ISO/TC37/SC 4 **N442 rev05**

Date: 2009-10-05

ISO/CD 24617-2-2009-10-05

ISO/TC 37/SC 4/WG 2

Secretariat: KATS

Language resource management — Semantic annotation framework (SemAF) — Part 2: Dialogue acts

Gestion des ressources linguistiques — Cadre d'annotation semantique (SemAF) — Partie 2: Actes de dialogue

Document type: International Standard
Document subtype: Not applicable
Document stage: (20) Working Draft
Document language: E

Reference Number
ISO/CD 24617-2-2009-10-05(E)

cd5-iii.tex 2009-10-05

ISO copyright office
Case postale 56 CH-1211 Geneva 20
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Language resource management — Semantic annotation framework (SemAF) — Part 2: Dialogue acts

Foreword

International Standard 24617-2 was prepared by Technical Committee ISO/TC 37, *Terminology and Other Language Resources*, Subcommittee 4, *Language Resource Management*, Working Group 2, *Representation schemas*, following up on the EU-supported project LIRICS (Linguistic Infrastructure for Interoperable Resources and Systems) in collaboration with TC 37/SC 4 ad-hoc Thematic Domain Group 3, Semantic content, and the ACL SIGSEM Working Group on the Representation of Multimodal Semantic Information.

NOTE The LIRICS project was coordinated by Laurent Romary, INRIA-LORIA (laurent.romary@loria.fr); the ACL SIGSEM Working Group is headed by Harry Bunt (harry.bunt@uvt.nl), Tilburg University.

ISO 24617 consists of the following parts, under the general title *Language resource management - Semantic annotation framework*:

— *Part 1: Time and events* — *Part 2: Dialogue acts*

Part 1 has been approved for registration as a DIS, draft international standard.

The main parts of ISO 24617-2 are:

a) Scope

b) Terms and definitions

c) Basic concepts and metamodel

d) Defining communicative functions

e) Dialogue segmentation

f) Core dimensions

g) Core dialogue acts

h) Specification of ISO-DiAML, a formal annotation (specification) language for dialogue acts

i) Principles for schema extension and restriction

In addition, there are two normative and five informative annexes. The normative Annex A provides annotation guidelines. Annex B (informative) provides annotated a number of completely annotated examples. The normative Annex C contains the definition of an XML-based representation format for DiAML-annotations. Annex D (normative) contains the data categories for the core communicative functions of this standard. This is a subset

(with minor modifications) of the semantic data categories that were compiled in the EU project LIRICS, and approved by ISOTC 37/SC 4/TDG 3, Semantic content for inclusion in the ISO Data Category Registry. Annex E (informative) contains a number of data categories for non-core communicative functions. Annex F (informative) summarizes a study by Petukhova & Bunt (2009) on criteria for distinguishing core dimensions and core communicative functions, partly based on a survey of 18 existing dialogue act annotation schemas. This study was performed as part of the project of establishing the current standard proposal. Annex F (informative) contains a very brief summary of a technical report by Volha Petukhova and Harry Bunt, published by the Tilburg Center for Creative Computing (TiCC) at Tilburg University. Annex G (informative) provides editorial and authorship information for the current document, with a list of editors, contributors, and meetings. The document concludes with a bibliography.

Introduction

This standard proposal results from preliminary studies conducted in the ad-hoc Thematic Domain Group 3 (TDG 3) of ISO subcommittee TC 37/SC 4. The TDG 3 group was created in order to investigate the feasibility and timeliness of standardization efforts in the area of semantic annotation and representation. Five areas of semantic information were chosen to focus the investigations:

a) temporal entities and relations

b) referential entities and relations

c) semantic roles and argument structures

d) dialogue acts

e) discourse relations

The investigations were reported and discussed in a series of TDG 3 meetings, listed in Appendix D. In order to obtain input and feedback from a larger community, the Working Group on the Representation of Multimodal Semantic Information was created within SIGSEM, the Special Interest Group on Computational Semantics of the Association for Computational Linguistics (ACL). The ISO TDG 3 and the SIGSEM WG have been held four joint workshops (see Appendix D).

In order to boost the TC 37/SC 4 work aiming at the development of interoperable language resources, a project was set up within the eContent framework of the European Union. This project, called LIRICS (Linguistic Infrastructure for Interoperable Resources and Systems) ran from January 2005 to July 2007. Its deliverables served as input for ISO TC 37/SC 4 activities. The project adopted the areas a) - d) mentioned above as its semantic focus; its deliverables included feasibility studies, comparative studies of related efforts, methodological proposals, and data categories for semantic annotation.

As an outcome of the ISO TDG 3 and LIRICS studies, the area of temporal information was chosen for the Part 1 of an international standard for semantic annotation, of which the present proposal forms Part 2.

It is envisaged that this standard will have more parts, concerned with other areas of semantic information (such as semantic roles, reference, spatial information, quantification,..) and one part concerned with semantic annotation and representation from an integrative point of view, dealing with the combined annotation of semantic information from several areas.

1 Scope

Utterances in dialogue have one or more communicative functions which characterize the type of dialogue act that is performed; these functions carry an essential part of the meaning of dialogue utterances. An adequate characterization of this aspect of meaning requires a coherent system of well-defined communicative functions. As part of the standard for a semantic annotation framework, this document focuses on the formulation of a normative set of guidelines for annotating dialogues by specifying its semantic units, their communicative functions, and the type of semantic content that is conveyed. Since the semantic units in dialogue are often multifunctional, annotating dialogue with communicative functions should take a multidimensional view.

The standard defines an articulate approach to multidimensional dialogue act annotation, providing criteria for identifying dimensions and defining communicative functions, and specifying data categories for a core set of dimensions and a core set of communicative functions. An annotation language is defined for multidimensional dialogue act annotation (DiAML: Dialogue Act Markup Language), with an XML-based concrete syntax and a semantics.

As an effort within ISO TC 37/SC4, Language resources management, this standard has a focus on the use of language; however, since natural dialogue combines the use of language with that of nonverbal elements (facial expressions, gestures, gaze direction, nonverbal vocal sounds,..), and since interactive meaning is often determined by the combination of verbal and nonverbal elements, the standard aims to provide concepts and representational means for the annotation of multimodal dialogue behaviour more generally.

2 Normative references

For this international standard there are two main normative references:

- ISO 8879: 1986 (SGML) as extended by TC2 (ISO/IEC JTC 1/SC 34 N029: 1998-12-06).
- ISO 19757-2, Document Schema Definition Language, part 2.
- ISO 12620-1, Data categories
- ISO DIS 24612 Linguistic Annotation Framework
- ISO 24601-1 Feature Structure Representation

The first reference allows the use of XML as a markup language for semantic annotation and the second the Relax NG schema language for designing XML representations of annotation structures.

3 Terms and definitions

For the purposes of ISO 24617-2, the following terms and definitions apply.

3.1
addressee
Dialogue **participant** oriented to by the **speaker** in a manner to suggest that his words are particularly intended for him or her, and that some response is therefore anticipated from him/her, more so than from the other participants.

NOTE Source: Goffman (1981).

3.2
allo-feedback act
Feedback act where the **sender** elicits information about the **addressee**'s processing of what was said in the dialogue, or where the **sender** provides information about his perceived processing by the **addressee** of what was said before.

3.3
auto-feedback act
Feedback act where the **sender** provides information about his processing by the **addressee** of what was said before.

3.4
communicative function
Property of a **dialogue act**, specifying how the act's **semantic content** changes the addressee's information state when (s)he understands the dialogue act.

3.5
context or context model
See **information state**.

3.6
dialogue act
Semantic unit in the description of dialogue behaviour, characterizing how the **information state**(s) of the participant(s) at whom the behaviour is directed are changed when he/they understands) the behaviour.

3.7
dimension
Aspect of participating in dialogue which can be addressed by **dialogue acts**.

3.8
feedback act
Dialogue act which provides or elicits information about the **sender**'s or the **addressee**'s processing of something that was said in the dialogue.

3.9
feedback dependence relation
Relation between a **feedback act** and the stretch of communicative behaviour, the processing of which is addressed by the **feedback act**.

3.10
functional dependence relation
Relation between a **dialogue act** used to respond to a previous **dialogue act**.

NOTE Examples: question - answer relation; offer - acceptance of offer relation.

3.11
functional segment
Minimal stretch of communicative behaviour that expresses one or more **dialogue act**s.

NOTE A functional segment does not need to be continuous, and may be spread over multiple turns.

3.12
information state
The totality of a dialogue **participant**'s beliefs, assumptions, expectations, goals, preferences, hopes, and other attitudes that may influence his interpretation and generation of communicative behaviour in dialogue.

NOTE The terms **context** and **context model** are considered as synonyms of **information state**.

3.13
participant
Person or artificial agent involved in dialogue.

3.14
semantic content (of a dialogue act)

Objects, propositions, events, actions, and other entities that a dialogue act refers to or uses as arguments of predicates.

3.15
semantic content type
Kind of objects, events, actions, and other entities that a dialogue act refers to or uses as arguments of predicates, according to some typology of kinds of entities.

NOTE See also **dimension**, and the discussion in sections 5.

3.16
sender
Dialogue **participant** who produces a dialogue act, through producing the communicative behaviour that expresses the dialogue act.

NOTE The term **speaker** is considered as a synonym of **sender**.

3.17
turn
Stretch of speech produced by one **participant**, bounded by periods of speech inactivity of that **participant**.

4 Purpose and justification

The purpose of this proposed standard is to support the creation of interoperable resources for the study of dialogue and the development of dialogue systems, through the establishment of an international standard for annotating dialogues with dialogue act information.

The notion of a dialogue act plays a key role in studies of dialogue phenomena, in the description of the interpretation of communicative behaviour of participants in dialogue, and in the design of dialogue systems. The idea of interpreting dialogue behaviour in terms of communicative actions such as statements, questions, promises, requests, and greetings, goes back to speech act theory (Austin, 1962; Searle, 1969), which has been an important source of inspiration for modern dialogue act theory. But where speech act theory is primarily an action-based approach to meaning within the philosophy of language, dialogue act theory is an empirically-based approach to the computational modeling of communication, in particular of linguistic and nonverbal communicative behaviour in dialogue.

As a way to describe meaning in communicative behaviour, dialogue acts are semantic concepts that can be defined by the way they are intended to affect the information state of an addressee when (s)he understands the behaviour. For instance, when an addressee understands the utterance *Do you know what time it is?* as a question about the time, then the addressee's information state is updated to contain (among other things) the information that the speaker does not know what time it is and would like to know that. If, by contrast, an addressee understands that the speaker used the utterance to reproach the addressee for being late, then the addressee's information state is updated to include (among other things) the information that the speaker *does* know what time it is. Distinctions such as that between a question and a reproach concern the *communicative function* of a dialogue act; the objects, properties, relations, events, etc. that are referred to, constitute its *semantic content*. The communicative function of a dialogue act specifies how an addressee should update his/her information state with the information expressed in the semantic content, when (s)he understands the speaker's dialogue act.

Dialogue act annotation is the activity of marking up stretches of dialogue with information about the dialogue acts which are performed in these stretches of behaviour, and is usually limited to marking up their communicative functions using a given set of such functions (a 'tagset').

During the 1980s and 1990s a number of alternative dialogue act annotation schemas have been developed, such as those of the TRAINS project in the US, the Map Task studies in the UK, and the Verbmobil project in Germany. These schemas were all designed for a specific purpose and a specific application domain; they contained partly different sets of communicative functions and made use of overlapping, often mutually inconsistent

terminology. In the 1990s a group of dialogue researchers came together as the Discourse Research Initiative, and drafted the general-purpose schema for multidimensional dialogue act annotation called DAMSL: Dialogue Act Markup using Several Layers (Allen and Core, 1997; Core et al., 1998). With its focus on multidimensionality and domain-independence, this represented an important step forward in dialogue act annotation.

The design of DAMSL was unfortunately left in an unfinished state, due to the lack of a stable organizational structure behind the initiative. Several researchers have constructed variations and extensions of the DAMSL schema for their own purposes, such as Switchboard-DAMSL (Jurafsky, 1997) and COCONUT (Di Eugenio et al., 1998).

Preliminary studies in the ISO TC 37/SC 4 ad-hoc Thematic Domain Group on Semantic Content (TDG 3) have indicated that the area of dialogue act annotation is sufficiently mature for a new and sustained effort to design a comprehensive general framework for dialogue act annotation, supported by the stable organizational structure of ISO. In the European LIRICS project, a spin-off of TDG 3, a set of data categories for communicative functions has been defined following ISO standard 12620. This set of data categories has been documented in LIRICS deliverable D4.3 and endorsed by TDG 3. The data categories have been tested for their usability and coverage in the manual annotation of dialogues in English, Dutch and Italian (documented in LIRICS deliverable D4.4), and forms an important part of the background of the present proposal.

While the current state of the art seems to make it feasible to develop standard annotation concepts for a comprehensive set of core dialogue acts, researchers and application designers should also be supported in adding their own concepts for specific domains or purposes, or selecting parts of a comprehensive schema. The project will therefore provide general principles and guidelines for extending its core concepts and for selecting coherent parts of it.

5 Basic concepts and metamodel

The term ‘dialogue act’ is often used rather loosely in the sense of ‘speech act used in dialogue’. Such a characterization hardly does justice to the semantic nature of dialogue acts. A more accurate characterization is:

(1) *A dialogue act is a unit in the semantic description of communicative behaviour in dialogue, specifying how the behaviour is intended to change the information state of a dialogue participant who understands the behaviour* (i.e. as intended by the speaker).

The specification of intended information state changes, also called ‘information state updates’ (or ‘context changes’), requires two ingredients: (1) a specification of the information with which the information state is to be updated; (2) a specification of the way in which that information is to be used in updating the information state. These two ingredients correspond to the semantic content and the communicative function of the dialogue act, respectively.

A dialogue act being a unit in the semantic description of communicative behaviour, the question arises what stretches of such behaviour are considered as corresponding to dialogue acts. The identification of meaningful stretches of dialogue is called the segmentation of the dialogue. Spoken dialogues are traditionally segmented into *turns*, defined as stretches of communicative behaviour produced by one speaker, bounded by periods of inactivity of that speaker. The term ‘turn’ reflects the idea that the participants in a dialogue take turns in speaking, and that a dialogue can be cut up into sequences of communicative activity of one speaker followed by that of another. This is not really the case. In natural communication, the participants often produce overlapping speech rather than sequences of single-speaker turns; this happens especially when more than two participants are involved (see e.g. Campbell, 2004). Moreover, in natural conversation the use of speech is combined with facial expressions; head, hand and shoulder gestures; body posture; gaze direction and nonverbal sounds (laughs, sighs, sucks, chuckles,..), and all participants are most of the time performing some nonverbal communicative activity.

Turns, in the sense defined above, can be quite lengthy and complicated, and are for most purposes too coarse as the stretches of behaviour to assign communicative functions to. Communicative functions can be assigned

more accurately to smaller units, which we call *functional segments*, and which we define simply as the functionally relevant minimal stretches of communicative behaviour. See further section 8 for more details about dialogue segmentation.

According to the definition given in (1), a dialogue act has at least two participants: (1) an agent whose communicative behaviour is interpreted, usually called the “speaker”, or “sender”; and (2) a participant to whom he is speaking and whose information state he wants to influence, called the “addressee” (also called “hearer" or "recipient"). There may of course be more than one addressee. For natural multimodal dialogue, where some of the dialogue acts are expressed in speech, some in a combination of speech and nonverbal elements, and some purely nonverbally, its is best to use the term “sender" for the agent who performs a dialogue act and the term “speaker" for a participant who produces a turn unit.

Besides sender and addressee(s), there may be various types of side-participants who witness a dialogue without participating in it. The presence of side-participants may influence the communicative behaviour of the participants, if these are aware of their presence, as in a television interview or a talk show. Clark (1996) distinguishes between ‘side-participants’, ‘bystanders’, and ‘overhearers’, depending on the role that they play in the communicative situation.

Of the two most central aspects of a dialogue act, the communicative function and the semantic content, the former corresponds intuitively to the *type of action* that is performed, and as mentioned above, the term “dialogue act annotation” is commonly used to describe the assignment of communicative function labels to stretches of dialogue. A semantically more complete characterization of a functional segment also provides information about the *type of semantic content*. For example, the DAMSL annotation schema distinguishes four layers: Information Level, Communicative Status, and Forward- and Backward-Looking Functions, of which Information Level has three possible values: Task, Task Management, and Communication, also known as ‘Dialogue Control’. These values indicate whether the semantic content of the dialogue act is concerned with performing the task that underlies the dialogue, or with discussing the performance of the task, or with the communication. This is a coarse 3-way distinction of semantic content types. This standard makes a more fine-grained distinction of semantic content type by distinguishing communication-related information (DAMSL’s ’Communication’ type) into a number of subtypes, such as information about the processing of something that was said before (feedback information), about the allocation of turns (turn management information), or about the structuring of the dialogue (topic and dialogue structure information). These types of semantic content are also called ‘dimensions’, and are discussed in more detail below in Section 9 7.

The distinction between communicative function and semantic content is basically the same is that between illocutionary force and propositional content in speech act theory. The use of a different terminology is justified by the differences between the respective theoretical frameworks in which the corresponding concepts play their roles, both in the different ways in which communicative behaviour is interpreted as speech acts/dialogue acts, and in the ways in which speech acts/dialogue acts themselves are interpreted. An essential difference is that, where speech act theory associates a functional dialogue segment with a single speech act, dialogue act theory interprets utterances as having multiple communicative functions (i.e. as expressing multiple dialogue acts), taking a multidimensional view on communication. This difference is reflected in the fact that existing dialogue act annotation schemas use taxonomies of communicative functions that are very different from the typology used in speech act theory, which distinguishes five categories of speech acts: *assertives, commissives, directives, expressives*, and *declarations*. As for the use of ‘semantic content’ rather than ‘propositional content’, the choice of the latter term is rather unfortunate since dialogue acts often are not about propositions but about other types of content. For example requests, suggestions, instructions, promises, and other directive and commissive acts are about actions; feedback acts are about processes; and WH-questions and their answers are about sets of objects. Semantically, a communicative function and a semantic content together define an update operation on information states.

Many dialogue acts have a responsive character, being semantically dependent on one or more dialogue acts that occurred earlier in the dialogue. This is for example the case for answers, whose meaning crucially depends on which question is being answered; but also for the acceptance or rejection of offers, suggestions, invitations, and requests; and for accepting an assignment of the turn, or responding to a greeting. For these dialogue acts, an important aspect that may be marked up is the relation to the ‘antecedent’ on which their meaning depends.

Feedback-providing and eliciting acts are in a sense also responsive, as they relate to what happened earlier in the dialogue, but in a different way. Feedback acts are concerned with the processing of what was said before - such as its perception, interpretation, or evaluation. The difference is that feedback acts are about the processing of *what was said* earlier, rather than responding to the dialogue acts that were expressed. The following examples illustrates this.

(2) 1. A; What happens to the result when the third condition is not met?
2. B: The third condition?
3. B: It will decrease with 15-20%.

B's utterance 3 is used to give an answer to the question in 1; its meaning (notably its semantic content) depends on that of the question, and it responds not to what was said but to its *interpretation* as a question. Utterance 2, by contrast, checks the correctness of what was understood, rather than responding to the question that it expresses.

Note that nonverbal feedback, for instance in the form of nodding of using backchannels like *m-hm*, may relate to what is *currently* said, rather than to what was previously said. This is also the case for speech editing acts like self-corrections (*on Tuesday I mean Thursday*) and completions of what the partner is trying to say.

These examples all show that, besides a semantic dependence relation between dialogue acts, for some types of dialogue acts (in particular for feedback acts) we should take into account how they relate to what was previously said. We call the former 'functional dependence relations' and the latter 'feedback dependence relation'.

In the characterization of the notion of a dialogue act and its realization, as given so far, the following key elements occur, which will form the backbone of the proposed metamodel for dialogue act annotation:

- sender (or 'speaker')
- addressee(s)
- participants in other roles (such as overhearers)
- functional segment
- dialogue act
- communicative function
- semantic content type
- functional dependence relation
- feedback dependence relation

The metamodel in Figure 1 shows a representation of the fundamental upper-level concepts that are involved in dialogue act annotation. Each dialogue act is related to one functional segment, but each functional segment is related to one or more dialogue acts, reflecting the possible multifuncitonality of functional segments (see Bunt, 2009; submitted). A turn unit is produced by a single speaker, regardless of whether an addressee interjects with nonverbal feedback or backchannels. (These are seen as separate though concurrent turns.) A functional segment is usually part of a single turn unit but may also spread over multiple turns, as we will see in Section 8. The speaker of a turn unit is the sender of the dialogue acts realized by functional segments of the turn unit, and vice versa: the sender of a dialogue act is the speaker of all the turn units that the functional segment which realizes the dialogue act may be part of – constraints that the metamodel cannot express. Because of this one-to-one relation between speakers and senders of dialogue acts, we will, if there is no danger of confusion, also speak of the speaker of a dialogue act, as is customary in the literature.

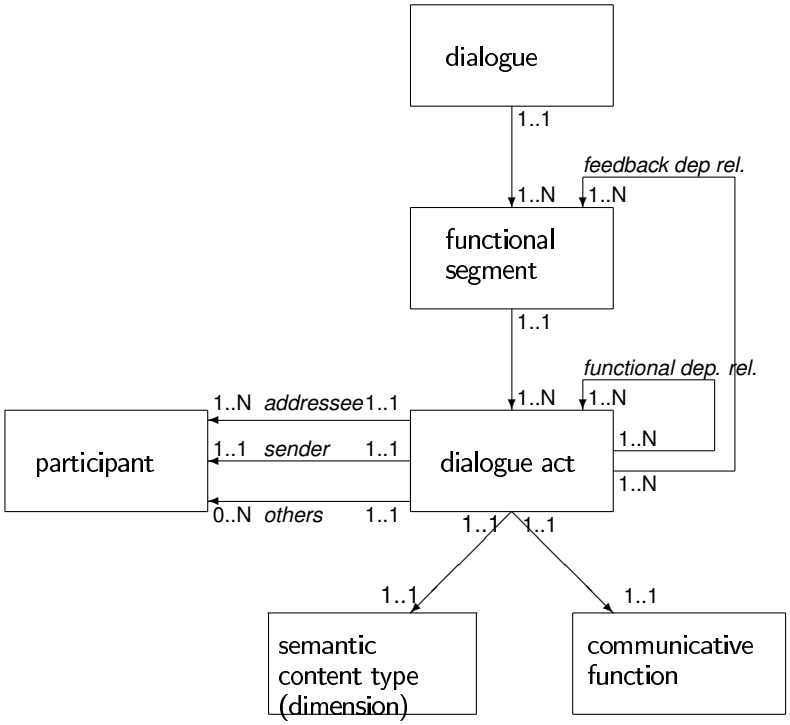


Figure 1 — Metamodel for dialogue act annotation.

6 Defining communicative functions

6.1 Approaches to the definition of communicative functions

Existing dialogue act annotation schemas use either one or both of the following two approaches to defining communicative functions: (1) in terms of the intended effects on addressees (notably in terms of the addressees beliefs about the speaker's intentions and associated attitudes, like beliefs and preferences); (2) in terms of properties of the signals that are used. For example, questions, invitations, confirmations, and promises are nearly always defined in terms of speaker intentions, while repetitions, hesitations, and dialogue openings and closing are typically defined by their form.

Defining a communicative function by its linguistic form has the advantage that its recognition can be straightforward, but has to deal with the fundamental fact that the same linguistic form can often be used to express different communicative functions. For example, the utterance *Why don't you start?* has the form of a question, and can be intended as such, but it can also be used to invite or encourage somebody to start. Similarly for *Do you know what time it is?*, which can be both an indirect request to tell the time and a reproach for being late, and for 'declarative questions' like *You're going home tomorrow* which look like statements but are intended as (check) questions.

Form-based definitions of communicative function are also in danger of being purely descriptive of certain forms of behaviour, without saying anything about their interpretation. For example, a description like 'response' only indicates that something is being reacted to, but it does not say what is the communicative function of the response. Similarly for 'hesitation', 'repetition', and 'reply'.

We will take a strictly semantic approach to the definition of communicative functions in terms of the effects on an addressee's information state that occur when the addressee understands the speaker's behaviour. 'Understanding' here means: as the speaker intended his behaviour to be understood. Note that, while we do not take linguistic form to be part of the definition of a communicative function, we do insist that for every communicative function that is distinguished there are ways in which a speaker can indicate that his behaviour should be understood as having that function, by shaping his (linguistic and nonverbal) behaviour so as to have certain observable features that are indicative for that function. This requirement puts all communicative functions on an empirical basis.

Dialogue act annotation schemas have by and large ignored the phenomenon known as 'indirect speech acts', which is the phenomenon that a speaker says one thing but is understood to mean something else or something additional. Questions of the form *Do you know [X]*, such as *Do you know what time it is?* illustrate this: while a question of this form seems to be asking whether the addressee knows something, it is mostly intended (and understood) to ask for the information $[X]$, such as what time it is.

The DIT^{++} taxonomy of communicative functions (Bunt 2004; Bunt & Girard, 2005) views indirect questions and requests as having a communicative function which is slightly different from, though closely related to, that of the corresponding direct form, because their performance is thought to have slightly different effects on information states of addressees. For example, the difference between *Where is Lee's office?* (SetQuestion) and *Do you know where Lee's office is?* (IndirectSetQuestion) would be that in the indirect version the speaker does not express an expectation that the addressee knows the answer to his question, whereas the direct version does (see Bunt, 2000 for further discussion).

The full complexity of the phenomenon of indirect speech acts is beyond the scope of this standard. However, an important class of indirect speech acts will be covered, as described in section (10.3).

6.2 Communicative function recognition

Successful communication depends on addressees understanding the communicative functions of the speaker's utterances in the way intended by the speaker. These functions are inferred from the utterance surface characteristics in combination with their models of the dialogue context. Such models include assumptions about each other's beliefs and goals, as well as knowledge of the dialogue history and about the activity which motivates the dialogue. It is in general not possible to recognize the communicative functions of utterances from their surface form only, since virtually every utterance form can be used with different functions; only in a particular dialogue context can the utterance features be interpreted unambiguously (and sometimes not even then).

The distinction between intention-based and form-based approaches to dialogue acts is relevant to consider in connection with the differences between human and automatic annotation. Human annotators are better at recognizing the intentions behind dialogue utterances, since they are experienced in interpreting intentional behaviour and they have more comprehensive context models. Since a general dialogue annotation schema should support human annotation, it should contain concepts with a depth and granularity that matches human understanding of the functions of dialogue utterances. In order to support automatic annotation, on the other hand, the schema should also contain concepts that are suitable for a more surface-oriented form of automatic annotation that relies less on deep semantic knowledge. In order to accommodate both requirements, this standard defines hierarchies of communicative functions and functions with optional qualifiers (see section 10.3), where functions deeper down in a hierarchy or carrying communicative function qualifiers correspond to more detailed intentions or assumptions on the part of the speaker than functions higher in the hierarchy (or being unqualified), and require deeper semantic knowledge for their recognition. An example is the recognition of whether an inform act should be interpreted as a justification, an explanation, or an elaboration of something that was said before.

6.3 Semantic content and communicative function

The idea of distinguishing communicative function and semantic content as the two main components of a dialogue act, inherited from the corresponding distinction in speech act theory (where they are called 'illocutionary force' and 'propositional content'), is that these describe separable and largely independent components of utterance meaning. For example, a propositional question (a.k.a. Yes/No-question) can be about a wide range of semantic content, as (3) illustrates.

(3) a. Do you need to be there before nine o'clock?

b. Did you say Thursday?

c. Did you hear me?

d. Can you wait a minute?

e. Would you like me to say something?

In (3a) the semantic content is related to the underlying task; in b) and c) it is about different types of feedback; in d) about the timing of the interaction; and in e) about who should take the turn. Informs, likewise, can also be about almost any imaginable kind of semantic content. Semantic content therefore cannot play a significant role in the definition of these communicative functions. Other communicative functions can only be used in relation to particular aspects of communication, such as the functions Stalling, Turn Keeping, and Initial Greeting. Dialogue acts with such communicative functions do not have an explicit semantic content; hence for these functions it is also the case that semantic content cannot play a role in their definition.

Some dialogue act annotation schemas employ highly specialized communicative functions such as "accept_date" and "suggest_exclude_location", which mix semantic content type into the communicative function definition. Such functions clearly do not belong in a general-purpose dialogue act annotation schema, and are not part of the set of core communicative functions defined in this standard; they may however be added as 'optional' functions for use in relation to particular domains.

7 Annotation schemas

7.1 Schema structure

A dialogue act annotation schema is a tag set consisting of communicative function tags with their definitions, plus guidelines for how the tags are intended to be used. Existing schemas differ not only in their sets of tags, but more importantly with respect to (1) the underlying approach to dialogue modeling; (2) the definitions of the basic concepts; (3) the coverage of aspects of interaction; and (4) the level of granularity of the tag set.

Dialogue act annotation schemas can be divided into one- and multidimensional ones. One-dimensional schemas have a set of mutually exclusive tags, and are mostly used for coding stretches of dialogue with a single tag. Their tag sets are often quite small (such as the LINLIN schema (Ahrenberg et al., 1995) and the HCRC schema (Carletta et al., 1996)), and have the form of a flat list. The simplicity of these tag sets is often considered to make them more reliable and to take less effort to apply consistently by annotators. Several researchers have noted, however, that one-dimensional annotation schemas also have serious disadvantages (see e.g. e.g. Klein et al., 1998; Larsson, 1998; Popescu-Belis, 2005).

Multidimensional schemas are intended for encoding stretches of dialogue with multiple tags. Such schemas typically have a relatively large tag set. A structuring of such a tag set into clusters of communicative functions has several advantages:

- Clustering semantically related tags improves the transparency of the tag set, as each cluster is concerned with a certain kind of information. The introduction of clusters of tags also makes the coverage of the tag set clearer, since each cluster typically corresponds to a certain class of dialogue phenomena.
- A structured tag set can be searched more systematically and more 'semantically' than an unstructured one, and this can clearly have advantages for dialogue annotation and interpretation.
- The tags within a cluster are typically mutually exclusive (such as 'signal understanding' and 'signal non-understanding'). This supports annotation multidimensional processes since the choice of a particular within a cluster means that the rest of that cluster does not need to be considered any further,

7.2 Multdimensionality and multifunctionality

Participation in a dialogue involves several activities beyond those strictly related to performing the task or activity for which the dialogue is instrumental (such as obtaining certain information, instructing another participant, or reaching an agreement). In natural conversation, among other things dialogue participants constantly "evaluate whether and how they can (and/or wish to) continue, perceive, understand and react to each other's intentions" (Allwood, 1997). They share information about the processing of each other's messages, elicit feedback, and manage the use of time and turn allocation, of contact and attention, and of various other aspects. Communication is thus a complex, multi-faceted activity, and for this reason dialogue utterances are often multifunctional.

Multifunctionality comes in a variety of forms. Allwood (1992) made a distinction between sequential and simultaneous multifunctionality, illustrated by the following example:

(4) A; Yes! Come tomorrow. Go to the church. Bill will be there. OK?
B: The church, OK.

Sequential multifunctionality occurs when an utterance has several parts which each have a different communicative function. (If they have the same function, then it may be wise to apply the 'maximum functional-segment priniciple' (Larsson, 1998) and avoid splitting up the markable into smaller parts.) In the example we see A's utterance containing five functional segments with respectively the communicative functions *feedback giving, request, request, statement,* and *response elicitation*. The occurrence of sequential multifunctionality depends on the way in which a dialogue is segmented (see also section 8), and disappears when sufficiently small segments of dialogue behaviour are considered as markables.

Simultaneous multifunctionality, by contrast, persists even when minimal segments are used as markables. The following example illustrates this:

(5) 1. A: Do you know what date it is?
2. B: Today is the fifteenth.
3. A: Thank you.

In the third turn of (5), A's utterance has the function of thanking, and will mostly be taken to imply that A has understood and accepted the information in B's answer - i.e., as having a positive feedback function. But "Thank you" does not *always* express positive feedback; a speaker who finds himself in a rather unsuccessful dialogue may just want to terminate the interaction in a polite way. The feedback function of the thanking behaviour in example (5) can be inferred along the following lines: A thanks B, so there must be something that A is thankful for, which can only be what B just said; that can only constitute a reason for thankfulness if A considers B's utterance as relevant and useful, which means that A accepted B's utterance as an answer to his question, which implies that A believes that B understood that question. The feedback function in such a case can be viewed as a conversational implicature, as it depends on the application of the cooperative principle (Grice, 1979).

The relation between thanking and giving positive feedback is different from that between a propositional answer ('yes' or 'no') and a confirmation – in this case the relation is one of *entailment*, since every confirmation by its very nature is also an answer. Entailment relations occur in an annotation schema if the definition of one communicative function is a special case of that of another. It is not obvious that such cases should be considered as instances of multifunctionality, however; a speaker who wants to issue a confirmation can hardly have the intention to additionally also give an answer, since the recognition of that intention is already part of the recognition of the confirmation.

There are also cases of multifunctionality where the different functions do not have a logical relation. This is for example the case for turn-initial hesitations, as in the following example:

(6) 1. A: Is that your opinion too, Bert?
2. B: Ehm,.. well,... I guess so.

In the first turn of (6), speaker A asks a question to B and assigns the turn to B (by the combined use of B's name, the intonation, and by looking at B). Speaker B performs a stalling act in order to buy some time for deciding what to say; now the fact that he starts speaking without waiting until he has made up his mind about his answer indicates that he accepts the turn, which was given to him. So the segment *Ehm,.. well,...* has both a stalling function and a turn-accepting function. Note, incidentally, that A's utterance is also multifunctional: it asks a question about B's opinion and it assigns the turn to B.

What implications does the phenomenon of multifunctionality have for the design of annotation schemas? Traum (2000) writes on this subject:

> Given that utterances in dialogue are generally multifunctional, the question arises as to how best to capture this multiplicity of functions in a taxonomy. There are two extremes: one is to separate out each function and code it separately, which requires multiple labels for each utterance, one for each function [...] The other extreme is to combine sets of coherent bundles of dialogue functions into complex labels [...] It is also possible to find taxonomies that take a more intermediate position. (Traum, 2000, p. 23)

Since an annotation schema has two parts, a tag set with definitions and a set of guidelines for their use, a multidimensional approach can be reflected in a schema in two ways: (1) in the tag set, by structuring it into clusters of functions; (2) in the guidelines, instructing annotators how to choose and apply multiple tags from a flat list of tags. If the tag set is fairly extended and does not have any structure, it is next to impossible to formulate good instructions for how to use the tags in multiple tagging, since there is no easy way to refer to groups of tags. Therefore, the recognition of the claim that utterances in dialogue tend to be multifunctional naturally leads to the introduction of dimensions in a dialogue annotation schema.

7.3 Multidimensionality, clustering, and dimensions

The clusters of communicative functions that can be found in existing annotation schemes are typically chosen on the basis of conceptual similarity of the constituent functions. A early version of the DIT schema, for example, has a cluster of 'information-seeking functions' for a range of question types, and a cluster of 'information-providing' functions for various kinds of informs and answers (Bunt, 1989).

The DAMSL schema is organized into 'layers' and 'dimensions'. Four layers are distinguished: Communicative Status, Information Level, and Forward and Backward Communicative Functions (FLF and BLF); the latter two are indeed clusters of communicative functions (the tags in the other layers are concerned with other kinds of information). The FLF cluster is subdivided into five clusters, including (roughly) the classes of commissive and directive functions, well known from speech act theory. The BLF cluster has four subclasses: Agreement, Understanding, Answer, and Information Relation. Core & Allen (1997) refer to these nine subclasses as 'dimensions'. While the DAMSL documentation does not discuss or motivate the choice of layers and dimensions, these are clearly useful for introducing structure in the tag set in a way that can help annotators to make their choices, supported by annotation guidelines which make use of this structure in the process of choosing tags. The DAMSL dimensions within the FLF and BLF clusters make up mutually exlcusive sets of tags, and make DAMSL a nine-dimensional schema.

Popescu-Belis (2005) mentions the following aspects of utterance function that could be relevant for choosing dimensions in a multidimensional schema: (1) the traditional clustering of illocutionary forces in speech act theory into five classes: Representatives, Commissives, Directives, Expressives and Declarations; (2) turn management; (3) adjacency pairs; (4) topical organization in conversation; (5) politeness functions; and (6) rhetorical roles.

Bunt (2005; 2006) suggests to structure a tag set for multidimensional annotation by using a well-founded notion of *dimension*, based on the observation that participation in a dialogue involves a range of communicative activities beyond those strictly related to performing the task that underlies the dialogue. Dialogue participants share information not only about the task that is pursued with the help of the dialogue, but also about the processing of each other's messages, about the allocation of turns, about contact and attention, about the use of time, and about various other aspects of the interaction. They thus perform various types of communicative

Dimension	Dimension-specific comm, functions	Typical expressions
Auto-Feedback	PerceptionNegative	*Huh?*
	EvaluationPositive	*True.*
	OverallPositive	*OK.*
Turn Management	TurnKeeping	final intonational rise
	TurnGrabbing	hold gesture with hand
	TurnGiving	*Yes.*
Time Management	Stalling	slowing down speech; fillers
	Pausing	*Just a minute*
Contact Management	ContactChecking	*Hello?*
Social Obligations Management	Apology	*I'm sorry.*
	Greeting	*Hello.*
		Good morning
	Thanking	*Thanks.*

Table 1: Examples of dimension-specific communicative functions and their expression for some of the dimensions distinguished in the LIRICS project.

activity, such as giving and eliciting feedback, taking turns, stalling for time, establishing contact, and showing attention. Each of these types of activity is concerned with a different kind of information which can can take as their semantic content. In this standard, we use the term 'dimension' to refer to these various types of semantic content or, equivalently, to the types of communicative activity concerned with these types of information. This leads to dimensions such as feedback, turn management, time management, and contact management, besides the dimension formed by the task that motivates the dialogue. Below, in section 9, we will discuss the specific set of dimensions that forms part of this standard, and how they are justified by empirical, theoretical, and practical considerations.

An annotation schema with dimensions that relate directly to activities in communication has the advantage that an annotator or analyst who is specifically interested in certain aspects of communication can choose to use only the corresponding dimensions of the schema. On the other hand, an annotator or analyst who is especially interested in an aspect that is not covered by the schema can add a dimension.

7.4 Dimension-specific and general-purpose functions

A 'dimension' in the context of dialogue act analysis being a type of communicative activity, corresponding to one of the multiple aspects of interacting that dialogue participants monitor and manage, not every cluster of related communicative functions qualifies as a dimension. For example, the cluster that can be formed consisting of the various kinds of information-seeking acts (such as Yes/No-questions, WH-questions, Check Questions, and Menu Questions) does not constitute a dimension, since questions can be concerned with any aspect of communication, such as feedback, task progression, change of topic, or contact. The same is true of the cluster of information-giving acts, which includes statements, warnings, answers, confirmations, and so on, and of commissive and directive acts like Request, Suggest, Instruct, Offer, Promise, and so on. Not only do these clusters of functions fail to qualify as separate dimensions, but the functions in these clusters can be combined with semantic content relating to every dimension. These functions are therefore called *general-purpose communicative functions*. When combined with a semantic content of a certain kind, they form a dialogue act addressing the dimension corresponding to that kind of content. In that sense, the general-purpose functions could be said to belong to every dimension. These functions are discussed further in section 10.1.

In contrast with the general-purpose communicative functions, other functions can only be used to address a specific dimension, such as Turn Keep and Turn Release which are specific for the dimension of Turn Management; and Stalling and Pause for the dimension of Time Management. Table 1 lists some examples of dimension-specific communicative functions in some of the dimensions distinguished in the LIRICS project.

8 Dialogue segmentation

Many studies in the annotation of dialogue with communicative functions have assumed the dialogue to be segmented at the level of turns or utterances. Turns are too coarse-grained for many purposes as the units to

assign communicative functions to, since they often contain smaller parts that have separate communicative functions; example (4) illustrates this. These smaller parts are often called 'utterances'.

Utterances are linguistically defined contiguous stretches of (linguistic) behaviour. Levinson (1983) writes: *'An utterance is the issuance of a sentence, a sentence-analogue, or sentence-fragment, in an actual context.'* Segmenting a dialogue into utterances has the advantage of more fine-grained units being annotated, allowing more precise annotation; however, the notion of an utterance as a smaller unit inside a turn has no clear definition, and the detection of utterance boundaries is a highly nontrivial task. Syntactic and prosodic features are often used as indicators of utterance endings, but are in general not very reliable (see e.g. Shriberg et al., 1998; Stolcke et al., 2000; Nöth et al., 2002). In the case of nonverbal or multimodal communication, the notion of an utterance as a linguistically defined unit is even more problematic.

Apart from the difficulties of automatic utterance boundary detection, from a semantic point of view the stretches of behaviour that are relevant for dialogue act annotation may be discontinuous, may overlap, and may even contain parts from more than one turn. They therefore do not always correspond to 'utterances' in the usual sense of the term, which is why we have introduced the notion of a *funcitonal segment* as a minimal stretch of communicative behaviour that has a communicative function (and possibly more than one).

Example (7) shows that a functional segment may be discontinuous:

(7) A: Do you know what time the next train leaves?
B: The next train is *... let me see...* at 7.48.

The segment *The next train is at 7.48*, which answers the preceding question, is interrupted by *... let me see...* which expresses that the speaker cannot answer immediately, but needs a little time (a 'Stalling' act). As a result, the stretch of communicative behaviour that expresses the answer is discontinuous.

The following example illustrates that dialogue acts may also be expressed by *overlapping* stretches of communicative behaviour:

(8) A: What time is *the first train to the airport on Sunday*?
B: *The first train to the airport on Sunday* is at 06:25.

In this example, B's response as a whole is an answer to A's question, and the repeated question part *The first train to the airport on Sunday* can be viewed as expressing a positive feedback act, displaying B's understanding of A's question. So the answer act and the feedback act are expressed by overlapping functional segments.

Example (9) shows that a dialogue act markable may spread over multiple turns. A asks a question, the answer to which consists of a list of items which B communicates one by one.

(9) A; Could you tell me what departure times there are for flights to Frankfurt on Saturday?
B: Certainly. There's a Lufthansa flight in the morning leaving at 08:15,
A: yes,
B: and a KLM flight at 08:50,
A: yes,

The complications of discontinuity, overlap, and spreading over multiple turns can be handled by taking the multidimensional view on communication that was described in Section 7, where participation in a dialogue involves performing actions in various dimensions, concerned with several aspects of the interaction, often performing these actions simultaneously.

The most natural way to take this into account in dialogue act annotation is to assign communicative functions to all those segments of behaviour that express a dialogue act, allowing these segments to overlap and to be discontinuous and to spread over multiple turns. For example, consider the 3-way segmentation of S's utterance in the following dialogue fragment, where the functional segments in each dimension are indicated in boldface:

U: What time is the first train to the airport on Sunday morning please?
S: The first train to the airport on Sunday morning is .... let me see... at 5:45.
(10) TA **The first train to the airport on Sunday morning is at 5:45** ; ..... let me see...
FB **The first train to the airport on Sunday morning** ; is ..... let me see... at 5:45
Ti **.... let me see...** ; The first train to the airport on Sunday morning is at 5:45

In the example (10) the second turn is segmented is three dimensions: (1) Task/Activity (TA); (2) Feedback (FB); and (3) Time Management (Ti). In the TA dimension, the turn is segmented into the discontinuous functional segment *The first train to the airport on Sunday morning is at 5:45*, which has the function of an answer in this dimension, and the intervening stretch *... let me see...*, which does not have a communicative function in this dimension. In the Time Management dimension the same segmentation applies, but now it's only the segment *... let me see...* which has a communicative function (Stalling). Finally, in the Feedback dimension the turn is segmented into the functional segment *The first train to the airport on Sunday morning*, which provides positive feedback on understanding the preceding question, and the contiguous stretch *is ... let me se... at 5:45*, which is not a functional segment.

9 Core dimensions

'Core dimensions' are those dimensions that are important for almost every task domain and for almost every kind of dialogue, and are therefore distinguished in many annotation schemas. Moreover, for being practically useful, dimensions should be recognizable with acceptable precision by human analysts, and preferably also by automatic systems. (For automatic systems, dimension recognition is important since different dimensions correspond to different types of semantic content, and such differences are often crucial for correctly interpreting a dialogue act with a general-purpose function.)

Petukhova & Bunt (2009a) formulate and test five criteria that a dimension should satisfy for being included in a standard for dialogue act annotation. First of all, only dimensions should be considered which can be distinguished according to empirically observable behaviour in dialogue. This places the notion of a dimension on an empirical basis. Second, each dimension should be theoretically justified, corresponding to well-studied and investigated communicative activities that dialogue participants perform, such as turn taking and feedback. In other words, every dimension should be both theoretically justified and empirically observed.

A third criterion is that each dimension should be recognizable with acceptable precision by human analysts, in particular human annotators, as well as by automatic annotation dialogue understanding and dialogue annotation systems. Recognizability by human and machine annotators is important in order to make the schema useful for the purpose of annotation, as discussed in more detail in the next two sections. Recognizability by dialogue understanding systems is important because the dimension corresponds to the kind of semantic information which is addressed by a dialogue act, and this is relevant for determining which aspects of the system's information state should be updated.

A fourth criterion, which is not so much applicable to the choice of an individual dimension, but rather the choice of a useful *set of dimensions*, is that of the *independence* (or 'orthogonality') of the set. This criterion says that the various dimensions in a multidimensional system can be addressed by dialogue acts independent from addressing other dimensions. More precisely, for every dimension D_i there should be forms of communicative behaviour which express a dialogue act that is concerned with information of the kind that is characteristic for D_i, without also expressing a dialogue act addressing one of the other dimensions. In other words, each dimension is *separately addressable* by dialogue acts. Notice that this is a criterion that can be tested empirically, as has been carried out by Petukhova and Bunt (2009a).

Finally, a fifth consideration specifically applies to the design of a multidimensional *standard* annotation schema, requiring that only dimensions should be included which are commonly present in existing dialogue act annotation schemes. This is a rather practical consideration, making explicit that an annotation standard should capitalize on what is already present in existing good practices.

In sum, the following criteria and considerations can help to make a well-motivated choice of the dimensions in a multidimensional dialogue act annotation schema:

(11) Each dimension in a dialogue act annotation schema should be:

a) theoretically justified, in the sense that it is a well-established and well-studied aspect of communication;
b) empirically observed in the functions of dialogue utterances;
c) addressable independently of the other dimensions.
d) recognizable with acceptable precision by human annotators and by automatic annotation systems;
e) present in a significant number of existing dialogue act annotation schemes.

In their study, Petukhova and Bunt (2009a) survey the literature and analyse the contents of 18 existing annotation schemes in order to verify the requirements (11a) and e) for a range of proposed dimensions. In order to examine the other three requirements, they present the results of annotation experiments and of a range of statistical and machine-learning tests, applied to dialogue corpora of various kind. These tests include empirical data on co-occcurrence relations among dialogue acts and dimensions, tests of independent addressability, measures of semantic relatedness, and data on human and machine recognition of dimensions. The main findings are summarized in Annex F.

Their study confirms that the following nine dimensions fulfil all the requirements (11) and qualify as core dimensions in a dialogue act annotation schema: *"They have been studied extensively, from both theoretical and practical points of view; they are observed in actual dialogues; they are reliably annotated and successfully classified automatically; they are defined in most existing annotation schemes; and they address a certain aspect of communication independently of others."* (Petukhova & Bunt, 2009a, Conclusions). Other dimensions could be considered as optional.

1. Task Dialogues are usually motivated by goals, tasks, or activities which are non-communicative in nature, such as obtaining certain information, solving a problem, improving relationships, acting in a game as a team mates, and so on. Of the nine emerging core dimensions, the one that corresponds to communication about the performance of the task/activity motivating the dialogue, or the task/activity domain is called the Task dimension.

2. Auto-Feedback The term 'feedback' (or more precisely, 'communicative feedback') is most often used to refer to the activity of dialogue participants signalling their attention, understanding, and evaluation of what the speaker says. Feedback is essential aspect of successful communication. Allwood (2000) agues that feedback morphemes and mechanisms, whether they occur as a single utterance or as a part of a large utterance, are probably the most important cohesion device in spoken language. Feedback mechanisms, their linguistics (verbal and non-verbal expressions, durational, temporal and prosodic properties) and related phenomena have been studied extensively, e.g. Duncan & Fiske (1977); Allwood et al. (1993); Clark & Krych (2004). Bales (1951) observed that dialogue participants address several levels of processing of the partner's previous utterances, taking each other into cognitive consideration and showing readiness to communicate, giving attention and receptiveness, recognition, interest and responsiveness to the partner's contribution(-s). Thus, feedback may be reported on various levels. Allwood et al. (1993), Clark (1996) and Bunt (2000) distinguish several feedback levels: attention (in Allwood, 1993 called contact); perception (in Clark (1996) called identification), understanding (in Bunt, 2000 called interpretation); evaluation (in Clark, 1996 called consideration and in Allwood et al., 1993 attitudinal reaction), and execution (Bunt, 2000). The term 'auto-feedback' is used here in order to make a distinction with 'allo-feedback', which is a similar but different different activity.

3. Allo-Feedback Dialogue participants do not only signal whether they are paying attention, to what extent they succeed in hearing and understanding what is being said and meant, and how they evaluate that in the light of their expectations and the context of the interaction (this is what is called 'auto-feedback'), but a participant who has the speaker role also monitors the attention, perception, understanding and evaluation of the addressees, considering such questions as: *Is the addressee paying attention? Does he seem to hear what I'm saying? Does he seem to understand what I mean? Does he accept/appreciate what I'm telling?*, and when appropriate, speakers confirm or correct an addressee's processing, or elicit information about that (feedback elicitation). This communicative activity, where the speaker elicits or volunteers information about the addressee's processing of what the speaker has said, is called *allo-feedback*. Examples of allo-feedback utterances are:

(12) a. Is this clear enough?
b. That's what I meant.

4. Turn Management Turn Management acts are concerned with the allocation of the main speaker role, also called the 'floor'. Allwood (1997) defines turn management as the distribution of the right to occupy the sender role in dialogue. He argues that this is rather normative than a behavioural unit. Accordingly, the decision to take the next turn or to offer the next turn to the partner(-s) depends on the speaker's needs or motivations and beliefs, and on the rights and obligations in a conversational situation.

In dialogues with two or three participants, normally only one participant is speaking at any given moment, while the other participants express their involvement through backchannels (like *m-hm*) nonverbal sounds, and other nonverbal activity. The greater the number of participants in a dialogue, the more one may find different simultaneous speakers, depending very much on the type of interaction. Spontaneous multi-party conversations may be fairly chaotic in this respect; more organized interactive situations, like meetings, tend to have a single participant as the main speaker at any moment.

5. Time Management Fluent speech is relatively rare in spontaneous conversation (Clark & Fox Tree, 2002). An aspect of communication which is concerned with disfluent speech production is time management, where the speaker suspends the dialogue for one of several possible reasons, and resumes it after minor (*stalling*) or prolongned (*pause*) delay. Delays take place at all major levels of planning - from retrieving a word to deciding what to talk about next (Clark & Fox Tree, 2002), in other terms 'micro-' (e.g. word searching problems) and 'macro-structure' delays (uncertainty - see Smith & Clark, 1993), new topic introduction (Swerts & Ostendorf, 1997), or turn-keeping (Stenström, 1990; Goldman-Eisler, 1968)

According to Clark's *theory of performance* (Clark, 1996) speakers in dialogue proceed along two tracks of communication simultaneously: (1) a primary track referring to the task or topic of the dialogue; and (2) a collateral track referring to the communication - to rephrasing, repairs (own communication management), intentions to speak (turn management), timing, delays (time management), and the like. Clark shows that stalling acts are not simply ways of holding the floor but signal imminent delays. He analysed monologues and observes that in monologue there is no issue of holding the floor, yet stalling acts are used just as in dialogues.

6. Discourse Structure Management A dialogue participant may give indications that he is going to close the discussion of a certain topic; or that he wants to concentrate the hearer's attention on a new topic, or that he wants to terminate the dialogue. Dialogue structuring acts are based on the speaker's view of the underlying task, on his plan for continuing the dialogue, and on the assumed need to structure the discourse for his partner.

7. Social Obligations Management Participating in a dialogue is a social activity, where one is supposed to do certain things and not to do others, and to act in accordance with norms and conventions for social behaviour. Dialogue participants have ethical tasks and obligations, and perform dialogue acts to fulfill these. The golden rule of ethics *'Do unto others what you would have them do unto you'* means in communication *'make it possible for others to be rational, motivated agents'* (Allwood, 2004). Bales (1951) pointed out the importance of social obligation acts such as *acts for giving help and reward*.

Bunt (1996) noticed that social obligation acts are often not just 'social', they are also used for improving the transparency of the dialogue. For example, people greet each other to establish and acknowledge their presence, and say good-bye to close the conversation.

8. Own Communication Management A communicative activity in spoken interaction which has been addressed in the literature as well as extensively studied in the context of designing spoken dialogue systems, concerns the speaker's speech production and monitoring.

Allwood et al. (2005), introduced the term 'Own Communication Management (OCM)' for describing the communicative activity of a speaker relating to his management, planning, and execution of his speech production. This activity is indispensible for the description of spoken dialogue, and is illustrated by the performance dialogue acts which are usually called '(self-)repairs', 'restarts', and other speech-editing acts.

9. Partner Communication Management Partner Communication Management (PCM) is concerned with monitoring the partner's speech by the speaker, either providing assistance by completing an utterance that

the partner is struggling to complete (*completion*), or correcting (part of) a partner's utterance, believing that partner made a speaking error (*correct-misspeaking*).

It satisfies all criteria for being a core dimension, although it is not recognized in many existing annotation schemas. This is perhaps related to its relatively low frequency in certain types of dialogue.

10 Core dialogue acts

The various annotation schemas for dialogue acts that have been proposed share a number of communicative functions which are of obvious importance in virtually any type of dialogue. These 'core dialogue acts' include various types of questions, answers, informs, requests, and acknowledgements. Traum and Hinkelman (1992) have used the term 'core dialogue acts' to refer to the types of acts that are most familiar from traditional speech act theory. These are often related to the use of performative verbs (such as *promise, invite,* and *confirm*) and include the commissive and directive act types (*offer, request, propose,...*), the 'reportative' speech acts used for stating facts (*assert, conclude*), and the 'expressive' ones for expressing psychological states (*apologize, thank, congratulate*). In this standard the terms 'core dialogue act' and 'core communicative function' are used to refer to the types of dialogue acts and their communicative functions that are most commonly found in dialogue and that are not specifically related to particular task domains; the data categories specifying names and definitions of these core communicative functions are part of this standard. These include the most common commissive, directive, and reportative acts known from speech act theory and some of the expressive ones, plus a set of other ones which have not been considered much in speech act theory, such as acts for turn taking and time management.

The choice of communicative functions to be included in a dialogue act annotation schema can be based on similar criteria as the choice of core dimensions. First of all, the criterion of *empirical validity* requires that for every communicative function there are linguistic or nonverbal means which are commonly used by speakers to indicate that their behaviour should be understood as having that function. Second, the criterion of *theoretical validity* requires that every communicative function has a precise definition, which clearly distinguishes it from other functions. In particular, the semantic approach taken in this standard requires precise definitions in terms of intended information state updates.

Another empirical requirement for including a communicative functions is that of coverage. For example, the phenomenon that conversational analysts have called 'adjacency sequences' means for an annotation schema that if it includes one element of such a pair, then it should preferably also contain the other. For example, a thanking act is often responded to by a 'downplayer', and an annotation schema which contains a function tag for encoding thankings should preferably also contain a tag for encoding the responding downplayers.

In order to be appropriate as elements in an annotation standard, two additional requirements, again comparable to requirements for dimensions, are (1) that each communicative function should be recognizable with acceptable precision by humans and preferably also by machines, and (2) that they commonly occur in existing annotation schemas.

Finally, where for dimensions we have the requirement of 'independence' or 'orthogonality', saying that every dimension in a multidimensional schema can be addressed independently of the other dimensions, for communicative functions we have the apparently opposite requirement of *functional dependence*, which says that *any two communicative functions that can be used for addressing a given dimension are either mutually exclusive, i.e. if one of them applies then the other one does not; or one is a specialization of the other.* This requirement has the effect that an annotator, when deciding that a functional segment addresses a given dimension D_i, can choose from the set of communicative functions available for D_i that function which, among the functions that might be applicable, is the most specific one for which there is sufficient evidence. For example, in (13) B's utterance clearly provides information to A in response to A's question (addressing the task dimension). This means that an annotator may consider assigning to the functional segment coinciding with B's turn such communicative functions as *Inform, Agreement, Disagreement, Correction, Answer, Confirm,* and *Disconfirm*. Clearly, the functions *Disagreement, Correction* and *Disconfirm* do not apply, since there is nothing adversary in what B says. Of the four remaining possibilities, *Inform* and *Agreement* are not optimally specific, since they miss the fact that B is responding to a question. Of the two remaining responsive functions, *Confirm* is more specific than *Answer*, and since the use of the word *"right"* is a sign of confirmation, expressing not only a positive reply but

also agreement with A's expectation (as opposed to *"Yes"*), the most appropriate function tag is *Confirm*. The functional dependence requirement allows an annotator to select a single function tag from the set of tags available for the dimension under consideration. In the example (13), the chosen tag *Confirm* is mutually exclusive with the tags *Disconfirm, Disagreement*, and *Correction*, while it is more specific than *Inform* and *Answer*.

(13) A: And that's the first flight tomorrow, right?
B: Right.

The requirement of functional dependency among the set of functions available for any given dimension is very useful property of a multidimensional annotation scheme, since the dimensions of the scheme are required to be orthogonal ('independent'). As a result, each stretch of communicative behaviour corresponds to at most as many functional segments as their are dimensions (viz. if it constitutes a functional segment in each dimension), and in each dimension the segment has at most one communicative function. A stretch of communicative behaviour can therefore be at most as many ways multifunctional as there are dimensions in the schema.

All in all, the communicative functions included in the present standard satisfy the following six requirements:

(14) Every communicative function is:
1. empirically observed in features of communicative behaviour in dialogue;
2. theoretically validated as an update operation on information states (i.e. clear semantics);
3. in accordance with the functional dependence principle for the dimensions which it may address;
4. relevant for obtaining a good coverage of the phenomena in the dimensions considered;
5. recognizable by humans and machines;
6. present in a significant number of annotation schemes.

In their survey of approaches to dialogue analysis and existing dialogue act annotation schemas Petukhova & Bunt (2009a) identify the most commonly used communicative functions for the dimensions defined in the previous section - see Annex G, which is the basis for choosing the core communicative functions of this standard.

Note that the definition of communicative functions in this standard should be seen in connection with the inclusion of data categories for these concepts in the ISO Data Category Registry (DCR) (`http://www.isocat.org`). The core dialogue act functions will as such also be registered in the 'Semantics' profile of the registry, which contains certified data categories for semantic annotation. Additional, optional data categories for communicative functions, and extensions sets for different domains or purposes, will in due time also be inserted in the ISO registry, as the result of such concepts being proposed by dialogue researchers and endorsed by the experts in the ISO DCR Board. in view of this connection, this standard includes only small numbers of core communicative functions for the various dimensions:

- for general-purpose functions:
  - 4 information-seeking functions,
  - 6 information-providing functions,
  - 4 commissive functions,
  - 5 directive functions;
- for dimension-specific functions:
  - 2 auto-feedback functions;
  - 3 allo-feedback functions;
  - 2 time management functions;
  - 6 turn management functions;
  - 3 discourse structuring functions;
  - 2 own communication management functions;

- 2 partner communication management functions; and
- 10 social obligation management functions.

The standard refers to the ISO Data Category Registry for more comprehensive sets of communicative functions, including domain-specific sets of functions.

10.1 General-purpose functions

The core general-purpose functions are those domain-independent functions which concern the exchange of information and the discussion of (communicative or other) actions. They can be divided into information-related and action-related. The information-related functions are subdivided further into information-seeking functions, where the speaker aims to obtain certain information from his addressee(s), and information-providing functions, where the speaker wants to make the addressee(s) aware of certain information. The action-related functions fall apart into those where the speaker is committing himself to perform certain actions, possibly under certain conditions (commissive functions), and those where the speaker aims to make the addressee(s) perform certain actions (directive functions).

The choice of core communicative functions within each of these four classes is based on the most frequently occurring functions in existing annotation schemas. The tables in Annex F provide an overview of the occurrence of general-purpose functions in 12 annotation schemas (the twelve richest of the 18 schemes surveyed in Petukhova & Bunt, 2009a).

The most obvious examples of *information-seeking* acts are questions. Many schemas distinguish several types of question, depending on the type of information that the speaker is looking for and on the speaker's expectations regarding the answer that he will get. These distinctions are supported in most natural languages in the distinction of different sentence types and interrogative constructions for different question types. In this standard a distinction is made between *propositional questions*, where the speaker puts forward a proposition of which he wants to know the truth (also known as 'yes/no questions); *check questions*, which are propositional questions where the speaker expects the answer to be positive; *set questions*, where the speaker wants to know which elements of a given set of entities have a certain property (also knnown as 'WH-questions'); and *choice questions* (also known as 'multiple-choice questions', 'menu questions', 'or-questions', or 'alternatives-questions'), where the speaker wants to know which one of a set of listed alternatives is applicable in the situation described by the question.

The most obvious case of an *information-providing* function is the *inform*, which also goes by the names *statement* and *assertion*, and which is the function of a dialogue act where the speaker has the aim to bring certain information to the addressee's attention. More specific cases are the functions *agreement* and *disagreement*, where the speaker believes that the addressee agrees or disagrees, respectively, with the information that is brought to his attention, and the *answer* function, where the speaker provides solicited information. In response to a check question, the speaker may either *confirm* or *disconfirm* the addressee's expectation underlying the check question. Common reactions to informs include *agreement, disagreement*, and *correction*.

Important *commissive* functions are *promise* and *offer*, which have in common that the speaker is prepared to commit himself to performing a certain action; the difference is that in the case of a promise this commitment is unconditional, whereas in the case of an offer the commitment will only occur if the addressee accepts the offer.

The prototypical case of a *directive* function is the *instruct*, also known as *command*, where the speaker unconditionally tells the addressee to do something. As in the case of commissives, there is also a conditional directive, namely the *suggestion*, which brings to the addressee's attention the possibility to perform a certain action, but does not put pressure on him to actually carry out the action. Note that accepting an offer or a suggesting is itself a commissive act,

Further subdivisions and more specific types of each of the functions mentioned here can be made; for instance, the DIT taxonomy distinguishes check questions with a positive and a negative expectation (*posi-check* and *nega-check*. Some taxonomies also distinguish different answer types, such as WH-answer and YN-answer for answers to set questions and propositional questions, respectively. However, these and other more specific

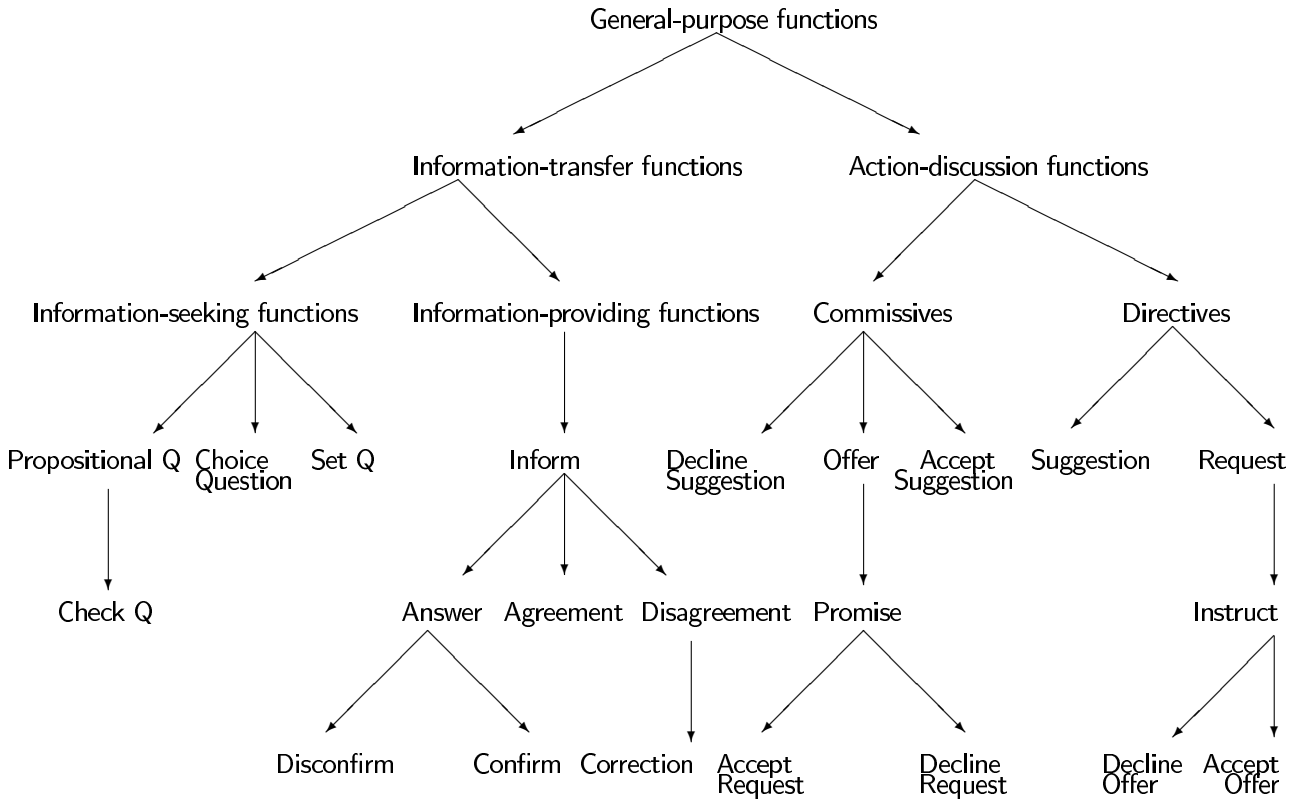


Figure 2 — General-purpose functions

functions may be regarded as optional refinements of the taxonomy of core general-purpose functions distinguished here, which is depicted in Figure 3.

The definitions of the core general-purpose functions are provided in annex D and have in most cases been taken over with minor adaptations from the set of data categories developed in the LIRICS project (see LIRICS deliverable D.4.3).

10.2 Dimension-specific functions

Besides the general-purpose communicative functions, which are available for building a dialogue relating to any of the dimensions that have been defined (by combining the function with a semantic content of the type defined by the dimension), there are also communicative functions which can only be used to address a particular dimension, as already noted above. These 'dimension-specific' functions often do not have a semantic content; for instance, a *Turn Keep* function signals that the current speaker wants to keep the turn; this dialogue act does not require any semantic content. The same is true of all other turn management acts, but also of time management acts, and of contact management acts. Many social obligation management acts like greetings and goodbyes likewise have not semantic content; others, like expressions of thanks or apologies, can have a semantic content, if the speaker indicates what he is thankful for, or what he apologizes for.

The following subsections describe the core communicative functions identified for each of the nine core dimensions.

10.2.1 The Task dimension

Dimension-specific communicative functions for the Task dimension are, by definition, functions that are specific for communication about a particular task domain. The present standard aims to be domain-independent, and therefore does not propose any task-specific functions as core communicative functions. The definition of such functions would be part of the non-core functions that may be added to the core ones, such as functions that are

specific for court room interaction (*sentence, plea, statement..*). The speech act categories called 'declarations' are typically restricted in their use to particular institutionalized situations, with participants that have specific authorizations, such as *baptize, christen, declare war, pronounce husband and wife*, and so on. In view of their domain-dependence, such functions are not included in the standard, but are considered as optional add-ons.

10.2.2 Feedback

Auto- and allo-feedback acts are often performed nonverbally, for instance by nodding, by looking at the speaker (indicating attention), by placing a hand behind an ear ('didn't hear you'), by raising eyebrows, or by frowning.

Feedback-providing acts for auto-feedback as well as for allo-feedback fall apart into positive and negative ones, signaling in the positive case for auto-feedback that the speaker successfully processed a previous utterance and for allo-feedback that the speaker believes the addressee processed the previous utterance successfully, and in the negative case for auto-feedback that the speaker encountered a processing problem and for allo-feedback that the speaker believes the addressee was unsuccessful in processing a previous utterance. In the case of allo-feedback, moreover, feedback elicitation acts express that the speaker is uncertain whether the addressee was successful in processing a previous utterance, and wants to obtain information about that.

Some annotation schemes distinguish various levels of processing to which feedback acts may refer. The GBG-IM schema distinguishes *contact, perception, understanding, evaluation*; the DIT++ schema, inspired by GBG-IM, distinguishes *contact, perception, interpretation, evaluation, execution*. Feedback signals may be specific about the level of processing they address; for instance, a repetition of what was said before in slightly different terms usually relates to the level of understanding, while a verbatim repetition more likely refers to the level of perception (reporting what was heard). Also, head nods may express positive feedback, and it has been observed that the intensity and duration/repetitivity of the nodding indicates the level of processing. Quite often, however, feedback signals are underspecified as to the level of processing they refer to. This is especially true of positive feedback messages. Expressions like *OK* and *Yes* either express positive auto-feedback at some (underspecified) level of processing, such as understanding, or provide positive feedback at the highest level of processing, and thereby by implication at all levels. For annotators it is often impossible to reliably indicate a specific level of processing for feedback messages, therefore the present standard includes the function *positive feedback*, understood as being unspecific in this respect, as well as the function *negative feedback* for the negative case. For expressing feedback at a specific level of processing, notably negative feedback, speakers often use general-purpose functions like Inform and Question, as in *What did you say?*, and *I don't understand what you mean*.

10.2.3 Turn Management

The majority of dialogue act annotation schemas define communicative functions dealing with turn management.

Taking the turn, assigning the turn, accepting the turn, keeping the turn, grabbing the turn (i.e. interrupting) and releasing the turn are the main turn-managment activities. Turn management functions can be divided into *turn-initial* functions, which can only occur at the beginning of a speaker turn and which are concerned with obtaining the speaker role, and *turn-final*functions, which can occur only at the end of a speaker turn end which are concerned with the end of having the speaker role.

DIT++ and **LIRICS** define 6 communicative functions in this dimension: *turn accepting, grabbing* and *taking* as turn-initial functions, and *turn keeping, assigning* and *releasing* as turn-final (or closing) functions.

The turn management functions in this standard are defined as the activities that a dialogue participant undertakes *explicitly and specifically* for obtaining, maintaining, or giving up the speaker role. Just starting to speak is not considered as a turn management act, and neither is ceasing to speak considered as giving up the speaker role, or continuing to speak as trying to keep the speaker role. See on this topic also the annotation guidelines in Annex A.

10.2.4 Time Management

Stalling for time is a widespread phenomenon in spoken interaction. Speakers often need a bit of time to decide on the content and form of what they are going to say, and do not always succeed in immediately finding the

right words. This is indicated by slowing down and using fillers like *ehm, let me see, you know, well,* and so on. Simply being silent in such cases would be felt as awkward. Using fillers and slowing down can be used only when the speaker needs just a few seconds, not for, say, several minutes. A speaker who needs more time then just a few seconds, for instance because he needs to find his agenda, or to look something up, or to make a calculation, or because he is interrupted by something urgent, should do something else then stalling. This is where expressions like *just a minute, hold on, momentito, un instant, veuillez patienter* are used. They signal that the speaker is briefly suspending his contribution to the dialogue but intends to resume soon. We call this function Pausing.

Interactive computer systems also use Pausing acts, often by means of the same expressions as used in natural conversation, to indicate that the user has to wait a little, that the system is busy and needs some time to complete its processing. Nonverbal means are also used for this purpose, such as an hour glass icon or a bar which gradually fills up.

10.2.5 Discourse Structuring

Dialogue participants may structure the interaction explicitly by opening and closing the dialogue, by introducing, changing, or closing a topic, by indicating what they intend to do next, or what they would like another participant to do next. When the discourse structure is addressed explicitly by dialogue acts, most often general-purpose functions are used.

Several existing annotations schemas include a Closing act for closing a dialogue. Most commonly, however, a dialogue is closed by using farewell greetings. Such greetings, which belong to the dimension of Social Obligations Management, can thus be seen as having the additional function of closing the dialogue. In human dialogue, no separate dialogue acts have been found that have as their only function to close a dialogue. Computer dialogue systems also commonly use farewell greetings (`bye`) to indicate termination of the interaction.

10.2.6 Own and partner communication management

Own communication management, occurring when a speaking edits his own speech while contributing to the dialogue, most commonly takes the form of self-corrections (also called 'repairs') and retractions. The most common forms of Partner Communication Management are the correction of speaking errors and the completion of an utterance which the partner is struggling to complete.

Speakers continuously monitor the utterances that they are producing or preparing to produce (Clark, 2004), and when problems or mistakes are discovered, they stop the flow of speech and signal that there is some trouble, and that a repair is following (*error signalling*). A speaker may make mistakes in verbal fluency, e.g. stuttering, or mispronouncing words and may wish to reformulate a part of his utterance or to start from the beginning of the phrase within the same turn (*retraction*). Retractions frequently occur at the beginning of an utterance and within other hesitations and phrasal breaks. Sometimes a speaker just repeats a phrase or part of it without reformulations within the same turn (*restart* or *refresh*), and this may have several reasons. When the speaker has produced a (partial) result, recognises that he made an error, and corrects it within the same turn, one speaks of *self-correction*.

10.2.7 Social obligations management

Of the numerous dialogue acts that can be performed for social functions, some are found very frequently in all kinds of dialogue. These include greetings and valedictions, which are often used to open and close a dialogue, respectively. Introducing oneself is also used in many interactive situations in order to get a dialogue going. Apologies are common when a dialogue participant has misunderstood one of the other participants, or is unable to fulfill a request or answer a question; thanking occurs frequently in those situations where one participant is performing a service or providing help, and is also often used to initiate the closing of a dialogue. All these dialogue acts tend to come in initiative-response pairs, such as an initial and a response greeting, an apology and its acceptance, and a thanks and a 'downplayer' (*de nada; pas de quoi*). These $2 \times 5 = 10$ highly domain-independent functions are proposed as core communicative functions (see annex D).

Additional functions such as Congratulation, Condolence, and Compliment can optionally be added to this category.

10.3 Function qualifiers

A limitation of virtually every dialogue act taxonomy is that it fails to capture certain subtleties in the performance of communicative actions relating to such phenomena as modality, conditionality, partiality, and accompanying emotions and attitudes. For instance, the set of action-discussion functions considered so far distinguishes only two possible responses to an offer: accepting it and refusing it. However, people may respond to an offer in less clear-cut ways, accepting an offer conditionally, or partly, as in the following examples:

(15) 1. A: Would you like to have some coffee and cakes?
2. B: Only coffee please.
3. B: Coffee would be nice, but do we have time for that?

Response 2 can be characterized as a *partial acceptance* of the offer in (15), and response 3 as a *modal acceptance*. Besides offers, also suggestions and requests can be accepted partly, conditionally, and with certain modalities. Answers to questions may exhibit similar qualifications:

(16) 1. A: Do you know who'll be coming tonight?
2. B: Peter, Alice, and Bert will come for sure.
3. B: I heard that Tom and Anne will not come.
4. B: I have a hunch that Mary will not come.

The responses 2, 3 and 4 in (16) all constitute partial (rather than exhaustive) answers; response 4 is in addition a modal (viz. *uncertain*) answer. Other information-providing dialogue acts besides answers may also be uncertain, since the agent who performs these acts may be uncertain about the correctness of the information that he provides.

Many dialogue acts can also be performed with the additional expression of the sender's emotional stance with respect to the semantic content of the act or his attitude toward the addressee, for instance:

(17) a. A; Can I offer you a cup of coffee?
B: Ah, lovely!

b. A: The first flight tomorrow morning is a seven-thirty.
B: Perfect!

c. A: What about a fresh cup of coffee?
B: Oh, you're wonderful!

In (17a), B's acceptance of A's offer carries the additional attitudinal information that B would very happy to have a cup of coffee; in (17b) B's positive feedback also carries information about B being very satisfied with the information he got from A; and in (17)c) B's acceptance of A's offer additionally expresses B's feelings toward A.

One way to enrich a dialogue act annotation schema in order to be able to represent such phenomena is to add a number of communicative functions, such as 'Uncertain Answer', 'Conditional Accept Request', 'Partial Accept Offer', 'Happily Accept Offer', and so on. This hardly seems adequate, however, in view of the fact that an answer can qualified in multiple ways, for instance being both uncertain and partial; an offer can be accepted both partly and with some emotion or attitude expressed, and so on. An alternative is to introduce a number of *qualifiers* that may be attached to a communicative function.

A study of these phenomena by Petukhova & Bunt (2009c) indicates that, for dealing with the most frequent cases, the following qualifiers indicated in Table 2 are sufficient. Each of the qualifiers concerns an aspect of qualification that may be relevant for certain categories of communicative functions, indicated in the rightmost column in the table. These qualifiers cover the following phenomena:

- modality: the qualifier 'uncertain' can be used with information-providing functions, in order to indicate that the speaker us uncertain about the correctness of the information that he provides, as exemplified in *It might get cold tonight* or in (16.4).
- conditionality: the qualifier 'conditional' can be used with action-discussion functions, which have in common that the speaker assumes that the action under discussion can be performed by the participant whose action is discussed (the speaker, in the case of commissives; the addressee, in the case of directives). The 'conditional' qualifier indicates that this assumption is dropped.
- partiality: the qualifier 'conditional' can be used to indicate that the semantic content to which the communicative function is applied is a proper part of the semantic content of the dialogue act with which there is a functional dependence relation.
- emotional/attitudinal qualification: these qualifiers indicate that the speaker has a certain emotional attitude to the semantic content that the communicative function is applied to, or towards the addressee.

aspect of qualification	qualifier values	communicative function category
modality	uncertain, certain	information-providing functions
conditionality	conditional, unconditional	action-discussion functions
partiality	partial, complete	responsive general-purpose functions; feedback functions
attitudinal type	happy, unhappy, surprised, ...	any communicative function

Table 2. Aspects of qualification, qualifiers, and relevant function categories.

The aspects *modality, conditionality* and *partiality* have simple binary values. For *emotional and attitudinal qualification* it seems more appropriate to have an open class of values, which are not part of the standard.

Note that some communicative functions are inherently conditional, for instance, a Request to do X can be seen as a conditional Instruct to do X (the condition being that the addressee agrees to do X), and an Offer to do X can be viewed as a conditional Promise to do X (the condition being that the addressee accepts the offer).

11 DiAML: Dialogue Act Markup Language

In the ISO Linguistic Annotation Framework (Ide et al., 2003), a distinction is made between *annotation* and *representation*. The term 'annotation' is used to refer to the process of adding information to segments of language data, or to refer to that information itself. This notion is independent of the format in which this information is represented. The term 'representation' is used to refer to the format in which an annotation is rendered, for instance in XML, independent of its content. In passing, Ide et al. note that the identification of a segment in the language data is in fact also a part of an annotation task. According to the Linguistic Annotation Framework, annotations rather than representations are the proper level of standardization. This standard defines a dialogue act markup language at two levels: the abstract level of annotation, and the concrete level of representation. This language is called DiAML: Dialogue Act Markup Language.

The distinction between annotations and representations is reflected in the definition of DiAML given below, where an *abstract syntax* is defined independent of a *concrete syntax*. The abstract syntax specifies the elements making up the information in an annotation and how these elements may be combined; these combinations are defined as set-theoretical structures. These structures can be represented concretely in many different ways. In line with other ISO TC 37/SC 4 standards, an XML-based concrete syntax is defined for representing DiAML annotations. Any other representation that is a faithful rendering of the abstract syntax of DiAML can readily be converted into this representation and vice versa. DiAML has a formal semantics associated with its *abstract* syntax, which explains why all concrete representations of DiAML annotations are semantically equivalent.

The Linguistic Annotation Framework recommends the use of stand-off annotation, where the annotation representations are kept separate from the primary data. Stand-off annotations refer to specific locations in the primary data by addressing byte offsets, linguistic elements such as words, or times associated with recorded data, to which the annotation applies. Compared to in-line annotation, stand-off annotation has the advantages

of respecting the integrity of the primary data and of allowing multiple annotations to be layered over given primary data. For dialogue act annotation, in-line annotation would moreover be fundamentally inadequate since functional segments can be discontinuous; moreover, the multiple segmentation that is used in multidimensional annotation would require alternative copies to be annotated for each dimension. Since semantic annotation typically occur at a relatively high level in a layered annotation structure, they do not necessarily refer directly to segments in the primary data, but rather to structures in other annotation layers.

11.1 DiAML abstract syntax

The abstract syntax of DiAML defines certain set-theoretical structures ("DiAML annotation structures") which contain all and exactly those elements that constitute the annotation of functional segments in dialogue with communicative functions according to the metamodel show in Figure 1. The definition of the abstract syntax of DiAML consists of two parts: (a) a specification of the elements from which annotation structures are built up, called a 'conceptual inventory', and (b) a set of rules which describe the possible ways of combining these elements ('annotaton construction rules').

Definition of DiaML, abstract syntax.

a) Conceptual inventory:

- a finite set $Parts = \{P_1, P_2, ..., P_k\}$of elements called 'dialogue participants';
- a finite set $Dim = \{D_1, D_2, ..., D_N\}$ of elements called 'dimensions';
- a finite set of sets $DSF = \{DSF_1, DSF_2, ..., DSF_N\}$, where each element DSF_i is a finite set $DSF_i = \{F_{i1}, F_{i2}, ..., F_{in_k}\}$ of elements called 'dimension-specific communicative functions';
- a finite set $GPF = \{F_{01}, F_{02}, ..., F_{0n}\}$ of elements called 'general-purpose communicative functions';
- a finite set QA = $\{A_1 ... A_k\}$ of elements called 'qualification aspects', and a set of finite sets $Q_1, ..Q_k$ of elements called 'qualifiers';

b) Annotation construction rules:

- an annotation structure is a pair $< \sigma, \delta >$ where σ is functional segment and δ is an act structure, or pair $< \sigma, \Delta >$ where σ is functional segment and Δ is a set of act structures, or a set of annotation structures;
- an act structure is one of the following:
  1) a quadruple $< S, A, d, f >$ where $S \in Part$ (the participant who is the sender/speaker of the dialogue act); $A \subset Parts$ (the set of addressees of the dialogue act); d is a dimension ($d \in Dim$); and f is a communicative function;
  2) a quintuple $< S, A, d, f, \delta' >$ with S, A, d, and f as before, and where δ' is an act structure;
  3) a quintuple $< S, A, d, f, \sigma' >$ with S, A, d, and f as before, and where σ' is a functional segment.
- a communicative function is an element of the set of (core) communicative functions, i.e. $f \in DSF \cup GPF$ (i.e., f is a dimension-specific or a general-purpose communicative function); or a pair $< f, q >$ where $f \in DSF \cup GPF$ and q is a qualifier structure;
- a qualifier structure $q$is a list of qualifiers in which no qualification aspect occurs more than once. Formally: $q \in Q_1 \cup Q_1 \times Q_2 \cup Q_1 \times Q_2 \times Q_3 \cup Q_1 \times Q_2 \times Q_3 \times Q_4$.

11.2 DiAML: concrete XML-based syntax

An XML-based concrete syntax for representing the information structures defined by the DiAML abstract syntax consists of (1) a vocabulary, specifying names of XML tags, attributes, and values for the various ingredients in the conceptual inventory, and (2) a specification of how to represent DiAML annotation structures.

1) Vocabulary:
   - XML attributes and values to represent the elements of the conceptual inventory (dimensions, communicative functions, function qualifiers, functional dependence relations, feedback dependence relations, speaker and addressees):
     * names to identify dialogue participants;

  - dimension names): `task, autoFeedback, alloFeedback, turnManagement, timeManagement, discourseStructuring, ownCommunicationManagement, partnerCommunicationManagement, socialObligationsManagement`
  - communicative function names (as attribute values): `inform, agreement, disagreement, correction, propositionalQuestion, setQuestion, checkQuestion, choiceQuestion, answer, confirm, disconfirm, inform, agreement, disagreement, correction, offer, promise, acceptSuggest, declineSuggest, suggest, request, instruct, acceptOffer, declineOffer, autoPositive, autoNegative, feedbackElicitation, alloPositive, alloNegative, turnTake, turnGrab, turnAccept, turnRelease, turnAssign, turnKeep, stalling, pausing, topicShift, completion, correctMisspeaking, selfCorrection, signalSpeakingError, opening, interactionStructuring, initialGreeting, returnGreeting, apology, acceptApology, initialSelfIntroduction, returnSelfIntroduction, thanking, acceptThanking, initialGoodbye, returnGoodbyeValediction`;
  - qualifier attribute and value names:
    - attribute `modality`; values `certain, uncertain`
    - attribute `conditionality`; values `conditional, unconditional`
    - attribute `partality`; values `partial, exhaustive`
    - attribute `emoAttitunality`; values `conditional, unconditional`

2) Representation of DiAML-structures:
The formal specification of the DiAML concrete syntax using XML, as well as the specification of a feature structure representation in XML following the TEI-ISO standard for annotation texts is given in Annex C. Here we just give a somewhat simplified example of the former kind of representation.

In this example a stretch of dialogue, corresponding to the second turn in the dialogue fragment (18) is contributed by participant P2 in the sender role, directed to participant P1 as the addressee, in response to a turn by P1. P2's utterance is segmented into two overlapping functional segments: one in the Auto-Feedback dimensions, with positive value, and one in the Task dimension, with value 'answer' qualified as 'uncertain'.

(18)
P1: *Do you know what time the next train to Utrecht leaves?*
P2: *The next train to Utrecht leaves at 8:32.*
AuFB The next train to Utrecht [positiveAutoFeedback]
TA The next train to Utrecht leaves I think at 8:32. [answer, uncertain]

Dialogue act annotations may be attached to primary dialogue data in a variety of ways. They may be attached directly to stretches of speech, defined by temporal begin- and end points, but often they will be attached to structures at lower levels of analysis and annotation, such as the output of a tokenizer. Here we will assume that the relevant functional segments are identified at another level of XML representation, for instance in the way in which information can be attached to digital documents according to TEI-ISO standard (ISO 24610-1; see Annex C for more details. For the example, we assume that P1's utterance is identified as the functional segment 'fs1', and the two functional segments in P2's turn as 'fs2' (in the Auto-Feedback dimension) and 'fs3' (in the Task dimension).

Another representational issue concerns the identification of the dialogue participants. Again, following ISO practices, we will assume that the dialogue fragment considered her forms part of a digital document in which the metadata contain the relevant information that identifies the participants.

With these assumptions, the DiAML-XML representation of the dialogue act information about (18) is as follows:

```
<diaml xmlns:"http://www.iso.org/diaml/">
      <dialogueAct xml:id="da1" sender="#p1" addressee="#p2"
                target="#fs1" communicativeFunction="setQuestion"
                dimension="task" conditionality="conditional"/>
      <dialogueAct xml:id="da2" sender="#p2" addressee="#p1"
                target="#fs2" communicativeFunction="autoPositive"
                dimension="autoFeedback" feedbackDependenceTo="fs1"/>
      <dialogueAct xml:id="da3" sender="#p2" addressee="#p1"
```

```
                    target="#fs2" communicativeFunction="answer"
                    dimension="task" functionalDependenceTo="da1"/>
</diaml>
```

11.3 DiAML semantics

A fundamental requirement on semantic annotation is that semantic markups should have a well-defined semantics (Bunt & Romary, 2002; Bunt, 2007b). The DiAML language has a formal model-theoretic semantics associated with its abstract syntax, which runs as follows. Below we will make use of the follow functional concept. For a given set of functions F, which all have the domain D and the range R (i.e. F is a set of functions from D to R, which is also commonly designated by $f \in R^D$), the notation F^* is used to designate the set of functions defined as: F^* = $\{f \in R^D | \exists f_i, ..., f_m \in F$, such that $f = \lambda x.f_i(x) \cup ... \cup f_m(x)\}$. (In particular, given a set F of update functions, this defines the set those update functions which are unions of update functions available in F.)

Definition. A model for DiAML is a pair $M = < D, F >$, where
D is the model structure, which is a triple $< IS, P_u, F_Q >$, where
- IS is a set of information states;
- P_u is a set of function schemas defining functions from IS to IS (update schemas);
- F_Q is a set of functions from P_u^* to P_u^* (from update schemas to update schemas).

F is the interpretation function of the model, which assigns to every element of $DSF \cup GPF$ an update schema which is a member of P_u^*.
In other words, every communicative function is interpreted as a way of updating an information state. An update schema is a function which, given a sender, an addressee, and a dimension, specifies how a given semantic content should be used to update the addressee's information state.

The DiAML semantic specification finally describes how the meaning of an annotation structure is computed recursively from that of its components. This specification runs as explained informally as follows.

Note, first, that the 'update schemas' in P_u in the model are parameterized schematic update operations, such as the following:

(19) P_{u1} : update Y's information state by extending it with the information that participant X wants participant Y to know that p
P_{u2}: update Y"s information state by extending it with the information that participant X believes that p

Using these two schemas, an Inform act where speaker S informs addressee A that AT(KL476, 19:15) could then be interpreted as A's information state being extended with two beliefs, applying the schemas)19), instantiated with X = S; Y = YA; and p = AT(KL476, 19:15). The effect of those updates is that (1) A now believes that S wants A to know that AT(KL476, 19:15), and (2) A now believes that S believes that AT(KL476, 19:15).

In this example we have interpreted the Inform function as the update schema $\lambda X.\lambda Y.\lambda z.P_{u1}(X,Y,z) \cup P_{u2}(X,Y,z)$. Applied to the participants S and A, the update function $\lambda z.P_{u1}(S,A,z) \cup P_{u2}(S,A,z)$ results, which can be applied to a semantic content in order to obtain a description of how A's information state is updated. This formalizes the definition of communicative functions given in section **??** as a specification of how a semantic content should be used to update an addressee's information state. (Note that this example is highly simplified as an account of the Inform function; for a more realistic treatment see Bunt et al., 2007.)

This example illustrates the backbone of the DiAML semantics. To this we have to add the interpretation of dimensions, functional and feedback dependence relations, and qualifiers.

In DiAML annotation structures there is no full representation of semantic content, only an indication of the *type* of semantic content by means of a dimension. From a semantic point of view, the importance of a dimension is that it provides information about which part of an information state is updated. Since information states are complex, structured representations where different kinds of information are represented in different parts of the structure (see e.g. Bunt, 2000; Cooper, 2004; Poesio & Traum, 1998) , the indication of where the update is situated is useful for making the update-based approach realistically feasible, especially when information of a rich variety is considered as is the case in this standard. The update schemas in P_u then have an additional parameter indicating the relevant part of an information state. Formally, addition of dimensions to the DiAML semantics means that interpretations are assigned to

quadruples $< S, A, d, f >$ consisting of a speaker, an addressee, a dimension, and a communicative function. Using V to designate the interpretation assignment, we would, for example, get for the interpretation of an Inform function applied in the Turn Management dimension:

(20) $V(<X, Y,$ TurnManagement, Inform$>)$ = $\lambda X.\lambda Y.\lambda z.P_{u1}(X, Y, tu, z) \cup P_{u2}(S, A, , tu, z)$

which says that Y's information about the allocation of the speaker role should be updated with the semantic content of the Inform act (e.g. when X informs Y that he is willing to take the turn.

The semantics of functional and feedback dependence relations adds to (20) that in the addressee's information state a link should be created between the dialogue act of which the communicative function is interpreted and the dialogue act to which it responds (in the case of a functional dependence relation) or to the functional segment to which it provides or elicits feedback (in the case of a feedback dependence relation). The addition of such links in an information state requires such a state to include a dialogue history, which represents dialogue acts that occurred earlier in the dialogue. The formal details of this are not spelled out here.

Finally, in the above outline of the semantics of DiAML annotation structures we have assumed the communicative function to be unqualified. Of the qualifiers introduced in section (10.3), the default values of the modality, conditionality, and partiality attributes ('certain', 'unconditional', and 'complete', respectively) have the semantic effect of leaving the communicative function unqualified. The other values have the following semantic effects:

1) 'uncertain': this qualifier is only applicable to information-providing functions, for which the update semantics includes an update schema describing that the speaker believes the semantic content to which the function is applied to be true. In the 'uncertain' case, this update schema instead says that the speaker has a *weak belief* concerning this.
2) 'conditional': this qualifier is applicable to action-discussion functions, for which the update semantics contains an update schema describing that the participant whose action is discussed (the speaker, in the case of commissives; the addressee, in the case of directives) is able to perform the action under discussion. In the 'conditional' case, this update schema is not applied.
3) 'partial': in this case the semantic content to which the communicative function is applied is a proper part of the semantic content of the dialogue act with which there is a functional dependence relation. This fact is added to the addressee's information state as an additional update.

Qualifying a communicative function with an emotional/attitudinal qualifier is semantically simply additive to the update semantics of the function without such a qualifier; for instance, if the speaker expresses being very pleased about the semantic content to which the function is applied, then this is just an additional fact to be added to the addressee's information state.

Note that the effects of the four types of qualifiers distinguished in this standard are semantically independent, so the semantics of a qualifier structure (in the sense of the abstract syntax of DiAML) containing several qualifiers is a simple composition of that of each of the qualifiers separately.

12 Principles for schema extension and restriction

cannot be expected to be ideal for every kind of dialogue analysis, for every task domain, for every kind of dialogue, and for every annotation purpose. The general principles underlying the design of the schema and the DiAML annotation language should however also be useful for accomodating extensions, modifications, or restrictions of the schema and the annotation language, as the need arises for particular applications. In this section we summarize the main design principles and formulate guidelines for schema extension and restriction.

12.1 Main design principles

The main principles underlying the annotation schema and the DiAML language can be summarized as follows:

- Dialogue behaviour is viewed as multifunctional, i.e. each stretch of communicatively meaningful behaviour may have more than one communicative function.

- Communicative functions are most accurately assigned to *functional segments*, minimal stretches of communicative behaviour that have a communicative function (one or more). Functional segments may be discontinuous, overlapping, and spreading over multiple turns.

  Segmenting a dialogue into functional segments is most accurately done in a 'multidimensional' fashion, identifying in each dimension the stretches of communicative behaviour that count as functional segments in the sense of having a communicative function in that dimension.
- Communicative functions are defined semantically in terms of how they use a semantic content to change the information state of a dialogue participant who understands the corresponding functional segment (understands: as intended by the speaker).
- Dimensions are required to be (1) theoretically justified; (2) empirically observed; (3) recognizable with acceptable precision by human annotators and by automatic annotation systems; (4) addressable independently from other dimensions ('orthogonal').
- Communicative functions are required to be: (1) empirically observed; (2) theoretically validated; (3) satisfy the functional dependence principle; (4) relevant for obtaining a good coverage of the phenomena in a given dimension; (5) recognizable with acceptable precision by human annotators and by automatic annotation systems.
- The *functional dependence principle*, mentioned in the previous item as a requirement on the set of communicative functions that can be used for addressing a certain dimension (i.e. the functions specific for this dimension and the general-purpose functions), stipulates that any two such functions should either be mutually exclusive alternatives or one should be a specialization of the other.
- Adherence to the requirement of orthogonality for dimensions and that of functional dependence for communicative functions has the effect that a functional segment should never be annotated with more than one function per dimension, when annotators follow the strategy of assigning the most specific communicative functions for which there is sufficient evidence, and as having maximally as many functions as there are dimensions in the annotation schema.

12.2 Guidelines for schema extension

- Dimensions were noted to satisfy certain requirements which are summarized above plus the requirement of occurring in a significant number of existing dialogue act annotation schemas. For specific purposes or domains, new dimensions may be added for which this additional requirement does not hold. Neither does the requirement of theoretical justification need to be observed, since the purpose may be to investigate dialogue phenomena which have not been well studied yet.
- New communicative functions may be added to the set of core functions, provided that they satisfy the requirements summarized above, except possibly the requirement of theoretical validity; in particular, the functional dependence principle should be observed. An additional requirement for including a communicative function, mentioned in section 10, is that the function should occur in a significant number of annotation schemes. As in the case of adding dimensions, this requirement may be dropped when adding functions for a specific purpose or a particular domain.
- Domain-specific communicative functions may freely be added, provided that they meet the general
- New communicative functions may be freely introduced which are more specific than a function already present in the schema.
- Communicative function qualifier values may freely be introduced for the attribute that takes emotional stance and attitude into account.
- Additional communicative function qualifier attributes (and their values) may be introduced provided that they capture information which relates in a well-defined way to what is captured by those attributes already present.

12.3 Guidelines for schema restriction

- Communicative functions may be freely left out for which there is a less specific function present in the schema.
- It is not recommended to leave out a communicative function for which the schema contains more specific functions while maintaining the more specific functions.
- A dimension and the corresponding set of dimension-specific communicative functions may be freely left out.

- Communicative functions may be left out which are considered irrelevant for a particular purpose, if that does not have undesirable limiting effects on the desired coverage of dialogue phenomena.
- The use of communicative function qualifiers is optional; these may be freely left out, in which case the results of analysis or annotation are equivalent to using the qualifiers with their default values.

Annex A
(normative)
Annotation guidelines
(First, preliminary draft)

A.1 Overview

This annex consists of three parts. First, a brief discussion is given of the purposes of applying this standard and of some exceptional situations that annotators may encounter. Second, some very brief general'tips for annotators' are formulated. These are copied with minor adjustments from he DIT^{++} annotation guidelines – see `http://dit.uvt.nl`. Third, a more detailed list of explanations is given for how to use each of the DiAML tags, using the representation format detailed in Annex C.

A.2 Annotation purposes and unusual annotation situations

This standard is primarily intended for use by human annotators and by automatic annotation systems. It has been tested for being useful for both these purposes.

If the purpose of an annotation effort is to achieve the most accurate annotations, then the annotators involved should use all the sources of information that are available. For a multimodal dialogue, where speech is used in combination with nonverbal behaviour that the participants may see, this means that not only the recorded speech should be available to annotators, but also a video recording of the nonverbal behaviour. Similarly, in the case of a dialogue over the telephone, annotators should not only have the transcribed speech at their disposal but also the original sound recording, for being able to judge the use of prosody, of pauses, and of nonverbal vocal sounds. One important source of information for annotators when deciding on the identification or annotation of a given functional segment may be the recording of how the dialogue continued after the segment under zonsideration. Therefore, if the purpose is to obtain the most accurate possible annotation, annotators should be allowed to use look-ahead.

One of the purposes of constructing annotated dialogue corpora may be to train automatic dialogue management systems. Such systems should be able to recognize the communicative functions of the user's dialogue behaviour. As an on-line participant in a dialogue, a dialogue system obviously cannot look ahead; therefore, it may be useful to train such a system on a corpus of which the annotations have been constructed without making use of look-ahead.

Dialogue annotators may encounter interactive situations where the sender and addressee roles are less straightforward then in spontaneous conversations. For example, in a televised interview the person who is interviewed is talking to the interviewer as the addressee, while in fact he may be trying to get something across to the audience in the TV studio and to the viewers at home. Or a member of parliament may be formally addressing the speaker of the house, while in fact he is understood to be conducting a debate with a cabinet minister. It is assumed that such unusual situations can be clarified in the metadata of the electronic resource describing the recorded dialogue, and that annotators dealing with such situations will receive instructions for how to deal with these roles as required by the situation.

A.3 Tips for annotators

Dialogue act annotation is about indicating the kind of intention that the speaker had; what kind of thing was he trying to achieve? When participating in a dialogue, this is what the addressees and possible other participants are trying to establish. The first and most important two guidelines for dialogue act annotation follow from this.

1) **"Do as an addressee would do!"**
   When assigning annotation tags to a dialogue utterance (a 'functional segment', more precisely), put yourself in the position of the participant(s) at whom the utterance was addressed, and imagine that you try to understand what the speaker is trying to achieve. Why does (s)he say what (s)he says? What are the purposes of the utterance? What assumptions does the speaker express about the addressee? Answering such questions should guide you in deciding which annotation tags to assign, regardless of how exactly the speaker has expressed himself. Use all the information that you

could have if you were the actual addressee, and like the addressee, try to interpret the speaker's communicative behaviour as best as you can.

2) **"Think functionally, not formally!"**
The linguistic form of an utterance often provides vital clues for choosing an annotation tag, but such clues may also be misleading; in making your choices you should of course use the linguistic clues to your advantage, but don't let them fool you - the true question is not what the speaker says but what he means.

For example, Set Questions are questions where the speaker wants to know which elements of a certain domain have a certain property. In English, such questions often contain a word beginning with "wh", such as *which* as in *Which books did you read on your holidays?* or *where* in *Where do your parents live?* But in other languages this is not the case; moreover, even in English not all sentences of this form express a Set Question: *Why don't you go ahead* is for instance typically a suggestion rather than a question.

Similarly, Propositional Questions are questions where the speaker wants to know whether a certain statement is true or false. Such sentences typically have the form of an interrogative statement, such as *Is The Hague the capital of the Netherlands?* or *Do you like peanut butter?* But not all sentences of this form express a Propositional Question; for example, *Do you know what time it is?* is most often used as an indirect way of asking a question (*What time is it?*). Similarly, *Would you like some coffee?* is most likely an offer, rather than a question, and *Shall we go?* a suggestion.

3) **"Be specific!"**
Among the communicative functions that you can choose from, there are differences in specificity. For instance, a Check Question is more specific than a Propositional Question, in that it additionally carries the expectation that the answer will be positive. Similarly, a Confirm act is more specific than an Answer, in that it carries the additional speaker assumption that the addressee expects the answer to be positive.

In general, try to be as specific as you can. But if you're in serious doubt about specific functions that you could choose between, then simply use a less specific function tag that subsumes the more specific tags.

4) **"Code indirect speech acts just like direct ones."**
Standard speech act theory regards indirect speech acts, such as indirect requests, as just an indirect *form* of the same illocutionary acts. By contrast, this standard incorporates the view that indirect forms signal subtly different packages of beliefs and intentions than direct ones. For example, the direct requestion *Tell me what time it is please* carries the assumption that the addressee is likely to know what time it is, whereas the indirect request *Do you know what time it is?* does not carry that assumption (it does at least not *express* that assumption; in fact it questions it), and is best interpreted as *Please tell me what time it is, if you know.*

This example shows that an indirectly formulated request has a conditional character: the speaker is expressing a request under the condition that the addressee is able to perform the requested action. In this case you should therefore make use of the option to annotate the utterance as having a qualified Request function, with the attribute 'conditionality' having the value 'conditional'.

Similarly, an indirectly formulated question like *Do you know where I should check in for Munich?* should be annotated as a Set Question with the qualification 'conditional'. (Which comes down to interpreting the question as: *Where should I check in for Munich, if you know that?*.)

5) **"Do not code implicit communicative functions, that can be deduced from functions that you have already assigned."**
Dialogue acts are often performed implicitly. This is in particular the case for positive feedback acts. For example, someone answering a question may be assumed to (believe to) have understood the question. So any time you annotate an utterance as an answer, you might consider annotating it also as providing positive feedback on the interpretation of the question that is answered. For most annotation purposes this would not be sensible, sense It would be redundant. Similarly for reacting to an offer, a request, a suggestion, etc.

6) *Guidelines for the annotation of feedback functions.*
Negative feedback, where the speaker wants to indicate that there was a problem in processing a dialogue utterance, is always explicit and as such mostly easy to annotate.

As already noted, positive feedback, by contrast, is sometimes given explicitly, but very often implicitly.
Examples of explicit positive auto-feedback are the following utterances by B, where he repeats part of the question by A:
A: *What time does the KLM flight from Jakarta on Friday, October 13 arrive?*
B: *The KLM flight from Jakarta on Friday, October 13 has scheduled arrival time 08.50*
B: *The flight from Jakarta on Friday has scheduled arrival time 08.50*
B: *The KLM flight from Jakarta on October 13 has scheduled arrival time 08.50*
B: *The flight from on October 13 has scheduled arrival time 08.50*

In such cases, the utterance by B should be annotated as having, besides the general-purpose function Answer in the Task dimension, also a function in the Auto-Feedback dimension (see below).
By contrast, the short answer: *At 08.50* would carry only implicit feedback information, and should therefore, following Guideline 5, rather not be coded in the Auto-Feedback dimension.

7) *Guidelines for the annotation of Turn Management functions.*
General guideline: **"Code Turn Management functions only when these are not just implied."**
In a spoken dialogue, the participants mostly do not all speak at the same time. Mostly, one participant is the main speaker ('holds the floor'). (The nonverbal behaviour of dialogue participants is *not* organised in turns; all or both participants use facial expressions and gestures more or less all the time.) Turn Management acts are the actions that participants perform in order to manage the allocation of the (main) speaker role. These acts are subdivided into acts for taking the turn (turn-initial acts) and those for keeping the turn or giving it away (turn-final acts). Usually only the first segment in a turn has a turn-initial function and only the last one a turn-final one. The non-final utterances in a turn do not have a turn-final function, except when the speaker signals (for example by using a rising turn-final intonation) that the utterance is not going to be the last one in the turn, that he wants to continue. In that case the utterance has a Turn Keeping function.
When a speaker accepts a turn that the addressee has assigned to him through a Turn Assign act, the relevant segment should be annotated as having the turn-initial function Turn Accept *only* when the speaker performs *a separate act* for the purpose of accepting the turn (such as nodding, or saying something like *Yes* or *OK*). So don't encode this when the turn is accepted implicitly by simply starting to speak.

Similarly, a segment should be annotated as having the turn-initial function Turn Take only if the speaker performs a separate act to that effect. If he just goes ahead and makes a contribution to the dialogue, without first signalling his intention to do so, then the segment should not be marked with a turn-initial Turn Management function.

The verbal as well as nonverbal activities that a speaker performs to seize the turn should be marked as Turn Grabbing, but the segment that follows *after* he has seized the turn should not be marked as having a turn-initial Turn Management function.

Guidelines for the annotation of Time Management functions.
When a speaker is buying time, using fillers such as *Well,..; Let's see,...* , then such a segment should be annotated as having the Stalling function in the Time Management dimension.
There may be several reasons why a speaker wants to have more time; it may be that the speaker has trouble completing his current utterance, or that he is interrupted by some urgent event that requires his attention before he can continue the dialogue. but it may also be that he needs some time to find some information (for instance, for answering a question). So when you encounter a Stalling act, you may well pay attention to the reason why the speaker is stalling. (For instance, Stalling often goes hand in hand with turn acceptance or turn keeping.) However, don't speculate; only code additional functions for which you have evidence.

Guidelines for the annotation of Contact Management functions.

The management of contact in the sense of both partners being ready to send and receive messages to and from each other, is important especially in other than face-to-face situations, such as telephone conversations, video-conferencing, and internet chatting.

Note that in many languages expressions used for establishing contact can often be used for other purposes as well, for example for greeting (*Hello!*). When annotating a dialogue where this happens, the utterance in question should be marked as having both a Contact Management function and a Social Obligation Management function.

Guidelines for the annotation of Own Communication Management functions.
Own Communication Management (OCM) functions should be coded whenever a speaker signals that he made a speech error and/or wants to edit what he is saying. Since this typically requires some extra effort and time, OCM acts often go hand in hand with acts whose function is to win time, such as hesitations (*Ehm...*), which have a Stalling function.

Guidelines for the annotation of Partner Communication Management functions.
Partner Communication Management (PCM) functions should be coded whenever a speaker signals that he believes the addressee made a speech error or has difficulty in completing an utterance, for instance being unable to recall a name or to find the right words to express something. The use of dimension-specific PCM functions for this purpose is typically only possible by interrupting the dialogue partner or in immediate response to a partner utterance.

Guidelines for the annotation of Dialogue Structuring functions.
These functions should be coded only when the speaker explicitly signals something about his intention to open or close the dialogue, or to continue in a particular way.

During a dialogue, the topic is often changed implicitly, simply by talking about a new topic. This happens especially if the new topic is closely related to the previous one, for instance by being a subtopic of the previous topic, or by being another subtopic of a more general topic. Implicit topic management should not be encoded; it would be redundant. Topic Management functions should be annotated only if the speaker explicitly introduces or closes a topic, or signals his intention to do so.

8) *8. Guidelines for annotating Social Obligation Management (SOM) functions.*
Utterances that serve a 'social' purpose such as greetings, thanks, and apologies can often be used for other purposes as well. As already mentioned, greetings like *Hello!* , for example, can be used also for establishing contact (Contact Management function) and/or for opening a dialogue (a Dialogue Structuring function). Also, an expression of thanks can be used to signal that the speaker wants to soon end the dialogue (Dialogue Structuring function Pre-Closing), and can also be used for overall positive feedback. Since these additional functions are not logical consequences that are inferrable form already assigned tags, in such cases the relevant segments should be coded with the appropriate functions in all these dimensions.

Guideline 8: **"When coding an utterance as having a SOM function, look out for additional functions in other dimensions."**

A.4 Note on segmentation

It is recommended to use software for making dialogue act annotations which allow you to segment the dialogue into functional segments in multiple dimensions, as explained in section 8 of this standard.

However, if you're working form a transcription of a spoken dialogue that has been pre-segmented, then the segmentation in the transcript is not necessarily perfect, or not as you would like it to be. First, you may run into cases where you would prefer a previously chosen segment to be segmented further into a sequence of smaller segments, that each have a functional interpretation. In such a case it is best to assign the various tags that you would prefer to assign to the parts to the segment as a whole.

Second, it may happen that a turn has been segmented into certain parts, where you would want to annotate a longer segment, formed by these parts. In such a case it is recommended to annotate all these parts with the same tags.

Third, a previously chosen segment may be 'self-interrupted' by a part that has a different communicative function, as in the following example:

(21) *When***, I mean what time,** *does the train to* **ehm,...** *Viareggio leave?*

Here we see a Set Question interrupted by a Self-Correction (*I mean what time*) and a Stalling segment (*ehm*). If the segmentation has not distinguished the discontinuous and intervening segments, then again, it is best to assign the tags for the intervening segments to the entire segment as a whole.

Fourth, it may happen that a dialogue act corresponds to (parts of) more than one turn, as in the following example, where the utterances in turns 1 and 3 together form an Answer:

1. A: *There are two flights early in the morning, at 7.45 and at 8.15,..*
2. B: *Yes*
3. A: *and two more in the evening, at 7.15 and at 8.30.*

If the presegmentation does not distinguish the segment formed by parts of several turns, then give each of these parts the same tag (Answer, in this example), and code them as having a functional dependency relation with the same question. In this way it is clear that they all respond to the same question.

A.5 Explanation of the use of DiAML tags

. This section will provide a systematic discussion and explanation of the intended use of the DiAML tags, as specified in Annex C. This part of the annex will be completed at a later stage.

Annex B
(informative)
Completely annotated examples

The first example is a two-turn dialogue fragment, also used in Annex C. In this case the two turns are not segmented further; they each constitute a single functional segment in only one dimension (Task). For the anchoring of the DiAML annotation to the primary text, it is assumed that these two functional segments are defined at another level of analysis (see Annex C) as having the XML identifiers "fs1" and "fs2", respectively.

(22) 1. P1: Do you know where I should check in for Munich?
2. P2: For Munich go to counters 31 to 40.

```
<diaml xmlns:"http://www.iso.org/diaml/">
<dialogueAct xml:id="da1" target="#fs1">
speaker="#p1"
addressee="#p2"
communicativeFunction="setQuestion"
dimension="task"
conditionality=""conditional"
</dialogueAct>
<dialogueAct xml:id="da2" target="#fs2">
speaker="#p2"
addressee="#p1"
communicativeFunction="answer"
dimension="task"
functionalDependencyTo="da1"
</dialogueAct>
</diaml>
```

In the second example a stretch of dialogue, again a two-turn dialogue fragment, P2's utterance is segmented into two overlapping functional segments: one in the Auto-Feedback dimension, with positive value, and one in the Task dimension, with value 'answer' qualified as 'uncertain'. As in the previous example and in all examples to follow, the relevant functional segments are assumed to be identified at another level of analysis, and to have XML identifiers like "fs1", "fs2", etc..

(23)

P1:	*Do you know what time the next train to Utrecht leaves?*
P2:	*The next train to Utrecht leaves at 8:32.*
AuFB	The next train to Utrecht [positiveAutoFeedback]
TA	The next train to Utrecht leaves I think at 8:32. [answer, uncertain]

```
<diaml xmlns:"http://www.iso.org/diaml/">
      <dialogueAct xml:id="da1" sender="#p1" addressee="#p2"
                target="#fs1" communicativeFunction="setQuestion"
                dimension="task" conditionality="conditional"/>
      <dialogueAct xml:id="da2" sender="#p2" addressee="#p1"
                target="#fs2" communicativeFunction="autoPositive"
                dimension="autoFeedback" feedbackDependenceTo="fs1"/>
      <dialogueAct xml:id="da3" sender="#p2" addressee="#p1"
                target="#fs2" communicativeFunction="answer"
                dimension="task" functionalDependenceTo="da1"/>
</diaml>
```

In the next example, again a two-turn dialogue fragment (taken from the HCRS Map Task corpus), the turns coincide with functional segments. The example illustrates the use of general-purpose functions for addressing another dimension than that of the task. Participant P1 checks that he understood the previous instruction correctly, producing a Chek Question in the Auto-Feedack dimension. Participant P2 confirms P1's understanding, thus addressing P1's processing of that same instruction, i.e. performing a Confirm act in the Allo-Feedback dimension.

The example also illustrates the use of feedback and functional dependency relations. P1's contribution has a feedback relation to the functional segment ("fs1") expressing the previous instruction, while P2's contribution has both a functional dependency relation to the Check Question that it reacts to, and a feedback dependency relation to the functional segment of that same previous instruction

(24) P1: Slightly northeast?
P2: Yeah very slightly.
(From HCRC Map Task)

```
<diaml xmlns:"http://www.iso.org/diaml/">
      <dialogueAct xml:id="da2" sender="#p1" addressee="#p2"
                target="#fs2" communicativeFunction="checkQuestion"
                dimension="autoFeedback" feedbackDependenceTo="fs1"/>
      <dialogueAct xml:id="da3" sender="#p2" addressee="#p1"
                target="#fs3" communicativeFunction="confirm"
                dimension="alloFeedback" functionalDependenceTo="da2"
                feedbackDependenceTo="fs1"/>
</diaml>
```

The next example, a dialogue fragment from the TRAINS corpus, illustrates the use of a dimension-specific function (Correct Misspeaking) in the dimension of Partner Communication Management (PCM). Notice that PCM act refer to what the main speaker is doing at that moment, as opposed to allo-feedback acts, which refer to what was said in a previous turn. Still, the relation between the Correct Misspeaking act and the functional segment that it refers to is of the same nature as the relation between a feedback act and its trigger, so we use the same 'feedback dependence' relation to indicate this relation.

(25) P1: *engine E3 is going to pick up the bananas, back to Avon, dro...*
P2: *to pick up the oranges*

```
<diaml xmlns:"http://www.iso.org/diaml/">
      <dialogueAct xml:id="da1" sender="#p1" addressee="#p2"
                target="#fs1" communicativeFunction="inform"
                dimension="task"/>
      <dialogueAct xml:id="da2" sender="#p2" addressee="#p1"
                target="#fs2" communicativeFunction="correctMisspeaking"
                dimension="oartnerCommunicationManagement"
                feedbackDependenceTo="fs1"/>
</diaml>
```

This annex will be extended with more annotated examples at a later stage, taking into account and illustrating the annotation guidelines in Annex 3 section A.3.

Annex C
(informative)
ISO-DiAML schema

This annex introduces the technical scheme for the Dialogue Act markup Language DiAML associated with this standard for the concrete representation of annotations of dialogue data with dialogue act information.

C.1 Overview

This representation relies on a three-level architecture:

1) a primary source, which may correspond to a speech recording, textual transcription or any further low-level annotation thereof (e.g. tokenisation or morphosyntactic annotation according to ISO 24611 - MAF);
2) the marking of functional segments from the primary source;
3) the actual dialogue act annotation associated with a functional segment.

This annex provides a specification for this third level (the dialogue act annotation) as well as implementation guidelines for the two others.

The annotation of a dialogue act associated to a functional segment is done by means of the `dialogueAct` element. This element has the following attributes:

- `sender`, `addressee`
- `communicativeFunction`, `dimension` - see section 11.2 for the possible values of these attributes;
- `partiality`, `conditionality`, `modality`, `attitudeEmotionality` – - see section 11.2 for the possible values of these attributes;
- `feedbackDependenceTo`, `functionalDependenceTo`.

Functional segments can be identified by means of the `functionalSegment` element, which is syntactic sugar for the TEI construct `<span type="functionalSegment">`. Functional segments can be grouped together within a functionalSegmentGrp, which is syntactic sugar for `<spanGrp type="functionalSegment">`. The `functionalSegment` element associates a functional segment directly with a span of text; `functionalSegmentGrp` (functional segment group) collects a number of `functionalSegment` tags. A functional segment can then be identified with reference to the corresponding span in the source document. The `target` attribute can be used to point to the functional segment. (The `target` attribute can denote any TEI pointer reference, thus allowing discontinuous segments to be referred to. (Another valid way of indicating the relations between functional segments and dialogue act information would be to use the `ana` attribute (for 'analysis'), which can be used as an attribute for a `functionalSegment` element, pointing to all the dialogue acts associated with the segment. These dialogue act elements should then be collected in a `dialogueActGrp` (dialogue act group) element.)

C.2 Example

The following excerpt exemplifies how the three above mentioned levels may be instantiated in the specific case of a tokenised primary source encoded according to the TEI guidelines. The source represents two sentences forming a small dialogue fragment as follows. Two alternative XML representations are shown of the dialogue act information associated with the primary data, one using the XML encoding of feature structures according to joint TEI-ISO standard 24610-1; the other using an XML encoding of the DiAML concrete syntax introduced in section 11 of this standard.

The `TEI header` contains metadata, including the identity of the dialogue participants.

```
<TEI xmlns:="http://www.tei-c.org/ns/1.0"
```

```
          xmlns:xlink="http://www.w3.org/1999/xlink">
    <teiheader>
      <fileDescr>
        <titleStmt>
          <title>DiAML annotation example </title>
        </titleStmt>
        <publicationStmt>
             <p>...</p>
        </publicationStmt>
          <sourceDescr>
             <p>...</p>
          </sourceDescr>
      </fileDescr>
      <profileDescr>
        <particDescr>
          <person xml:id="p1">
             <p>the first speaker</p>
          </person>
          <person xml:id="p2">
             <p>the second speaker</p>
          </person>
        </particDescr>
      </profileDescr>
  </teiheader>
<text>
   <body>
     <div>
       <head>Simple dialogue fragment</head>
       <u>Do you know where I should check in for Munich?</u>
       <u>For Munich go to counters 31 to 40</u>
     </div>
     <div>
       <head>The dialogue turns, segmented into words (TEI-compliant)</head>
       <u>
          <w xml:id="1">Do</w>
          <w xml:id="2">you</w>
          <w xml:id="3">tknow</w>
          <w xml:id="4">where</w>
          <w xml:id="5">I</w>
          <w xml:id="6">should</w>
          <w xml:id="7">check</w>
          <w xml:id="8">in</w>
          <w xml:id="9">for</w>
          <w xml:id="10">Munich</w>
          <pc>?</pc>
       </u>
       <u>
          <w xml:id="11">Forn</w>
          <w xml:id="12">Munich</w>
          <w xml:id="13">go</w>
          <w xml:id="14">to</w>
          <w xml:id="15">counters</w>
          <w xml:id="16">32</w>
          <w xml:id="17">to</w>
          <w xml:id="18">40</w>
       </u>
         </div>
```

```
    <div>
    <head>Identification of functional segments</head>
     <spanGrp type="functionalSegments"
          <span xml:id="fs1" from="1" to="11" />
                          <span xml:id="fs2" from="12" to="18" />
     </spanGrp>
    </div>
<div>
   <head>Representation by means of feature structures
             in TEI/ISO-compliant format</head>
   <fs type="dialogueAct" xml:id="da1" target="#fs1">
        <f name="speaker" fVal="#p1"/>
        <f name="addressee" fVal="#p2"/>
        <f name="communicativeFunction"
<symbol value="setQuestion"/></f>
        <f name="dimension" <symbol value="task"/></f>
        <f name="conditionality" <symbol value=âĂŽ"conditional"/></f>
   </fs>
   <fs type="dialogueAct" xml:id="da2" target="#fs2">
        <f name="speaker" fVal="#p2"/>
        <f name="addressee" fVal="#p1"/>
        <f name="communicativeFunction" <symbol value="answer"/></f>
        <f name="dimension" <symbol value="task"/></f>
        <f name="functionalDependencyTo"="da1"/></f>
   </fs>
</div>
<div>
   <head>Representation using DiAML syntax</head>
   <dialogueAct xml:id="da1" target="#fs1">
        speaker="#p1"
        addressee="#p2"
        communicativeFunction="setQuestion"
        dimension="task"
        conditionality="conditional"
   </dialogueAct>
   <dialogueAct xml:id="da2" target="#fs2">
        speaker="#p2"
        addressee="#p1"
        communicativeFunction="answer"
        dimension="task"
        functionalDependencyTo="da1"
   </dialogueAct>
             </div>
          </body>
     </text>
</TEI>
```

Annex D
(informative)
Data categories for core communicative functions

D.1 Overview

Data categories may have a *conceptual domain* which consists of its specializations; in this case the functions Propositional Question and Set Question, also known in the literature as Yes/No-question and WH-question, respectively, and described by the data categories /propositionalQuestion/ and /setQuestion/. Besides the definition we see a *source*, which is rather unspecific in this case in view of the commonplace character of the Question function, and a *note* which may be used for all kinds of remarks, in particular for mentioning related concepts or related terminology.

The second example, that of the Inform function, illustrates two elements that a data category entry may have, in addition to what we have seen in the first example: (1) an *explanation*, which may provide useful information that does not strictly belong to the definition of the concept, but helps to understand it and place it in perspective; (2) an *example*.

D.2 General-purpose functions

D.2.1 Information-seeking functions

	/**propositionalQuestion**/
Conceptual domain	/checkQuestion/
Definition	Dialogue act where the speaker, S, wants to know whether a certain proposition is true or false. S assumes that the addressee, A, posseses that information, and puts pressure on A to inform S whether the proposition is true or false.
– Source	Commonplace
– Note	Related terminology in other schemes: QUERY-YN (HCRC MapTask), Yes-No-Question (SWBD-DAMSL) and YNQ (TRAINS).
Explanation	A propositional question corresponds to what is commonly termed a YN-question or polarity question in the literature. The term 'propositional' is preferred because: (a) it clearly separates form from function by removing any oblique reference in the label to syntactic criteria for the identification of such acts; and (b) it is not a language specific term.SWBD-DAMSL for example conflates form and function by distinguishing between propositional questions that are marked explicitly by subject inversion (yes-no questions) and those that are marked by intonation alone (declarative questions). However, though they may have different realisations, these are in fact performing the same function.
Example	"Have you got a haystack on your map?"
–Source	HCRC MapTask

	/**checkQuestion**/
Broader concept	/propositionalQuestion/
Definition	Dialogue act where the speaker, S, wants to know whether a given proposition is true, about which S holds an uncertain belief that it is true S. S assumes that A knows whether the proposition is true or not, and puts pressure on A to provide this information
– Source	Commonplace
– Note	Related terminology in other schemes: Check (DIT, HCRC MapTask), Tag Question (SWBD-DAMSL), Tag (TRAINS) and Request_Comment (Verbmobil)
Example	"The meeting starts at ten, right?"

	/setQuestion/
Conceptual domain	/choiceQuestion/
Definition	Dialogue act where the speaker, S, wants to know which elements of a certain set have a named property. S puts pressure on the addressee, A, to provide S with this information. S believes that at least one element of the set has the named property, and S assumes that A knows which are the elements of the set that have the property.
– Source	Commonplace
– Note	Related terminology in other schemes: QUERY-W (HCRC MapTask), WH-Question (SWBD-DAMSL) and WHQ (TRAINS).
Explanation	A set question corresponds to what is commonly termed a WH-question in the literature. The term set is preferred because: (a) it clearly separates form from function by removing any oblique reference in the label to syntactic criteria for the identification of such acts; and (b) it is not a language specific term (it may be further noted that even in English, not all questioning words begin with 'wh', e.g. "How?").
Example	"What time does the meeting start?"

	/choiceQuestion/
Broader concept	/setQuestion/
Definition	Dialogue act where the speaker, S, wants to know which one from a given list of alternative propositions is true; S believes that exactly one element of that list is true; S assumes that the addressee, A, knows which of the alternative propositions is true, and S puts pressure on A to provide this information.
– Source	DAMSL; DIT
– Note	Related terminology in other schemes: Alternatives Question (DIT, LIRICS), QUERY-W (HCRC MapTask), Or-Question/Or-Clause (SWBD-DAMSL, MRDA).
Explanation	It is not very common in annotation schemes to specifically distinguish the concept of choice questions from that of set questions (although it is common in the literature on interrogatives, see for instance: Tsui 1994). However, whereas it is common for the concept set question to carry the expectation that all members of the set with a given property should be returned by the addressee, for a choicequestion the expectation is that there will be exactly one. The different preconditions and effects indicate that these are semantically different concepts, and they have been treated here as such.
Example	"Does she live in Amsterdam or in Rotterdam?"

D.2.2 Information-providing functions

	/inform/
Conceptual domain	/agreement/ /disagreement/
Definition	Dialogue act where the speaker, S, wants to make certain information known to the addressee, A; S believes that the information is correct.
– Source	Commonplace
– Note	Related terminology in other schemes: Assert (DAMSL), Explain (HCRC MapTask), Update (LINLIN), Statement (SWBD-DAMSL) and Inform (DIT, TRAINS, Verbmobil).
Explanation	The inform function may also have more specific rhetorical functions such as: explain, elaborate, exemplify and justify, but these all fall under the more generic function here defined.
Example	"The 6.34 to Breda leaves from platform 2."

	/agreement/
Definition	Dialogue act where the speaker, S, wants to inform the addressee, A, that the information which S has reason to believe that A believes is correct, is in fact correct.
– Source	DIT
– Note	Related terminology in other schemes: Accept (DAMSL, SWBD-DAMSL, TRAINS, Verbmobil).
Explanation	DAMSL and SWBD-DAMSL use "Agreement" to refer to various degrees in which a speaker accepts some previous proposal, plan, opinion or statement; "accept" is one of these degrees; "reject" is another.Note: in this definition /agreement/ inherits the elements in the definition of /inform/.
Example	"Exactly"; Dutch" "Precies!"; Danish: "Netop!"
–Source IT	D

	/**disagreement**/
Conceptual domain	/correction/
Definition	Dialogue act where the speaker, S, wants to inform the addressee, A, that the information which S has reason to believe that A believes is correct, is in fact incorrect.
– Source	DIT
– Note	Related terminology in other schemes: Reject (DAMSL, SWBD-DAMSL, Verbmobil) and Denial (TRAINS).
Explanation	DAMSL and SWBD-DAMSL use "Agreement" to refer to various degrees in which a speaker accepts some previous proposal, plan, opinion or statement; "accept" is one of these degrees; "reject" is another.Note: in this definition /disagreement/ inherits the elements in the definition of /inform/.
Example	"I'm afraid you're wrong."

	/**correction**/
Definition	Dialogue act where the speaker, S, wants to inform the addressee, A, that the information which S has reason to believe that A believes is correct, is in fact incorrect and should be replaced by the information that S offers.
– Source	Commonplace
– Note	In this definition /correction/ inherits the elements in the definition of /disagreement/.
Example	"To Montreal, not to Ottawa."

	/**answer** /
Broader concept	/inform/
Conceptual domain	/confirm/ /disconfirm/
Definition	Dialogue act where the sender, S, wants to make certain information available to the addressee, A, which S believes A wants to know.
– Source	Commonplace

	/**confirm**/
Broader concept	/answer/
Definition	Dialogue act where the sender, S, believes that the addressee, A, wants to know whether his (A's) uncertain belief that the information queried by a check is correct.
– Source	DIT; Verbmobil
– Note	Related terminology in other schemes: Reply-Y (HCRC MapTask), Yes-Answer (SWBD-DAMSL).
Example	"Indeed"

	/**disconfirm**/
Broader concept	/answer/
Definition	Dialogue act where the sender, S, believes that the addressee, A, wants to know whether his (A's) uncertain belief that the information queried by a check is incorrect.
–Source	DIT
–Note	Related terminology in other schemes: Reply-N (HCRC MapTask) and No-Answer (SWBD-DAMSL).
Example	French "Si"; Danish "Jo"; Dutch: "Toch niet" and "Toch wel"

D.2.3 Commissives

	/**promise**/
Conceptual domain	/acceptRequest/ /declineRequest/
Definition	Dialogue act where the sender, S, commits himself unconditionally to perform a certain action in the manner or with the frequency described. S believes that the addressee, A, prefers that the action be performed (rather than not be performed).
– Source	DIT, Searle (1969)
– Note	Related terminology in other schemes: Promise (TRAINS)
Example	"I will send you an email"

	/**offer**/
Definition	Dialogue act where the sender, S, commits himself to perform a certain action, conditional on A's consent that S do so.
– Source	DAMSL; DIT
– Note	Related terminology in other schemes: Offer (TRAINS).
Example	"Shall I start?"; "Would you like to have some coffee?"

	/**acceptRequest**/
Definition	Dialogue act where the sender, S, commits himself to perform an action that was requested.
– Source	DIT
– Note	Related terminology in other schemes: Accept (DAMSL, SWBD-DAMSL, TRAINS, Verbmobil).
Example	"Sure"

	/**declineRequest**/
Definition	Dialogue act where the sender, S, indicates unwillingness to perform an action that was requested.
– Source	DIT
– Note	Related terminology in other schemes: Reject (DAMSL, SWBD-DAMSL, TRAINS, Verbmobil).
Example	"Not now"

	/**acceptSuggest**/
Definition	Dialogue act where t.
– Source	DIT
– Note	Related terminology in other schemes: Accept (DAMSL, SWBD-DAMSL, TRAINS, Verbmobil).
Example	"Let's do that"

	/**declineSuggest**/
Definition	Dialogue act where t.
– Source	DIT
– Note	Related terminology in other schemes: Reject (DAMSL, SWBD-DAMSL, TRAINS, Verbmobil).
Example	"I don't think so"

D.2.4 Directives

	/**instruct**/
Broader concept	/request/
Conceptual domain	/acceptOffer/ /declineOffer/
Definition	Dialogue act where the sender, S, wants the addressee, A, to carry out a named action in the manner or with the frequency described; S assumes that A is able and willing to carry out the action.
– Source	DIT; HCRC Map Task
– Note	Related terminology in other schemes: Action-directive (DAMSL).
Example	"Go right round until you get to just above that."
–Source	HCRC MapTask

	/suggest/
Definition	Dialogue act where the sender, S, wants the addressee, A, to be aware that a named action is potentially promising for achieving a certain goal, which is either named explicitly or contextually salient.
– Source	DIT; TRAINS; Verbmobil
– Note	Related terminology in other schemes: Open-option (DAMSL).
Example	"Let's wait for the meeting to finish."

	/request/
Conceptual domain	/instruct/
Definition	Dialogue act where the sender, S, wants the addressee, A, to perform a named action in the manner or with the frequency described, conditional on A's consent.
– Source	DIT; TRAINS; Verbmobil
Example	"Please turn to page five"

	/acceptOffer/
Definition	Dialogue act where the sender, S, informs the addressee, A, that S agrees to A performing the action that A has offered to perform.
– Source	DIT
– Note	Related terminology in other schemes: Accept (DAMSL, SWBD-DAMSL, TRAINS, Verbmobil).
Example	"Yes please"; French: "Je vous en prie"

	/declineOffer/
Definition	Dialogue act where the sender, S, informs the addressee, A, that S does not agree to A performing the action that A has offered to perform.
– Source	DIT
– Note	Related terminology in other schemes: Reject (DAMSL, SWBD-DAMSL, TRAINS, Verbmobil).
Example	"No thanks"

D.3 Feedback functions

	/positiveAutoFeedback/
Definition	Dialogue act where the sender, S, wants the addressee A to know that S believes that S's attention to, perception, interpretation, evaluation or execution of the previous utterance was successful.
– Source	DIT
– Note	Related terminology in other schemes: Signal-Understanding (DAMSL), Acknowledge (HCRC MapTask, SWBD-DAMSL) Ack (TRAINS) and Feedback_Positive (Verbmobil).This type of feedback could be further broken down into more specific areas (dealing with the sender's attention, perception, interpretation, evaluation and execution), as exemplified in the DIT schema. Such fine distinctions have hitherto not been made in other annotation schemes however, so a simplified top level data category is defined here.
Example	"Uh-huh"; Nonverbally: nodding; "Yes"

	/positiveAlloFeedback/
Definition	Dialogue act where the sender, S, wants the addressee, A, to know that S believes that A's attention to, perception, interpretation, evaluation or execution of the previous utterance was successful.
– Source	DIT
– Note	The distinction between whether feedback is about S's (auto) understanding or A's (allo) is only made within the DIT scheme. This type of feedback could be further broken down into more specific areas (dealing with the addressee's attention, perception, interpretation, evaluation and execution).
Example	"You got that right"

	/**negativeAutoFeedback**/
Definition	Dialogue act where the sender, S, wants the addressee, A, to know that S believes that S's attention to, perception, interpretation, evaluation or execution of the previous utterance encountered a problem.
– Source	DIT
– Note	Related terminology in other schemes: Signal-Non-Understanding (DAMSL) and Feedback _Negative (Verbmobil). This type of feedback could be further broken down into more specific areas (dealing with the sender's attention, perception, interpretation, evaluation and execution), as is exemplified in the DIT schema. Such fine distinctions have hitherto not been made in other annotation schemes however, so a simplified top level data category is defined here.
Example	"Sorry?"; "What?"; Spanish: "Que?"; Italian: "Como?"

	/**negativeAlloFeedback**/
Definition	Dialogue act where the sender, S, wants the addressee, A, to know that S believes that A's attention to, perception, interpretation, evaluation or execution of the previous utterance encountered a problem.
– Source	DIT
– Note	The distinction between whether feedback is about S's (auto) understanding or A's (allo) is only made within the DIT scheme. This type of feedback could be further broken down into more specific areas (dealing with the addressee's attention, perception, interpretation, evaluation and execution).
Example	"No no no no no"

	/**feedbackElicitation**/
Definition	Dialogue act where the sender, S, wants to know whether A's attention to, perception, interpretation, evaluation or execution of the previous utterance was successful.
– Source	DIT
– Note	Feedback elicitation could be further broken down into more specific areas dealing with the addressee's attention, perception, interpretation, evaluation and execution.
Example	"Okay?"; Italian: "Capisce?"

D.4 Turn management functions

	/**turnAccept**/
Definition	Dialogue act where the sender, S, agrees to take the turn when he is requested to do so.
– Source	DIT
– Note	Related terminology in other schemes: Take-Turn (TRAINS).
Example	A: "Would you like to say something at this point?" B: "Certainly."; Nonverbally: nodding

	/**turnAssign**/
Definition	Dialogue act where the sender, S, wants the addressee, A, to take the turn.
– Source	Common in literature on turn taking in conversation
– Note	Occurs especially in multiparty dialogue.Related terminology in other schemes: Assign-Turn (TRAINS).
Example	"Adam?", characteristically accompanied by the pseaker directing his gaze to Adam, possibly also nodding or pointing in his direction and raising the eyebrows.

	/**turnGrab**/
Definition	Dialogue act where the sender, S, wants to take the turn from another participant.
– Source	DIT
– Note	Related terminology in other schemes: Take-Turn (TRAINS).
Example	"Hold on"

	/**turnKeep**/
Definition	Dialogue act where the sender, S, wants to keep the turn.
– Source	DIT
– Note	Related terminology in other schemes: Keep-Turn (TRAINS).
Explanation	Utterances used for turn keeping often also have a stalling function.
Example	"Erm"

	/**turnRelease**/
Definition	Dialogue act where the sender, S, wants to give other dialogue participants the opportunity to take the turn
– Source	Common in literature on turn taking in conversation
– Note	Related terminology in other schemes: Release-Turn (TRAINS).
Example	Sender uses declining intonation towards the end of a contribution and subsequently pauses.

	/**turnTake**/
Definition	Dialogue act where the sender, S, wants to take the turn when it is available.
– Source	DIT
– Note	Related terminology in other schemes: Take-Turn (TRAINS)
Example	"Ehm..." as a turn-initial segment

D.5 Time management functions

	/**stalling**/
Definition	Dialogue act where the sender, S, wants to have a little more time to construct his contribution.
– Source	DIT
– Note	Related terminology in other schemes: Hold (DAMSL).
Example	"Let me see...", "Erm..."; Nonverbally: slowing down
–Note	Utterances used for stalling often also have a turn keeping function.

	/**pausing**/
Definition	Dialogue act where the sender, S, wants to suspend the dialogue for a while because he needs some time to do something.
– Source	DIT
– Note	Related terminology in other schemes: Hold (DAMSL).
Explanation	Pausing occurs either in preparation of continuing the dialogue, or because something else came up which is more urgent for the sender to attend to.
Example	"Just a moment"; Dutch: “een ogenblikje"

D.6 Own and partner communication management functions

	/**completion**/
Definition	Dialogue act where the sender, S, wants to assist the addressee, A, by finishing or adding to the clause that A is in the middle of constructing.
– Source	DAMSL; DIT; TRAINS
Example	S: "which should leave us plenty of time to uhhh", A: "get to city H"
– Source	TRAINS

	/**correctMisspeaking** /
Definition	Dialogue act where the sender, S, wants to correct (part of) an utterance by the addressee, A, assuming that A made a speaking error.
– Source	DAMSL; DIT
– Note	Related terminology in other schemes: Correction suggestion (TRAINS).
Example	S: "second engine E3 is going to uhh city H to pick up the bananas, back to A, drop...", A: "to pick up the oranges", S: "sorry, pick up the oranges"
–Source	TRAINS

	/**signalSpeakingError**/
Definition	Dialogue act where the sender, S, wants the addressee, A, to know that S has made a mistake in speaking.
– Source	DIT
Example	"We're going out on Tues- no, er, not on Tuesday"

	/**selfCorrection**/
Definition	Dialogue act where the sender, S, wants to correct an error that he made, or to improve on an infelicitous formulation that he used, within the same turn.
– Source	Common in literature on conversation studies
– Note	Related terminology in other schemes:
Example	"We're going out on Tues- no, er, Thursday"

D.7 Discourse structuring functions

	/**interactionStructuring**/
Definition	Dialogue act where the sender, S, wants to explicitly indicate to the addressee, A, the function or topic of his next contribution(s).
– Source	DIT
– Note	Interaction structuring covers such phenomena as topic introduction, dialogue act announcement and topic closing.
Examples	"A question"; Dutch: "vraagje"

	/**opening**/
Definition	Dialogue act where the sender, S, wants the addressee, A, to know that S is ready and willing to engage in a dialogue with A.
– Source	DIT
Example	""

D.8 Social obligations management functions

	/**initialGreeting**/
Definition	Dialogue act where the sender, S, wants the addressee, A to know that S is present and aware of A's presence; S puts pressure on A to acknowledge this.
– Source	DIT
– Note	Related terminology in other schemes: Greet (Verbmobil).
Explanation	Greetings usually come in initiative-response pairs within a dialogue; this data category corresponds to the first element of such a pair.
Example	"Hello!"; "Good morning"

	/**returnGreeting**/
Definition	Dialogue act where the sender, S, wants to acknowledge that S is aware of the presence of the addressee, A, and of A having signalled his presence to S; S has been pressured to respond to an initialGreeting by A.
– Source	DIT
– Note	Related terminology in other schemes: Greet (Verbmobil).
Explanation	Greetings usually come in initiative-response pairs within a dialogue; this data category corresponds to the second element of such a pair.
Example	"Hello!"; "Good morning"

	/**initialSelfIntroduction**/
Definition	Dialogue act where the sender, S, wants to make himself known to the addressee, A; S puts pressure on A to acknowledge this.
– Source	DIT
Explanation	Introductions usually come in initiative-response pairs within a dialogue; this data category corresponds to the first element of such a pair.
Example	"I'm Jack"

	/**returnSelfIntroduction**/
Definition	Dialogue act where the sender, S, wants to make himself known to the addressee, A; S has been pressured to respond to an initialSelfIntroduction by A.
– Source	DIT
Explanation	Introductions usually come in initiative-response pairs within a dialogue; this data category corresponds to the second element of such a pair.
Example	"And I'm Jill"

	/**apology**/
Definition	Dialogue act where the sender, S, wants the addressee, A, to know that S regrets something; S puts pressure on A to acknowledge this.
– Source	DIT; SWBD-DAMSL
Example	"Sorry about that."

	/**acceptApology**/
Definition	Dialogue act where the sender, S, wants to mitigate the addressee, A's feelings of regret; S has been pressured to respond to an apology by A.
– Source	DIT
Example	"No problem."

	/**thanking**/
Definition	Dialogue act where the sender, S, wants the addressee, A, to know that S is grateful for some action performed by A; S puts pressure on A to acknowledge this.
– Source	DIT
– Note	Related terminology in other schemes: Thank (Verbmobil).
Explanation	Utterances used for thanking often also indicate that the sender wants to end the dialogue.
Example	"Thanks a lot."

	/**acceptThanking**/
Definition	Dialogue act where the sender, S, wants to mitigate or respond to the addressee, A's feelings of gratitude; S has been pressured to respond to an act of thanking by A.
– Source	DIT
Example	"Don't mention it"; Spanish: "De nada".

	/**initialGoodbye**/
Definition	Dialogue act where the sender, S, wants the addressee, A, to know that S intends the current utterance to be his final contribution to the dialogue; S puts pressure on A to acknowledge this.
– Source	DIT
– Note	Related terminology in other schemes: Bye (Verbmobil).
Explanation	Goodbyes usually come in initiative-response pairs within a dialogue; this data category corresponds to the first element of such a pair.
Example	"Bye, see you later"

	/**returnGoodbye**/
Definition	Dialogue act where the sender, S, wants to acknowledge his awareness that the addressee, A, has signalled his final contribution to the dialogue and S signals in return his agreement to end the dialogue; S has been pressured to respond to an initialGoodbye by A.
– Source	DIT
– Note	Related terminology in other schemes: Bye (Verbmobil).
Explanation	Goodbyes usually come in initiative-response pairs within a dialogue; this data category corresponds to the second element of such a pair.
Example	"Bye, see you later"

Annex E
(informative)
Data categories for non-core communicative functions

E.1 Contact management functions

	/**contactIndication**/
Definition	Dialogue act where the sender, S, wants to make it known to the addressee, A, that S is ready to send messages to, and receive messages from, A.
– Source	DIT
Example	"Yes?"

	/**contactCheck**/
Definition	Dialogue act where the sender, S, wants to establish whether the addressee, A, is ready to receive messages from, and send messages to, S.
– Source	DIT
Example	"Hello?!"

E.2 Other functions

	/**directQuestion**/
Conceptual Domain	/set Question/ /propositional Question/ /alternative Question/ /check Question/
Definition	Dialogue act where the sender, S, wants to know something which S assumes the addressee, A, to know. S puts pressure on A to provide this information
– Source	Commonly used as contrasting with /indirectQuestion/

Annex F
(informative)
Results from survey of dimensions and communicative functions in existing annotation schemas

Summary of results of Petukhova & Bunt (2009) to be provided here.

As part of the project to establish the present standard, a detailed study was conducted in order to provide theoretical and empirical arguments for identifying core dimensions and communicative function. The study included a survey of the literature on dialogue analysis and of the use of functions and dimensions in 18 existing annotation schemas. Moreover, a number of statistical and machine-learning test were carried out in order to identify dependencies among potential dimensions.

From this study, published by the Centre for Creative Computing at Tilburg University (Petukhova & Bunt, 2009a), a highly condensed version of which was presented at the 2009 NAACL-HLT conference (Boulder, Colorado, May 2009) and published as Petukhova & Bunt (2009b), a number of tables are copied and the conclusions of the study are quoted entirely.

Table F in this annex shows the relative frequencies of functional segments in 10 dimensions (9 of which are core dimensions in the present standard proposal) for three different dialogue corpora. The variation between the corpora is worth noting.

Table F shows the relative frequencies of functional segments addressing only one dimension. From this table it may be concluded that the 10 dimensions considered in the table are all independently addressable, which is a crucial element in the definition of a dimension.

The tables 3-6 show the occurrence of dimension-specific communicative functions in various dimensions in 18 existing annotation schemas. The tables 7-10, which were not part of Petukhova & Bunt (2009a) but which have been compiled more recently, show the occurrence of general-prupose functions in these annotation schemas.

The conclusion form this study was as follows:

"The analysis shows that eight dimensions, namely **Task**, **Feedback**, **Turn Management**, **Social Obligations Management**, **Own Communication Management**, **Discourse Structuring**, **Partner Communication Management** and **Time Management** fulfil all five criteria, and can be considered as 'core' aspects of dialogue communication. Our conclusion with respect to Feedback is moreover that a distinction should be made at least between **Feedback giving** and **Feedback eliciting** aspects, since dialogue participants not only report about successes and failures of their own processing of previous utterances, but also constantly evaluate the partner's cognitive state, message processing, and degree of involvement in the communication, and may elicit information about these aspects. Making only the distinction between feedback-giving and feedback-eliciting acts, however, does not to justice to the fact that feedback-giving acts can report not only on the speaker's own processing of previous dialogue but also on the speakers beliefs about the addressee's processing - a distinction which is semantically important and which is captured by the distinction between **Auto-** and **Allo-Feedback**. Note also that the phi-coefficient (-0.3) indicates that Auto- and Allo-Feedback are not very closely related. These arguments support the suggestion to distinguish the two as separate dimensions.

Time Management acts are also close to Turn Management acts, since speakers often need a bit of time to formulate their contribution when they take (or have and want to keep) the turn. This consideration applies only to *stallings* under certain context conditions, however; *pausing*, by contrast, does not imply that the speaker wants to keep the turn. It should be also noticed that *stallings* do not always imply that the speaker wants to keep the turn; extensive amounts of protraction accompanied by certain non-verbal behaviour may indicate that the speaker needs assistance. It was noticed by Butterworth (1980) that an excessive amount of gaze aversion may also lead a listener to infer that the speaker is having difficulty formulating a message. Moreover, as Clark (1996) in shows, time delays are not always are used for turn-keeping purposes, because even in monologues where speakers do not need to keep the turn, time delays

are frequently used. Time and Turn Management are therefore better kept apart rather than considered as one dimension.

A third view on **Time Management** acts is that they are produced unintentionally, *stallings* in particular. They should therefore perhaps not be regarded as dialogue acts. An act that is not consciously intentional may still be relevant, however; for example, humans produce a lot of facial expressions unconsciously, but they display the emotional or cognitive state of the dialogue participant, which is obviously important for dialogue analysis. In other words, they affect the information states of dialogue participants if they have shared encoded meaning. Goffman (1963) points out that the receiver is always responsible for the interpretation of an act as intentional or not. Kendon (2004) also notices that whether an action is deemed to be intended or not is something that is dependent entirely upon how that action appears to others. So this does not provide a good argument against viewing Time Management as a dimension of dialogue communication.

Our conclusion is that **Contact Management** could be considered as an 'optional' dimension, since this aspect of communication is not reflected in most existing dialogue act annotation schemes (6 out of 18). It was noticed, however, that for some types of dialogues, e.g. phone conversations or tele-conferences (as in the OVIS corpus), this aspect may be important."

Contact Management is considered as an optional, non-core dimension. The standard is open to the addition of other non-core dimensions, provided that they meet the requirement of being addressable independently of the other dimensions.

	AMI corpus	DIAMOND corpus	OVIS corpus
Task	33	47.7	48.8
Auto-Feedback	20	14	18
Allo-Feedback	0.7	3.8	39
Turn Management	15	14	1
Social Obligation Management	0.3	5	3.8
Discourse Structuring	2.2	2.3	2.4
Own Communication Management	8.7	0.7	0.3
Time Management	16.8	10.7	0.6
Partner Communication Management	0.3	0.3	0.1
Contact Management	0.1	1.3	12.3

Table 1 — ***Distribution of utterances across dimensions for analysed dialogue corpora in (%).***

Dimension	Frequency (in %)		
	AMI	OVIS	DIAMOND
Task	28.8	37.9	29.9
Auto-Feednack	14.2	16.3	20.9
Allo-Feedback	0.7	4.1	6.8
Turn Management	7.4	0.9	8.5
Time Management	0.3	0.4	0.7
Contact Management	0.1	0.3	0.7
Discourse Structuring	1.9	1.8	2.7
Own Communication Management	0.5	0.8	2.7
Partner Communication Management	0.2	3.1	0.4
Social Obligation Management	0.3	6.4	0.7

Table 2 — ***Distribution of functional segments addressing a single dimension for three dialogue corpora in (%).***

DIT	Positive attention	Positive perception	Positive interpretation	Positive evaluation	Positive execution
LIRICS	Positive Auto-Feedback				
DAMSL	Signal understanding		Acknowledgment		
SWBD-DAMSL	Signal understanding		Acknowledgment Repeat-rephrase		Summarize-reformulate
MRDA	Signal understanding		Acknowledgment Appreciation	Assessment	
Coconut	Signal understanding		Acknowledgment Repeat-rephrase		
AMI	Comment-about-understanding POS			Assess	Inform POS
HCRC MapTask			Acknowledgment		
Verbmobil	Backchannel			Acknowledge	Positive feedback
SLSA	Pos.contact	Pos.perception	Pos.understanding	Pos. acceptance/attitude	
TRAINS	Acknowledgement			Pos.evaluation	
SPAAC	Echo		Acknowledge	Appreciate	
MALTUS	Pos. attention	Repeat-rephrase		Appreciation	
Chiba	Follow up: pos. understand			Pos. response	
Alparon			Acknowledgement		
C-Star			Acknowledgement		

Table 3 — *Positive Auto-feedback functions in different dialogue act taxonomies.*

DIT	Negative attention	Negative perception	Negative interpretation	Negative evaluation	Negative execution
LIRICS	Negative Auto-Feedback				
DAMSL	Signal-non-understanding				
SWBD-DAMSL	Signal-non-understanding				
MRDA	Signal-non-understanding		Understanding Check		
Coconut	Signal-non-understanding		Clarification Check		
AMI	Comment-about-uderstanding NEG				Inform NEG
HCRC MapTask	Check				
Verbmobil	Request clarify				Neg.feedback
SLSA	Neg.contact	Neg.perception	Neg. understanding	Neg. attitude	
TRAINS			Neg. evaluation		
SPAAC	Pardon				
MALTUS	Neg.attention				
Chiba	Follow up: understand			Neg. response	
Alparon					
C-Star					

Table 4 — *Negative Auto-feedback communicative functions in different dialogue act taxonomies.*

DIT	Turn take	Turn grab	Turn accept	Turn keep	Turn assign	Turn release
LIRICS	Turn take	Turn grab	Turn accept	Turn keep	Turn assign	Turn release
DAMSL				Turn maintain		
SWBD-DAMSL			Hold before answers	Turn maintain	Turn exit	
MRDA	Regain turn	Grabber	Hold before answers	Holder		
Coconut				Turn maintain		
SLSA	Turn take	Interruption	Turn opening	Turn holding	Turn closing	
TRAINS	Turn take			Turn keep	Turn assign	Turn release
SPAAC				Hold		
MALTUS		Turn grabber		Turn holder		Back-channel
Primula		Turn grabber		Turn holder		Back-channel
Chiba				Hold		

Table 5 — ***Turn Management communicative functions in different dialogue act taxonomies.***

DIT	Greeting/ return greeting	Self-introduction/ return self-introdunction	Goodbye/ return goodbye	Apology/ accept apology	Thanking/ accept thanking
LIRICS	Greeting/ return greeting	Self-introduction/ return self-introdunction	Goodbye/ return goodbye	Apology/ accept apology	Thanking/ accept thanking
DAMSL	Greeting		Goodbye		
SWBD-DAMSL	Greeting			Apology/ downplayer	Thanking/ downplayer
MRDA				Downplayer/ sympathy	Thanking
Coconut	Greeting		Goodbye		
AMI	Be-positive/be-negative				
Verbmobil	Greet	Introduce	Bye	Polite (apologies and compliments)	Thank
SLSA	Greet				
TRAINS	Greet				
MALTUS	Politeness				
Primula	Politeness; face-threatening/face-saving				
Alparon	Greet		Bye		
C-Star	Greeting	Self-introduction		Apologize	Thanking

Table 6 — ***Social Obligation Management communicative functions in different dialogue act taxonomies.***

DIT	Opening	Pre-closing	Topic introduction	Topic shift	Topic shift announcement
LIRICS	Interaction structuring				
DAMSL	Opening	Closing			
SWBD-DAMSL	Opening	Closing			
MRDA			Topic change		
Coconut	Opening	Closing	Topic		
AMI	Argument structure and topic segmentation schemes				
HCRC MapTask			Ready (for topic shifts)		
Verbmobil		Task close	Task initiate	Digress	
LinLin	Opening	Closing	Topic layer		
SLSA	Opening	Closing	Opening	Continuation	
SPAAC			Initiate: release issue	Topic	
MALTUS				Topic change	
Primula	Opening	Closing	Topic opening	Topic closing/change	
Chiba	Opening	Closing	Topic break		
C-Star		Closing	Introduce topic		

Table 7 — *Discourse Structuring communicative functions in different dialogue act taxonomies.*

	Own Communication Management			Partner Communication Management	
DIT	Error signalling	Retraction	Self-correction	Correct-misspeaking	Completion
LIRICS	Error signalling	Self-correction		Correct-misspeaking	Completion
DAMSL		Speech repair		Correct-misspeaking	Completion
SWBD-DAMSL		Speech repair		Correct-misspeaking	Completion
MRDA		Speech repair		Correct-misspeaking	Collaborative completion
Coconut		Correct assumption; Speech repair		Correct-misspeaking	Completion
SLSA	Change				
TRAINS		Repair			
SPAAC		Correct-self		Correct	Complete
MALTUS		Restated info with repetition/correction		Restated info with correction	

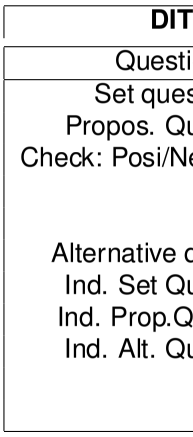
Table 8 — *Own and Partner Communication Management communicative functions in different dialogue act taxonomies.*

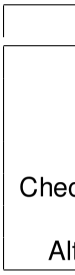
	Time Management		Contact Management	
DIT	Stalling	Pausing	Contact check	Contact indication
LIRICS	Stalling	Pausing	Contact check	Contact indication
DAMSL	Communication management: delay		Communication channel	
SWBD-DAMSL	Stalling; delay; Hold before answers		Communication channel	
MRDA	Hold before answers			
Coconut	delay		Communication channel	
AMI	Stall			
Verbmobil	Deliberate		Refer-to-settings	
SLSA	Choice			
TRAINS	Keep			
SPAAC	Hold			
Alparon		Pause		
C-Star		Please wait		

Table 9 — *Time and Contact Management communicative functions in different dialogue act taxonomies.*

DIT
Inform Uncertain I
Agreeme
-
-
Disagreen
-
-
Correcti
Inform Elab
Answe (Uncertain) Se Uncertain Prop Unc. con Unc. disco Prop. ans   Confirm Disconfi

Uncert In
Agre
Disag
Cor
Inform
Ar (Unc Set (Unc Prop. (Unc Co (Unc Disc





DIT
Offer
Promise
Address Req
Accept Req
Decline Req
Address Sugg
Accept Sugge
Decline Sugg
Other commis
Ind. Reque Instruct
Address O
Accept Of
Decline Of
Suggestio
Other direct

Annex G
(informative)
Editors, contributors and meetings

G.1 Editors and contributors

This document has been produced by Harry Bunt with support from members of the editorial project group and the external consultancy group, making use of documents from the LIRICS project, and of presentations and discussions at ISO TC 37/SC 4 meetings and joint workshops with the ACL-SIGSEM Working Group on the Representation of Multimodal Semantic Information.

The editorial project group for this project consists of: Jan Alexandersson, Harry Bunt, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, Claudia Soria, David Traum,

Co-authors of background LIRICS documents and publications include Amanda Schiffrin, Jeroen Geertzen, Volha Petukhova and Laurent Romary; LIRICS participants in discussions include Claudia Soria, Nicoletta Calzolari, Tomaso Casselli, Monica Monachini, Valeria Quocchi, Anna Joan Casademont, Koiti Hasida, Thierry Declerck and Gil Francopoulo. Through participation in joint TDG 3/SIGSEM WG meetings many other people have directly or indirectly contributed to this work, including Jens Allwood, Jae-Woo Choe, Chu-Ren Huang, Nancy Ide, Dafydd Gibbons, Jerry Hobbs, Simon Keizer, Staffan Larsson, Andrei Popescu-Belis, James Pustejovsky, David Traum, and Thorsten Trippel.

As of December 1, 2008, an 'Expert Consulting Group' has been formed for the project, consisting of James Allen, Jens Allwood, Thierry Declerk, Raquel Fernández, Gil Francopoulo, Julia Hirschberg, Staffan Larsson, Kiyong Lee, Oliver Lemon, Paul Mc Kevitt, Michael McTear, David Novick, Tim Paek, Massimo Poesio, Laurent Romary, Ielka van der Sluis, Pavel Smrz, Yorick Wilks, Aesun Yoon.

G.2 Meetings and Workshops

ISO TC 37/SC 4/TDG 3 had its inaugural meeting in Lisbon, Portugal on May 24-25, 2004. The following meetings have taken place since:

- January 10-11, 2005 in Tilburg, The Netherlands.
  ISA-1: Joint workshop with the ACL SIGSEM Working Group on the Representation of Multimodal Semantic Information, in conjunction with the 6th International Conference on Computational Semantics (IWCS-6).
- September 22-23, 2005, at INRIA-LORIA in Nancy, France during the Nancy Inference Week.
  ISA-2Joint workshop with the ACL SIGSEM Working Group on the Representation of Multimodal Semantic Information.
- January 20-22, 2006 at Jeju island, Korea (as part of TC 37/SC 4 meeting).
- April 20-22, 2006, Marina del Rey, California, Institute for Information Sciences (ISI).
  Invitation-only joint workshop with the ACL SIGSEM Working Group on the Representation of Multimodal Semantic Information.
- August 22-24, 2006, Beijing (at ISO TC 37 annual meeting).
- October 26-28, 2006, Brandeis University, Boston, USA.
  Meeting to start the work on temporal annotation in the form of ISO TC 37/SC 4 project "Semantic Annotation Framework (SemAF) Part 1, Time and Events".
- January 8-9, 2007, Tilburg, The Netherlands.
  ISA-3: Joint workshop with the ACL SIGSEM Working Group on the Representation of Multimodal Semantic Information in conjunction with the 7th International Conference on Computational Semantics (IWCS-7).
- May 7-9, 2007, Paris, France, at AFNOR, in conjunction with a meeting of the LIRICS project.
- January 12-13, 2008, Hong Kong, City University (in conjunction with ISO TC 37/SC 4 meeting and ICGL conference).

- September 29 - October 1, 2008, Pisa, Italy, CNR Istituto di Linguistica Computazionale.
  First meeting of the project editorial group.
- January 5-6, 2009, Tilburg, The Netherlands.
  ISA-4: 4th International Workshop in Interoperable Semantic Annotation. Joint workshop with the ACL SIGSEM Working Group on the Representation of Multimodal Semantic Information, in conjunction with the 8th International Conference on Computational Semantics (IWCS-8), January 7-9.
- May 29-31, 2009, Boulder, Colorado, USA
  ISO TC 37/SC 4/WG 2 meeting, including meeting of the project editorial group and External Consultancy Group.
- September 15-16, 2009, Tilburg, The Netherlands.
  ISO TC 37/SC 4/WG 2 meeting, including meeting of the project editorial group and External Consultancy Group.

Bibliography

[1] Ahrenberg, L., N. Dahlback & A. Jonsson (1995) Codings schemas for Studies of Natural Language Dialogue. In: *Working Notes from the AAAI Spring Symposium*, Stanford.

[2] Allen, J. & M. Core (1997) DAMSL: Dialogue Act Markup in Several Layers (Draft 2.1). Technical Report, Multiparty Discourse Group, Discourse Resource Initiative, September/October 1997. http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/

[3] Allen, J. and R. Perrault (1979) Plans, Inferences, and Indirect Speech Acts. In *Proceedings of the 17th Annual Meeting of the ACL*.

[4] Alexandersson, J., B. Buschbeck-Wolf, T. Fujinami, M. Kipp, S. Koch, E. Maier, N. Reithinger, B. Schmitz & M. Siegel (1998). *Dialogue acts in VERBMOBIL-2 (second edition)*. Verbmobil Report 226. Saarbrücken: DFKI.

[5] Allwood J. (1997) A critical look at speech act theory. In: O. Dahl (ed.) *Logic, Pragmatics, and Grammar*.

[6] Allwood, J. (2000) An activity-based approach to pragmatics. In: Harry Bunt and William Black, eds. *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics.* John Benjamins, Amsterdam, pp. 47–80.

[7] Allwood, J., J. Nivre & E. Ahlsen (1994) Semantics and Spoken Language Manual for Coding Interaction Management. Report from the HSFR project Semantik och talsprak.

[8] Appelt, D. (1982) Planning Natural Language Utterances to Satisfy Multiple Goals. SRI International AI Center Technical Note 259.

[9] Austin, J.L. (1962) *How to do things with words.* Clarendon Press, Oxford, UK..

[10] Bales, R.F. (1951 *Interaction process analysis: a method for the study of small groups.* Addison-Wesley, Cambridge.

[11] Bunt, H. (1995) Dynamic Interpretation and Dialogue Theory. In M.Taylor, D. Bouwhuis & F. Neel (eds.) *The Structure of Multimodal Dialogue, Vol. 2*. Amsterdam: Benjamins, 139-166.

[12] Bunt, H. (2000) Dialogue pragmatics and context specification. In Harry Bunt & William Black (eds.) *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics.* Amsterdam: Benjamins, 81-150.

[13] Bunt, H. (2004). A Framework for Dialogue Act Specification. Joint Workshop of ACL SIGSEM Working Group on the Representation of Semantic Information and ISO TC 37/SC 4 Thematic Domain Group 3: Semantic Content. Tilburg, January 2005. http://sigsem.uvt.nl/wg/meeting4.fdas.pdf

Bunt, H. (2005) A Framework for Dialogue Act Specification. 4th Joint ISO-SIGSEM Workshop on the Representation of Multimodal Semantic Information, Tilburg, January 2005. http://let.uvt.nl/research/ti/sigsem/wg/discussionnotes4.htm

[14] Bunt, H. (2006). Dimensions in Dialogue Act Annotation. In Proceedings 5^{th} International Conference on Language Resources and Systems (LREC 2006), Genova, May 2006, pp. 919-924..

[15] Bunt, H. (2007a) Multifunctionality and multidimensional dialogue act annotation. In: E. Ahlsén et al., editors, *Communication - Action - Meaning*. Gothenburg University, pp.237-259.

[16] Bunt, H. (2007b) The Semantics of Semantic Annotation. In *Proceedings of the 21^{st} Pacific Asia Conference on Language, Information and Computation (PACLIC-21)*, Seoul, November 2007, pp. 13-28.

[17] Bunt, H. (2009) The DIT^{++} taxonomy for functional dialogue markup. In D. Heylen, C. Pelachaud, R. Catizone and D. Traum (eds.) *Proceedings AMAAS 2009 Workshop "Towards a Standard Markup Language for Embodied Dialogue Acts'*, Budapest, May 2009.

[18] Bunt, H. (2009) *Multifunctionality and multidimensional dialogue semantics. Proceedings DiaHolmia, 2009 Workshop on the Semantics and Pragmatics of Dialogue'*, Stockholm, June 2009.

[19] Bunt, H. (*submitted*). Multifunctionality in dialogue.

[20] Bunt, H. & Y. Girard (2005) Designing an open, multidimensional dialogue act taxonomy. In C. Gardent & B. Gaiffe (eds.) *DIALOR'05, Proceedings of the 9th Workshop on the Semantics and Pragmatics of Dialogue*, Nancy, June 2005, 37-44.

[21] Bunt, H. and L. Romary (2002). Towards multimodal content representation. In *Proceedings LREC 2002 Workshop on International Standards of Terminology and Language Resources*, Las Palmas, May 2002. ELRA, Paris, pp. 54-60..

[22] Bunt, H. & L. Romary (2004) Standardization in Multimodal Content Representation: Some Methodological Issues. In *Proceedings 4th International Conference on Language Resources and Systems (LREC 2004)*, Lisbon, May 2004, pp. 2219-2222.

[23] Campbell, N. (2008) Individual Traits of Speaking Style and Speech Rhythm in a Spoken Discourse. In *Verbal and non-verbal features of human-human and human-machine interaction.* Revised papers from COST Action 2102 International Conference, Patras, Greece. Berlin: Springer.

[24] Carletta, J., A. Isard, S. Isard, J.Kowtko & G. Doherty-Sneddon (1996) HCRC dialogue structure coding manual. Technical Report HCRC/TR-82.

[25] Clark, H.H. (1996) *Using language.* Cambridge University Press, Cambrudge, UK.

[26] Clark, H., and Fox Tree, J. (2002). *Using 'uh' and 'um' in spontaneous speech.* Cognition, 84, pp. 73–111.

[27] Clark, H. and Krych, M.A. (2004). *Speaking while monitoring addressees for understanding*. Journal of Memory and Language, 50: 62-81.

[28] Clark, D. & A. Popescu-Belis, A. (2004) Multi-level Dialogue Act Tags. In *Proccedings of SIGDial 2004, 5th Workshop on Discourse and Dialogue*, Cambridge, MA, 163-171.

[29] Cooper, R. (2004) A type-theoretic approach to information state update in an issue-based approach to dialogue management. In *Proceedings CATALOG workshop on the semantics and pragmatics of dialogue*, Barceloma.

[30] Core, M., M. Ishizaki, J. Moore, C. Nakatani, N. Reithinger, D. Traum & S. Tutiya (1998) The Report of The Third Workshop of the Discourse Resource Initiative, Chiba University, Chiba, Japan. hhttp://cogsci.l.chiba-u.ac.jp/$\sim$tutiya/Publications/DRI98Report.pdf}%http://cogsci.l.chiba-u.ac.jp/$\sim$tutiya/Publications/DRI98Report.pdf\bibitem{}Core\&Allen1997) Coding Dialogs with the DAMSL Annotation schema. *AAAI Fall Symposium on Communicative Action in Humans and Machines*, Boston, MA.

[31] Di Eugenio, B., P.W. Jordan and L.Pylkkänen (1998) The COCONUT project: dialogue annotation manual. ISP Technical Report 98-1, December 1998, University of Pittsburgh.

[32] Duncan, S. and Fiske, D. (1977). *Face-to-face interaction: research, methods, and theory*. Lawrence Erlbaum Associates, New Jersey.

[33] Geertzen, J. (2009) *The recognition and prediction of dialogue acts.* PhD Thesis, Tilburg University.

[34] Geertzen, J. & H. Bunt (2006) Measuring annotator agreement in a complex, hierarchical dialogue act schema. In *Proceedings SIGDIAL 2006, 7th Workshop on Discourse and Dialogue*, Sydney.

[35] Geertzen, J.; V. Petukhova, and H. Bunt (2007), *A multidimensional approach to utterance segmentation and dialogue act classification.* In *Proceedings of SIGDIAL 2007, the 8th Workshop on Discourse and Dialogue*, Antwerp, September 2007, pp. 140–149.

[36] Goffman, E. (1981) Footing. In *Forms of talk.* University of Pennsylvania Press, Philadelphia.

[37] Goldman-Eisler, F. (1968). *Psycholinguistics: experiments in spontaneous speech*. New York, Academic Press.

[38] Grice, H.P. (1975) Logic and conversation. In J. Cole and P. Morgan (eds.) *Syntax and Semantics, Vol 3*. New York: Academic Press, pp. 183-198.

[39] Jekat, S., A. Klein, E. Maier, I. Maleck, M. Mast & J.J. Quantz (1995) Dialogue acts in Verbmobil. Verbmobil Report 65, April 1995. DFKI, Saarbrücken.

[40] Jurafsky, D., E. Schriberg, and D. Biasca (1997) Switchboard SWBD-DAMSL Shall-Discourse-Function Annotation Coders Manual. http://stripe.colorado.edu/~jurafsky/manual.august.html

[41] Larsson, S. (1998) Coding schemas for dialog moves. Unpublished paper, available at http://www.ling.gu.se/sl

[42] Levinson, S. (1983) *Pragmatics.* Cambridge University Press, Cambrudge, UK.

[43] LIRICS (2006a) LIRICS deliverable D4.2, Preliminary set of semantic data categories. Available from `http://lirics.loria.fr`. http://lirics.loria.fr

[44] IRICS (2006b) "Documented compilation of semantic data categories". LIRICS Deliverable D4.3. `\http://let.uvt.nl/research/ti/iso-tdg3/D4-3.pdf`

[45] IRICS (2007) "Multilingual test suites for semantically annotated data", LIRICS Deliverable D4.4. `http://let.uvt.nl/research/ti/iso-tdg3/D4-4.pdf`

[46] Petukhova, V. (2005) *Multidimensional interaction of multimodal dialogue acts in meetings."* MA thesis,Tilburg University.

[47] Petukhova, V. and H. Bunt (2009a) Dimensions of communication. Technical Report TR 2009-002, Tilburg center for Creative Computing, Tilburg University.

[48] Petukhova, V. and H. Bunt (2009b) The independenc of dimensions in multidimensional dialogue act annotation. *Proceedings 2009 NAACL-HLT Conference*, Boulder, Colorado, June 2009.

[49] Petukhova, V. and H. Bunt (2009c) Who's next?.Speaker-selection mechanisms in multiparty dialogue. *Proceedings of DiaHolmia, 2009 Workshop on the Semantics and Pragmatics of Dialogue*, Stockholm, June 2009.

[50] Poesio, M. and D. Traum (1998) Towards an axiomatization of dialogue acts. In *Proccedings of the Twnete Workshop on the semantics and pragmatics of dialogue,* Enschede.

[51] Popescu-Belis, A. (2005) Dialogue Acts: One or More Dimensions? ISSCO Working Paper 62, ISSCO, Geneva. November 2005.
http://www.issco.unige.ch/publicaitons/working-papers/papers/apb-issco-wp62b.pdf

[52] Sacks, H., E. Schegloff & G. Jefferson (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 53:696–735.

[53] Searle, J.R. (1969) *Speech Acts.* Cambridge, UK: Cambridge University Press.

[54] Smith, V.L., and Clark, H. (1993). *On the course of answering questions*. Journal of Memory and Language, 32, pp. 25–38.
bibitem Soria, C. & V. Pirrelli (2003) A multi-level annotation meta-schema for dialogue acts. In A. Zampolli, N. Calzolari & L. Cignoni (eds.) *Computational Linguistics in Pisa. Linguistica Computazionale, Special Issue, XVIII-XIX*, Pisa-Roma: IEPI, 865-900.

[55] Stenström, A.-B. (1990). *Pauses in monologue and dialogue*. In: Svartvik, J. (ed.). London-Lund Corpus of Spoken English: Description and Research. Lund, Lund University Press.

[56] Swerts, M., and Ostendorf, M. (1997).*Prosodic and lexical indications of discourse structure in human-machine interactions*. Speech Communication, Vol. 22 , pp. 25–41.

[57] Traum, D. (1999) Speech Acts for Dialogue Agents. In: M. Wooldridge & A. Rao (eds.) *Foundations of rational agency*, Kluwer, Dordrecht, 169–201.

[58] Traum, D. (2000) 20 Questions on Dialogue Act Taxonomies. *Journal of Semantics* 17(1): 7–30.

[59] Traum, D. & E. Hinkelman (1992) Conversation acts in task-oriented spoken dialogue. *Computational Intelligence* 3 (8): 575-599.

[60] Traum, D. & S. Larsson (2003) The Information State Approach to Dialogue Act Management. In J. van Kuppevelt & R. Smith (edr.) *Current and new Directions in Discourse and Dialogue*, Kluwer, Dordrecht.