segmentation structure. The dialogue's linguistic structure is in correspondence with both the intentional structure of the dialogue and the structure of attentional states.

a *Dialogue segments*

- How many segmentation levels are distinguished?

Because of the correspondence between intentional structure and segmentation

---

[13] Although co-operative dialogue performance can be handled within *human factors*, it is inherently related to the dynamics of dialogue expressed by its structure of attentional states.

structure, dialogue segmentation within the system crucially depends on how many intentional levels are distinguished (see above).

b. *Speech act types*

- Which speech act types (e.g. questions, assertions, commands, etc.) can be handled by the system?[14]

- How are the different types identified?

- How successful is the system with respect to the interpretation of indirect speech acts (see also above), and how does this depend on the dialogue management approach chosen?

c. *(Co-)reference*

- How does the system behave with respect to different linguistic realisations of the same referential object (pronouns, anaphoric descriptions, etc.)?

- Does the system keep a history of co-referential expressions?

- How does co-reference relate to the dialogue's intentional and attentional structure?[15]

3.3 Dialogue Context

A. Dialogue History

a. *Representation of Intentional Structure*

- Does the system incrementally construct a representation of the intentional structure as the result of the preceding context?

- What is the format of representation of the intentional structure (tree representation, matrix representation, etc.)?

- Which submodule of the dialogue component is responsible for this representation (e.g. the dialogue planner) and where is the representation stored (e.g. in the dialogue memory)?

- Which other system component does the stored intentional information contribute to and how?

b. *Representation of Attentional States*

- Does the system provide an incremental representation structure of the set of attentional states constituting the dynamic development of the dialogue, thereby providing a representation of the set of open alternatives ('focus set') at any dialogue information state?

---

[14] What is meant here are not dialogue acts which, as illustrated above, are very often also defined in terms of propositional content (see above).

[15] See, in particular, Grosz & Sidner (1986) on this phenomenon.

- What is the format of representation (e.g. an updated SIL object or a list of such objects in Sundial and the Daimler-Benz dialogue system)?

- How does the incremental representation structure of attentional states account for the set of *referentially accessible objects* at any dialogue state?

- Which other system component does the representation of attentional structure contribute to and how?

c. *Representation of Linguistic Structure*

- Does the system represent the segmentation structure of the preceding dialogue, and what is the relation with the incremental representation of the corresponding intentional structure?

- Does the system also represent referential structure (see above)?

- Which other system component does the represented linguistic structure contribute to and how?

B. User Model

a. *Modelling of performance-relevant aspects of the user*

- To what extent does the system model performance-relevant aspects of the user, such as user beliefs (his own beliefs as well as beliefs about the system), user desires (preferences), user expertise, etc.? What is precisely their function within the system?

- Does the system also allow for inferencing over initial beliefs?

- What is the status of, in particular, the user beliefs, desires and intentions? And, what is most important, how do these interact in the process of dialogue control?

b. *Contribution to co-operativity*

- What contribution is provided by the user model with respect to co-operative system behaviour?

- Give an indication of the remaining, most relevant problems in this respect.

C. Domain Model

a. *Data*

- Give a specification of the data types determined by the task(s).

- Provide the global characteristics of objects, properties and relations relevant to these tasks.

b. *Rules*

- Give a global characterisation of world knowledge rules used by the system, as well as the characteristic properties of their application.

D.   Dialogue Settings[16]

    a.   *Physical setting*

- What are the characteristic physical circumstances in which the dialogues take place (face-to-face situation, communication through electronic devices, etc.), and how is their influence on co-operative communication modelled in the system?

    b.   *Social setting*

- In what type of communicative situation do the dialogues take place (expert-novice, employer-employee, etc.), what roles do user and system have in these situations, and how is social knowledge of this type modelled in the system?

---

[16] Although dialogue settings are directly relevant to dialogue management, in fact they are part of the human factor description.

3.5 Human factors grid

Input from user (system understanding)

Speech style

- Does the system understand connected words such as connected digits?
- Does the system restrict input to isolated words ?
- Does the system allow continuously spoken utterances?
- Is grunt detect / silence detect available?
- Is there a limit on the length of an utterance?
- Is it possible to adjust max. utterance length?
- Is prosodic information interpreted?
- Are anaphora understood by the system?
- Are ellipses understood by the system?
- Are confidence levels supported
- Are command words provided for dialogue navigation?

Natural language processing

- Is syntactic parsing carried out?
- Is semantic parsing carried out?
- Is the dialogue specified *a priori*?
- Is the dialogue dependent on the user providing information, e.g. a destination as well as a starting point?
- Are discourse segments/topics used?

Multimedia input/output

- Are the following channels of user input supported:
  Speech?
  Gesture?
  DTMF?
- Does the system support text and / or graphic input (e.g. via the WWW)?
- Is a mix of the above possible?
- Is the interaction telephone-based?

Output to user

Channel

- Are non-verbal sounds (e.g. music, background noise, auditory icons) used by the system?
- Does the system make use of engineered sounds (e.g. earcons)?
- Is text and graphical output possible?
- Does the system support a mix of outputs (e.g. speech and text concurrently)?
- Is synthesised speech used?
- Was any evaluation of speech output quality carried out?

Signalling

- Can the system be interrupted?
- Does the system make use of a beep to prompt users?
- Does the system use an explicit prompt structure (i.e. menu structure)?
- Is a command structure supported (e.g. flat hierarchy)?
- Is there use of programmable intonation patterns in system output?
- Is a curt/denatured style used?
- Does the system use a transactional mode?
- Or an interaction mode?
- Is a form filling/frame-based dialogue structure used?

Help

- Is a help facility available?
- What kind of help facilities are used?
    Single shot?
    Incremental?
    Context Dependent?
- Is it possible to change the input modality?
- Are prompts worded such that it is possible to interrupt them before everything is said (i.e. semantically valid cut-off points)?
- Is an human operator backup provided?
- Is additional system information required, end-user guide?
- Are escape routes out of the dialogue available?
- Are explicit confirmations used?

User model

- Are user preferences reflected in the dialogue?

- Is a interaction history maintained?
- Is a change of dialogue history possible?

Dialogue style

- Is the dialogue driven by the system?
- By the user?
- Mix

Novice and expert

- Does the dialogue explicitly cater for novices and experts?

User description

Group

- For which group(s) does the system provide an interface:
  Agents?
  Caller/User?
  System Administrators?

Experience

- Is the system designed for naive / novice users?
- Is the system designed for expert users?

Frequency

- What frequencies of use are mainly supported by the system:
  Often?
  Occasional?
- Does this apply to the whole application?
- Only to certain Functions?

Domain description

- Was this studied before system design?
- Are the tasks transactional?
- Is information provided from a database?

System knowledge

- Is knowledge of SLDSs required?
- Is knowledge of previous service (e.g. agent based system) required?
- Do first time callers receive guidance from the system?

Domain knowledge

- Is knowledge of application domain required?

Use location

- Where is the system to be called from?/used
    - Home
    - Work
    - Mobile, e.g. car, train
- Are the users subscribers?
- Are the users paying premium rates for the service?

Language/culture

- Does the system cater for different languages,?
- Are different dialects supported?
- Is the dialogue dependent on the user's gender?
- Were cultural dependencies identified?

3.6 Systems integration grid

Functional components

- Does the system use a syntactic parser?
- Does the system use a semantic parser?
- Does the system use a separate Dialogue Manager?
- Does the system use a Text-to-Speech system?
- Does the system use pre-recorded full prompts?
- Does the system use pre-recorded concatenated prompts?

Application resources

- What format are the acoustic models (e.g. binary)?
- Are speaker-dependent acoustic models used?
- What grammar formalism is used?
- Are different grammars used?
- Is an application independent lexicon used?
- Are application specific databases used (e.g. user models, dialogue history)? How many?
- Is a standard interface used with the application-specific database(s)?

Logical integration

- Does the system use a public speech application API?
- Is information communicated in two directions?
- Is information communicated directly between functional modules?
- Is access to information restricted to particular modules?
- Can new functional modules be added?
- Is the form and content of the information communicated application-dependent?
- Can the communication module be used for other applications without changes?
- Can non-speech input and/or output modules be integrated?
- Can system internal data be accessed from the outside?
- Is system backward compatible?

Architectural integration

- Is a standard telephony protocol used?
- Is external LAN/WAN access provided?
- Is external database access provided?
- Are other peripheral devices supported?
- Is support provided for Intelligent Networks?
- Can control of the dialogue be surrendered to a Intelligent Network (or other)?
- Are application resources accessible from outside the SLDS (via LAN/WAN)?

4. Draft High-Level DISC Life Cycle

The DISC Consortium

Overall design goal(s): *What is the general purpose(s) of the design process?*

E.g. to build a product, a research prototype, to achieve excellence in a certain area of research, to explore a particular approach, other; to particularly study certain aspects of the system or component. Comment, if needed, on whether the design goal is worthwhile, e.g. from the pov. of innovative research.

Hardware constraints: *Were there any a priori constraints on the hardware to be used in the design process?*

Constraints could be economical, derived from performance demands on the system (e.g. real-time), other. Describe the effects, if any, of the constraints (e.g. on vocabulary size, recognition quality). Comment, if needed, on the constraints and their effects from the pow. of best practice.

Software constraints: *Were there any a priori constraints on the software to be used in the design process?*

E.g. use of in-house or off-the-shelf speech recognisers, synthesisers, other. Describe the effects, if any, of the constraints. Comment, if needed, on the constraints and their effects from the pow. of best practice.

Customer constraints: *Which constraints does the customer (if any) impose on the system/component? Note that customer constraints may overlap with some of the other constraints. In that case, they should only be inserted once, i.e. under one type of constraint.*

E.g. hardware constraints, adequacy evaluation criteria, other. Note that research prototypes may be built to hypothetical customer constraints. The basic advantage of assuming hypothetical customers is that the developers force themselves to face realistic problems and hence to be accountable for any deviations from a realistic development life-cycle. Such deviations may be justifiable from many different points of view but they are not likely to be recognised as such unless the project has (simulated) realistic boundary conditions. Describe the effects, if any, of the constraints. Comment, if needed, on the constraints and their effects from the pov. of best practice.

Other constraints: *Were there any other constraints on the design process?*

E.g. on cost, personpower, purchase price, development time, development phases, standards conformation, knowledge in the developer team. Describe the effects, if any, of the constraints. Comment, if needed, on the constraints and their effects from the pow. of best practice.

Design ideas: *Did the designers have any particular design ideas which they would try to realise in the design process?*

E.g. innovative product features, innovative experimental features, other? Describe the effects, if any, of the ideas. Comment, if needed, on the ideas and their effects from the pow. of best practice.

Designer preferences: *Did the designers impose any constraints on the design which were not dictated from elsewhere?*

E.g. programming language preferences, development methodology. Describe the effects, if any,

of the preferences. Comment, if needed, on the preferences and their effects from the pow. of best practice.

Design process type: *What is the nature of the design process?*

E.g. exploratory research, product development, redesign, other. Comment on any peculiarities of the design process type from the pow. of best practice.

Development process type: *How was the system/component developed?*

E.g. through Wizard of Oz, based on human-human or human-computer dialogues, using development methodology X (describe it) using no particular methodology except "to see if things work", other. Comment on any peculiarities of the development process from the pov. of best practice.

Requirements and design specification documentation: *Is one or both of these specifications documented?*

Describe the specifications. In the absence of requirement and design specifications, the developers have no guidance wrt. when or to what extent they will have achieved their development objectives. Comment on any peculiarities of the specifications from the pov. of best practice.

Development process representation: *Has the development process itself been explicitly represented in some way? How?*

E.g. bits and pieces can be found in scientific papers, the entire process was carefully documented in semi-formal notation, most of the process has been systematically represented in reports or meeting protocols, other. The advantages of explicit development process representations are that these can be re-used, possibly in revised form, in new projects and with new developers coming on the team, and can support re-design and maintenance. The main disadvantage is that this represents an additional project cost. Comment on any peculiarities from the pov. of best practice.

Realism criteria: *Will the system/component meet real user needs, will it meet them better, in some sense to be explained (cheaper, more efficiently, faster, other), than known alternatives, is the system/component "just" meant for exploring specific possibilities (explain), other (explain)?*

Most interactive speech systems have something to do with real user needs. However, to appropriately address real user needs, the development process often needs to include extended end-user contact, extensive work on domain delimitation, clear up-front performance criteria, final adequacy criteria, extended quantitative and qualitative evaluation throughout the development process, an explicit development methodology etc. Comment on any peculiarities from the pov. of best practice.

Functionality criteria: *Which functionalities should the system/component have (this entry expands the overall design goals)?*

E.g. "allow users to do tasks X and Y", "include barge-in", "respond in real-time". Note that this entry is more general than, but may partially overlap with, the "grid" properties. Comment on any peculiarities from the pow. of best practice.

Usability criteria: *What are the aims in terms of usability?*

E.g. spontaneous unconstrained dialogue, usable with no training, usable with training in .. (explain), naturalness, high user acceptance, intuitively well-defined task domain, other. Comment on any peculiarities from the pow. of best practice.

Organisational aspects: *Will the system/component have to fit into some organisation or other, how?*

E.g. partially replace the switchboard operator, require backup for difficult or incomprehensible queries.

Customer(s): *Who is the customer for the system/component (if any)?*

E.g. the system/component is custom-built, addresses a specific market segment, has not customers but produces spin-off products, has "simulated" customers, other. Comment on any peculiarities from the pov. of best practice.

Users: *Who are the intended users of the system/component?*

E.g. users speaking High German, or Swedish, walk-up-and-use users, specialised user group X. Comment on any peculiarities from the pov. of best practice, e.g. is walk-up-and-use an appropriate paradigm for the application?

Developers: *How many people took significant part in the development? Did that cause any significant problems, such as time delays, loss of information, other (explain)? Characterise each person who took part in terms of novice/intermediate/expert wrt. developing the system/component in question and in terms of relevant background (e.g., novice phonetician, skilled human factors specialist, intermediate electrical engineer).*

Comment on any peculiarities from the pov. of best practice.

Development time: *When was the system developed? What was the actual development time for the system/component (estimated in person/months)? Was that more or less than planned? Why?*

Comment on any peculiarities from the pov. of best practice.

Requirements and design specification evaluation: *Were the requirements and/or design specifications themselves subjected to evaluation in some way, prior to system/ component implementation? If so, how?*

However difficult this may be to do in any formal way, it is essential to good development practice to make explicit and systematically evaluate the requirements or design specifications. Comment on any peculiarities from the pov. of best practice.

Evaluation criteria: *Which quantitative and qualitative performance measures should the system/component satisfy?*

The definition, from early on in the development process, of clear, relevant and appropriate evalualation criteria, and the continuous evaluation of progress using those criteria, are main

characteristics of best practice in development and evaluation of spoken language dialogue systems.

For the relevant criteria below, state their definitions, describe the performance targets and state whether these were achieved.

(SR) Word error rate:

(SR) Coverage of vocabulary

(SS) Synthesis quality:

(NLUG) Perplexity:

(NLUG) Grammatical coverage:

(NLUG) Coverage of lexicon:

(NLUG) Complete parses:

(NLUG) Robustness - wrt. lack of lexical items:

(NLUG) Robustness - wrt. parsing errors:

(DM) Robustness - wrt. topic identification:

(DM) Robustness - wrt. unexpected deviations from the dialogue plan:

(DM) Number of turns:

(DM) Average and max utterance length (for user and for system):

(DM) Average number of "long" turns per dialogue turns (for user and for system):

(DM) Average number of word types and word tokens per dialogue:

(DM) Cumulative type/token ratio:

(DM) Number of questions in relation to total number of turns (for user and for system):

(DM) Complexity of interaction model, e.g. in terms of number of nodes if a graph representation is used:

(DM) Number of ad hoc generated phrases in relation to total number of turns (WOZ only):

(DM) Average number of ad hoc generated jumps per dialogue (WOZ only, system-directed dialogue):

(DM) Naturalness - mixed initiative dialogue:

(DM) Naturalness - no interaction problems:

(DM) Transaction success:

(DM) How succesful is the system wrt. the interpretation of indirect speech acts, and how does this depend on the dialogue management approach chosen:

(HF) User satisfaction and other subjective parameters (explain):

(HF) Average time for task completion:

(HF) Do the system phrases conform to the cooperativity guidelines (individually as well as in context):

(System) Re-usability:

Robustness - other (explain):

Naturalness - other (explain):

Multimodality: Describe the evaluation of the multimodal aspects of the system, if any. Method(s)? Results? Was the evaluation procedure appropriate?

Comment on any peculiarities from the pov. of best practice.

Evaluation: *At which stages during design and development was the system/component subjected to testing/evaluation? How? Describe the results.*

Describe, one-by-one, the aspects that were evaluated, when, the set-up and the methodologies used, e.g. Wizard of Oz scenario-based, glassbox, blackbox, progress (comparing successive measurements), diagnostic, performance, adequacy, acceptance, field, objective, subjective. Number of subjects/users involved in each test.

How was data collection done (logfiles, corpora, questionnaires, interviews, other)? Describe the corpora, etc. Was data annotation done? How? Which information has been extracted from the data?

What were the results? Is test material/data/test suites (and/or a description of the test conditions) available? Can the test be replicated? Can anybody perform the tests?

Is anything stated about comparability of the test(result)s with those of other systems/components of similar scope? Comment on any peculiarities from the pov. of best practice.

Mastery of the development and evaluation process: *Of which parts of the process did the team have sufficient mastery in advance? Of which parts didn't it have such mastery?*

Note that lack of mastery of parts of the development process is a normal condition in research projects which often serve in part as competence-building exercises. Comment on any peculiarities from the pov. of best practice.

Problems during development and evaluation: *Were there any major problems during development and evaluation? Describe these.*

E.g. problems of collaboration in the team, major delays caused by ?, difficulties in satisfying specification requirement X, developer Y left the team, lack of quality of what was delivered by some in the team. Comment on any peculiarities from the pov. of best practice.

Development and evaluation process sketch: *Please summarise in a couple of pages key points of development and evaluation of the system/component. To be done by the developers.*

Comment on any peculiarities from the pov. of best practice.

Component selection/design: *Describe the system components and their origins.*

E.g. off-the-shelf, based on somebody else's parser (specify), built in-house for the application, other (specify). Comment on any peculiarities from the pov. of best practice.

Robustness: *How robust is the system/component? How has this been measured? What has been done to ensure robustness?*

Comment on any peculiarities from the pov. of best practice.

Maintenance: *How easy is the system to maintain, cost estimates, etc.*

Note that maintenance may include continued development and re-design. Are there guidelines for maintenance (of, e.g., lexicon and grammar)? Comment on any peculiarities from the pov. of best practice.

Portability: *How easily can the system/component be ported?*

E.g. OS dependencies, machine dependencies. Comment on any peculiarities from the pow. of best practice.

Modifications: *What is required if the system is to be modified?*

Comment on any peculiarities from the pow. of best practice.

Additions, customisation: *Has a customisation of the system been attempted/carried out (e.g. modification of a part of the vocabulary, new domain/task, etc.)? Has there been an attempt to add another language? How easy is it (how much time/effort) to adapt/customise the system to a new task? Is there a strategy for resource updates (e.g. a predefined sequence of update steps to be performed if a new item is added to the lexicon or if a new grammatical description is added to the grammar)? Is there a tool to enforce that the optimal sequence of update steps is followed (e.g. a menu-driven update interface, etc.)?*

Comment on any peculiarities from the pov. of best practice.

Property rights: *Describe the property rights situation for the system/component.*

Comment on any peculiarities from the pov. of best practice.

Documentation of the design process

E.g. specification documents or parts thereof, architecture diagram (mandatory), user scenario(s), transcribed dialogue(s), other.

References to additional project/system/component documentation

Please refer to relevant information.

5. Draft Specific DISC Life Cycles

The DISC Consortium

5.1 Speech recognition life cycle

1. Evaluation criteria for SR

- Word error rate
- Human fall back rate

5.2 Speech generation life cycle

Development process type (modification of the general question in the life cycle above)

How will the system/component be developed? Is the system developed from scratch for or modified for the present use?

5.3 Natural language understanding and generation life cycle

Environment constraints

What is the usage environment of the system (e.g. noisy, telephone, ...)?)

Test material

- Is resource-building based on a lexical/grammatical analysis of a corpus? How is the lexical fragment defined. Is this definition usable as test material? Is there a testsuite for systematic checking of the grammatical coverage?

- Maintenance Additions, Customization: are there guidelines for lexicon and grammar maintenance? Is there a strategy for resource updates (e.g. a pre-defined sequence of update steps to be performed, if a new item is added to the lexicon or if a new grammatical description is added to the grammar)? Is there a tool to enforce that the optimal sequence of update steps is followed (e.g. a menu-driven updated interface, etc.)?

- Robustness: is there test material to check certain aspects of robustness, such as for example:

- robustness towards lack of lexical items;

- robustness towards parsing errors;

- robustness towards unexpected deviations from the dialogue plan.

5.4 Dialogue management life cycle

Evaluation criteria for Dialogue Management

The list below shows a series of evaluation criteria that are relevant to dialogue manager evaluation. Whilst not all of them are applicable to all systems, each criterion may help throw light on the adequacy and performance of some dialogue managers. They need not all be included in the "official" list of evaluation criteria stated in the requirements specification. Even if not included, they may still be relevant for evaluating how good or bad the dialogue manager is and what progress is being made during its development.

- Adequacy of (non-) distinction between novice and expert users.

- Adequacy of dedicated processing of ellipsis.

- Adequacy of domain inferences.

- Adequacy of dialogue manager support for the speech and/or language layers.

- Adequacy of information to the users on how to interact with the system.

- Adequacy of information to the users on the system's domain and task coverage.

- Adequacy of initiative distribution among user and system relative to the task.

- Adequacy of multimodal input fusion, i.e. of combining more or less simultaneous input messages expressed in different modalities into a single semantic representation or sub-task contribution.

- Adequacy of output distribution over speech and other output modalities in multimodal SLDSs.

- Adequacy of operator fallback strategy.

- Adequacy of strategy for identifying and responding to out-of-task and/or out-of-domain input.

- Average time for task completion as compared to other ways of solving the same task.

- Complexity of the interaction model expressed, e.g., in terms of number of nodes if a graph representation is used.

- Conformance of system phrases to the cooperativity guidelines (individually as well as in context).

- Co-reference interpretation success.

- Cost per transaction as compared to other ways of solving the same task.

- Database information sufficiency.

- Degree of utilisation of the knowledge sources available to the dialogue manager.

- Dialogue segmentation adequacy.

- Ease of maintenance/modification of the dialogue manager and/or of its individual modules.

- Feedback strategy sufficiency: information feedback.

- Feedback strategy sufficiency: processing feedback.

- Indirect speech acts interpretation success.

- Naturalness through mixed initiative dialogue.

- Number of dialogue interaction problems.

- Number of turns to complete a task.

- Real-time performance.

- Relevance and success of predictions.

- Re-usability of the dialogue manager and/or of its individual modules.

- Robustness - wrt. error loops and graceful degradation.

- Robustness - wrt. unexpected (user) deviations from the dialogue plan.

- Sub-task or topic identification success.

- Sufficiency of dialogue histories (linguistic, topic, task, performance).

- Sufficiency of meta-communication facilities: system-initiated repair, system-initiated clarification, user-initiated repair, user-initiated clarification.

- Task and domain model coverage: Sufficient? Delineated in a principled and intuitive way?

- Transaction success.

- Translation success of spoken language translation (support) systems.

- User model adequacy.

- User satisfaction and other subjective parameters (explain).

5.5 Human factors life cycle

Context analysis

User study
- Representative user defined?
- Personal details?
- Experience considered?
- Language/dialectical issues considered?
- Knowledge defined; domain, system, general?
- Organisational position?
- Job/role of users considered?
- Studies formally documented ?Domain/environment study
- Working environment description, e.g. ambient noise etc. ?
- Organisational effects of system considered

Requirements capture

Task analysis
- Vocabulary definition done?
- Core task description done?
- Task description formalism used?

Usability measures
- Quantitative measures identified?
- Qualitative measures ?

Prototyping
- Pen&paper WOZ
- WOZ test?
- Were subjects aware there was a wizard?
- Use of WOZ recordings as speech corpus
- Use of WOZ recordings to refine dialogue design?
- Were walk-through techniques used?
- Subjective evaluations carried out?

- Evaluation of prototypes documented?

- Prototypes thrown away?

- Documents prototyped?

- Did evaluations involve subjects using scenarios?

- Were the experimenters aware that scenario wording may influence user utterance wording?

Evaluation and field trials

Evaluation description
- When
  - During requirements capture
  - In iterations after implementation
  - Once on implementation
- Who with?
  - In-house
  - Friendly client
  - Cold client
- Size
  - Alpha
  - Beta

Evaluation methods
- WOZ
- Cognitive walk-through
- Heuristic
- Statistical

Evaluation type
- Performance
- Diagnostic
- Subjective (adequacy)

Typical industrial life cycle wrt. human factors

Initial Client Meeting: The meeting is an opportunity for the client to outline its requirements.

Tender: The client offers the contract. This will contain a high level 'service' description which will specify functionality and a rudimentary dialogue. No user information will be included.

Response: The third part provides a response specifying which aspects of the tender can be complied with. Since no dialogue requirements have been specified there is no assessment of how user requirements will be met.

Requirements Documentation: In conjunction with the client, requirements will be specified. This will specify the functionality to be supported and will *not* assess whether this is actually the functionality that is required by the end-users. This document will be updated on a regular basis. The client and contractor will sign this agreement.

Functional Requirements: A contractor document which will specify system and dialogue design, including suggested system outputs. The dialogue and prompt design will not be validated in any way, save with the management level (not end-users) of the client company. No prototyping of dialogue flow and/or prompting will be carried out.

Testing: Functional testing. No usability testing as no usability targets are specified in the functional specification.

Implementation and Trials: The system will be integrated. System problems (excluding usability) will be ironed out first. It is often only much later that it is realised problems are a result of poor dialogue design. This includes the functionality offered, command vocabularies and tasks decompositions.

Revision: At this stage users will finally be asked what are the problems. Invariably, customer confidence has been damaged.

5.6 Systems integration life cycle

System testing and maintenance

- Is a test-bed available for the SLDS?

- Are tools for system maintenance available?

- What communication protocol do these use?

- Is runtime (debug) output available?