# On the many benefits of a multidimensional approach to the analysis of spoken dialogue

**Harry Bunt and Volha Petukhova**
Tilburg Center for Creative Computing
Tilburg, the Netherlands
harry.bunt@uvt.nl; v.petukhova@uvt.nl

**Abstract**

In this paper we argue that the study of spoken dialogue can benefit in many ways from the use of a multidimensional, action-based analysis framework. We outline such a framework and show the benefits of its application to (1) the segmentation of spoken dialogue into functional units; (2) the analysis and explanation of the phenomenon that dialogue utterances often have more than one communicative function; (3) the annotation of spoken and multimodal dialogue with dialogue act information; (4) the semantic analysis of discourse markers; and (5) the generation of deliberately multifunctional utterances by a dialogue system.

## 1 Introduction

Participating in a dialogue is a complicated business. Not only do you want to achieve something, such as obtaining certain information, gaining support, giving advice, or improving relationships, but you also have to pay attention to whether you're being understood correctly, to whether you understand well what your partners are saying and doing, to whether you're expressing yourself well, to not speaking simultaneously with others all the time, and so on. Lots of expressions which are very common in dialogue reflect this; some examples are:

(1) a. OK?
b. Before four, you said?
c. Could you repeat that please.
d. ... to Canada ... to Toronto, Canada
e ... let me see...

In utterance (1a), the speaker elicits feedback about the successful processing of his previous utterance(s); in (1b) and (1c) the speaker performs negative feedback acts, signaling a problem in his processing of what was said before, in the form of a check question (1b) and a request (1c), respectively. In (1d), the speaker performs a self-correction, in order to express himself more accurately, and in (1e) the speaker stalls for time and indicates that he is keeping the turn.

An obvious requirement for successful participation in dialogue is the ability to understand what the dialogue partners are saying and doing. The characterizations that we just gave of the utterances in (1) are all descriptions of how the linguistic behaviour can be interpreted in terms of *actions*, performed by the speaker. Attempts to give an account of the meanings of dialogue utterances by applying traditional approaches to sentence meaning fail rather spectacularly because they are rooted in the truth-conditional view of meaning, while dialogue utterances like those in (1), but also many others like *Good morning*, *Excuse me*, *Quite so*, and *Thank you* have meanings that cannot be captured in terms of the truth or falsity of propositions. The meanings of dialogue utterances have, besides a propositional component, an equally important *functional* component. This view is reflected in *information-state update* (ISU), also called *context-change* (CC) approaches to utterance meaning (Bunt, 2000; Traum & Larsson, 2003), which formalize utterance meanings in terms of changes in the 'information states' (or 'context models') of the dialogue participants. These approaches closely relate to speech act theory, which regards the use of language as the performance of communicative actions that have an illocutionary function and a propositional content.

A complication that these approaches have to face is that, contrary to what speech act theory tells us, dialogue utterances often have multiple communicative functions, such as answering a question but also providing feedback on the understanding of the question, and also taking the turn. The following example illustrates this.

(2)
1. A: What time is the next train to Amersfoort?
2. B: Let me see.... That will be at 11:25.
3. A: Is there no train to Amersfoort before 11:25?
4. B: Amersfoort? I'm sorry, I thought you said Apeldoorn.

Utterance 3 shows that A assumes that B understood the question 1, when he answered it in 2. He did not question B's understanding of the question, even though the answer surprised him.

The first part of B's utterance 2 is also worth considering: why does B stall for time by saying *Let me see....*? This is because he needs a bit of time to find the information that A asked for, but then why doesn't he just wait until he has found that information before starting to speak? This must be because he has decided to take the turn, so the utterance *Let me see* in fact has two functions: B signals that (1) he takes the turn; and (2) that he needs a bit of time to formulate his contribution (the answer to A's question).[1]

This example illustrates that dialogue utterances often do not correspond to a single speech act, but to sets of speech acts. Moreover, some of these speech act types, such as feedback acts and turn-taking acts have hardly been studied in speech act theory, and do not easily fit within that theory. Approaches to dialogue semantics in terms of updating models of information states or dialogue contexts have therefore in fact not related closely to speech act theory, but rather to modern, data-driven versions of 'dialogue act' theory. In Section 2 we outline the multidimensional dialogue act-theoretical framework that we will apply in the rest of this paper.

The analysis of spoken dialogue in terms of communicative actions requires a way to identify those stretches of speech that correspond to the performance of one or more such acts. The identification of such stretches is often called the *segmentation* of the dialogue. In Section 3 we will discuss the segmentation of dialogue into functional units, which is a prerequisite for analysing the functions of parts of a dialogue. We will argue that a multifunctional approach naturally needs to abandoning the idea that segmentation should aim at cutting up a dialogue into a linear sequence of stretches of speech, and that one should allow functional units to overlap, to be discontinuous, to include other functional segments, and to spread over multiple turns, leading to a more accurate form of segementation that other approaches.

The rest of this paper is organized as follows. After the outline of our approach in Section 2, and the discussion of dialogue segmentation in Section 3, we discuss in subsequent sections the following claimed benefits of a multidimensional approach to dialogue analysis.

In Section 4 we will show that this approach is beneficial for obtaining a good qualitative and quantitative insight into the nature and the pervasiveness of the phenomenon of the multifunctionality of dialogue utterances, which has been one of the main motivations behind multidimensional approaches.

In Section 5 we turn to the potential advantages of multidimensional annotation (over one-dimensional approaches), arguing that these potential advantages are only realised fully if the notion of a dimension is very carefully defined, and multidimensional annotation schemas are equally carefully designed. We also pay some attention to the relation between the work reported here and an ongoing effort by the International Organisation for Standardisation ISO to establish an international standard for dialogue act annotation.

[1]This is common for a turn-initial stalling act. A turn-*internal* stalling act, by contrast, usually has a turn-*keeping* rather than a turn-*taking* function, as in *That will be... let me see... at 11:25.*

In Section 6 we describe a successful application of the proposed multidimensional approach to the semantic analysis of discourse markers.

Section 7 is more speculative than the preceding sections, discussing the potential benefits of a multidimensional approach to the generation of utterances by spoken dialogue systems. It is argued that a multidimensional approach opens the perspective of developing dialogue act generators that produce utterances which are multifunctional by design, rather than by accident, as in current systems. Section 8 ends with some general conclusions.

## 2 A multidimensional, action-based approach to dialogue utterance meaning

### 2.1 Dialogue acts and utterance meanings

The semantic framework of Dynamic Interpretation Theory (DIT, see Bunt, 2000; 2009) ) takes a multidimensional view on dialogue in the sense that participation in a dialogue is viewed as performing several activities in parallel, such as pursuing the task or activity that motivates the dialogue, providing and eliciting communicative feedback, taking turns, managing the use of time; and taking care of social obligations. The activities in these various dimensions are called *dialogue acts* and are formally interpreted as update operations on the information states (or 'context models')[2]; of the dialogue participants. Dialogue acts have two main components: a *semantic content* which is to be inserted into, to be extracted from, or to be checked against the current information state; and a *communicative function*, which specifies more precisely how an addressee updates his information state with the semantic content when he understands the corresponding aspect of the meaning of a dialogue utterance.

DIT distinguishes the following 10 dimensions (for discussion and justification see Petukhova & Bunt 2009a; 2009b):

1. Task/Activity: dialogue acts whose performance contributes to performing the task or activity underlying the dialogue;
2. Auto-Feedback: dialogue acts that provide information about the speaker's processing of the previous utterance(s);
3. Allo-Feedback: dialogue acts used by the speaker to express opinions about the addressee's processing of the previous utterance(s), or that solicit information about that processing;
4. Contact Management: dialogue acts for establishing and maintaining contact;
5. Turn Management: dialogue acts concerned with grabbing, keeping, giving, or accepting the sender role;
6. Time Management: dialogue acts signalling that the speaker needs a little time to formulate his contribution to the dialogue;
7. Discourse Structuring: dialogue acts for explicitly structuring the conversation, e.g. announcing the next dialogue act, or proposing a change of topic;
8. Own Communication Management: dialogue acts where the speaker edits the contribution to the dialogue that he is currently producing;
9. Partner Communication Management: the agent who performs these dialogue acts does not have the speaker role, and assists or corrects the speaker in formulating a contribution to the dialogue;
10. Social Obligations Management: dialogue acts that take care of social conventions such as greetings, apologies, thanking, and saying goodbye.

One of the products of DIT is a multidimensional taxonomy of communicative functions, called the DIT$^{++}$ taxonomy[3], which was designed for the purposes of dialogue act annotation and dialogue system design across a wide range of domains, and which includes elements from various other annotation

[2]In the rest of this paper, we will use the terms 'information state', and 'context' (or 'context model') interchangeably, as also the terms 'information state update, 'context change' and 'context model update'.

[3]See `http://dit.uvt.nl`

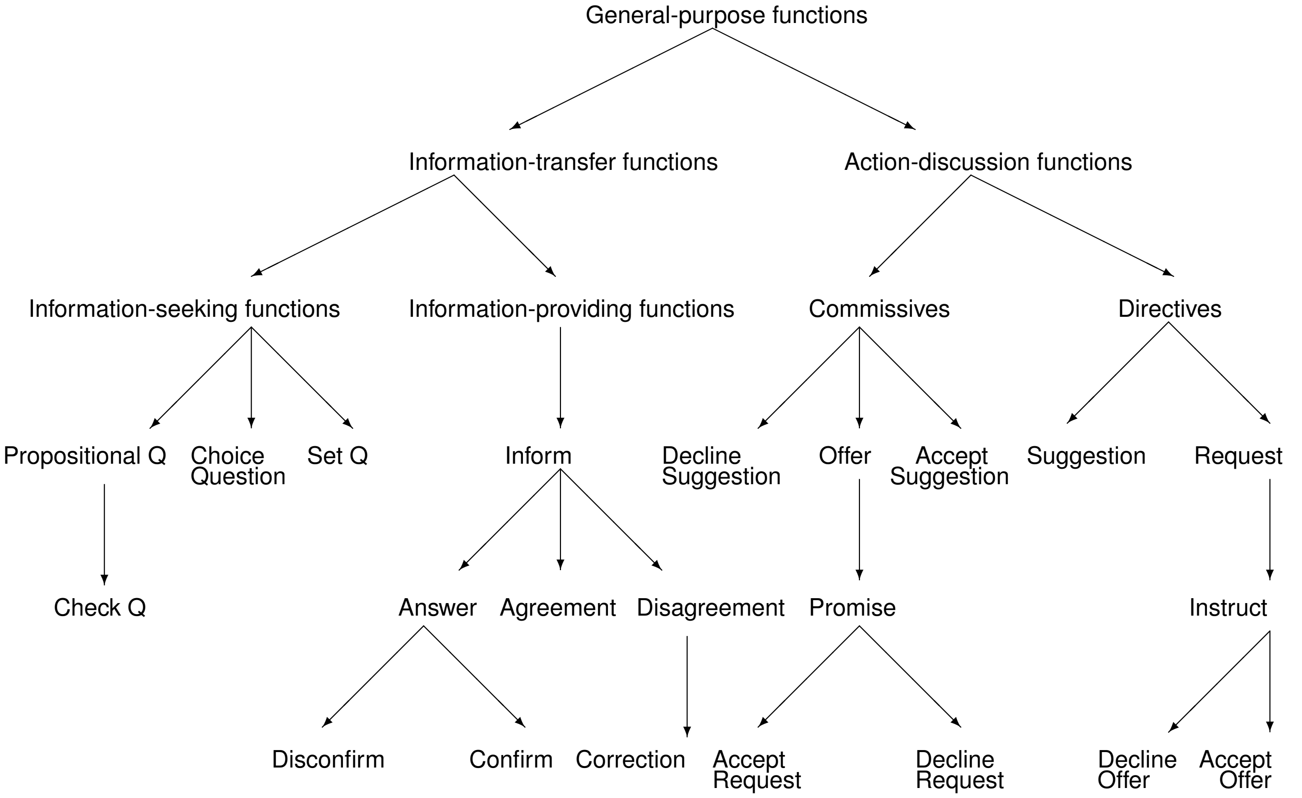


Figure 1: General-purpose functions

schemas, such as the DAMSL, TRAINS, and Verbmobil taxonomies (Allen & Core, 1997; Allen et al., 1994; Alexandersson et al., 1998).

Multidimensional taxonomies support dialogue utterances to be coded with multiple tags and have a relatively large tag set; such a tag set may benefit in several respects from having some internal structure. First, clustering semantically related tags improves the transparency of the tag set for human users, as the clusters indicate the kind of semantic information that is addressed. Second, introducing a hierarchical or taxonomical structure which is based on semantic clustering may support the decision-making process of human annotators: an initial step in such a process can be the decision to consider a particular cluster, and subsequently more fine-grained distinctions may be tested in order to decide on a specific tag within the cluster. Third, a hierarchical organisation in the tag set may also be advantageous for automatic annotation and for achieving annotations which are compatible though not identical with those of human annotators (namely, the automatic annotation may use less specific tags than the human annotation). In general, 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, interpretation, and generation.

Bunt (2005; 2006) suggests that the structure of a multidimensional annotation schema should be based not just on a clustering of intuitively similar functions, but on a well-founded notion of *dimension*, and proposes to define a *set of dimensions* as follows.

(3) Each member of a set of dimensions is a cluster of communicative functions which all address a certain aspect of participating in dialogue, such that:

1. dialogue participants can address this aspect through linguistic and/or nonverbal behaviour which has this specific purpose;
2. this aspect of participating in a dialogue can be addressed independently of the aspects corresponding to other members of the set of dimensions, i.e., an utterance can have a communica-

| *Dimension* | *Dimension-specific functions* | *Representative expressions* |
|---|---|---|
| Task/Activity | OpenMeeting, CloseMeeting; Appoint, Hire, Fire | domain-specific fixed expressions |
| Auto-Feedback | PerceptionNegative | *Huh?* |
| | EvaluationPositive | *True.* |
| | OverallPositive | *OK.* |
| Allo-Feedback | InterpretationNegative | *THIS Thursday.* |
| | EvaluationElicitation | *OK?* |
| Turn Management | TurnKeeping | final intonational rise |
| | TurnGrabbing | hold gesture with hand |
| | TurhGiving | *Yes.* |
| Time Management | Stalling | slowing down speech; fillers |
| Contact Management | ContactChecking | *Hello?* |
| Own Communication Man. | SelfCorrection | *I mean...* |
| Partner Communication Man. | PartnerCompletion | completion of partner utterance |
| Discourse Structure Man.t | DialogueActAnnouncement | *Question.* |
| | TopicShiftAnnouncement | *Something else.* |
| Social Obligations Man. | Apology | *I'm sorry.* |
| | Greeting | *Hello!, Good morning.* |
| | Thanking | *Thanks.* |

Table 1: Examples of dimension-specific communicative functions and representative expressions for each dimension.

tive function in one dimension, independent of its functions in other dimensions.

The first condition means that only aspects of communication are considered that are observed in actual communicative behaviour; the second that dimensions should be independent. A set of dimensions that satisfies these requirements can be useful for structuring an annotation schema, especially if the set of functions within each dimension is defined in such a way that any two functions are either mutually exclusive or have an entailment relation. In that case a functional unit can be annotated with (maximally) as many tags as there are dimensions, one function at most (namely the most specific function for which there is evidence that it should be marked) for each dimension.

When we view a dimension in dialogue analysis in accordance with (3) as a particular aspect of interacting, like the 10 dimensions mentioned above, we see that dialogue acts like question and answer do not belong to any particular dimension. This is because one can ask a question about something concerning the task, or about agreeing to close a topic, or about whose turn it is to say something, or about any other aspect of interacting, so questions can belong to *all* these dimensions. Every *occurrence* of a question function, as the function of a dialogue act that is performed, falls within one of the dimensions; which dimension is determined by the type of semantic content.

Similarly for answers, statements, requests, offers, agreements, (dis-)confirmation, and so on. Clusters of such general types of dialogue acts therefore do not form a dimension, but can be used in any dimension; they are called *general-purpose functions*. This in contrast with communicative functions that are specific for a particular dimension, such as Turn Keep and Turn Release; Introduce Topic and Change Topic; Apology and Thanking. The DIT$^{++}$ taxonomy therefore consists of two parts: (1) a taxonomy of *general-purpose functions*; (2) a taxonomy of *dimension-specific functions*. Figure 1 shows the structure of the taxonomy of general-purpose functions; table 3 lists examples of dimension-specific communicative functions in each of the DIT$^{++}$ dimensions.

In order to define a context-change semantics for all the types of dialogue acts in the DIT$^{++}$ taxonomy, the context models on which the semantics is based should contain all the types of information addressed by these dialogue acts. Table 3 lists these types, and illustrates their use by dialogue utterances whose update semantics involves these types of information.

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 functions may be regarded as optional refinements of the taxonomy of core

| *example utterance* | *dialogue act type* | *information category* |
|---|---|---|
| *Can I change the contrast now?* | Task-related propositional question | task information |
| *Please press reset first* | Task-related request | task information |
| *Did you say Thursday?* | Feedback check question | own processing success |
| *Okay?* | Feedback elicitation | partner processing success |
| *Let me see,...* | Stalling | processing time estimates |
| *Just a minute* | Pause | processing time estimates |
| *Well,...* | Turn Accept | turn allocation |
| *Tom?* | Turn Assign | turn allocation |
| *Let's first discuss the agenda* | Dialogue structure suggestion | dialogue plan |
| *Can I help you?* | Dialogue structure offer | dialogue plan |
| *On june first I mean second* | Self-correction | own speech production |
| *.... you mean second* | Partner correction | partner speech production |
| *Hello?* | Contact check | presence and attention |
| *You're welcome* | Thanking downplayer | social pressure |

Table 2: Semantic information categories as related to dialogue act types, and example utterances.

general-purpose functions, which is depicted in Figure 1.

## 3 Dialogue segmentation

Analyses of spoken dialogue often segmented dialogues at the level of turns or utterances. A *turn* can be defined as a stretch of speech produced by one speaker, bounded by periods of inactivity of that speaker or by activity of another speaker.[4]. Since the 'turn' is not only used in this sense, but also in the functional sense of 'having the turn', i.e. occupying the speaker role, we will use the term *turn unit* instead of 'turn' in the former sense. An advantage of segmentation at turn level is that turn unit boundaries can be detected relatively easily. A disadvantage is that a turn unit often contains smaller parts, called 'utterances', that have separate communicative functions. Turns units are therefore too coarse-grained for many purposes as the units of dialogue analysis.

Utterances are linguistically defined contiguous stretches of (linguistic) behaviour, which have a certain grammatical well-formedness (like being a phrase or a clause), and are assumed to correspond to units that have a separate, well-defined meaning. Segmentation into utterances has the advantage of more fine-grained units being interpreted, allowing more precise analysis than a turn-based segmentation; however, the notion of an utterance as a smaller unit inside a turn unit does not have a 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).

Apart from the difficulties of automatic utterance boundary detection, from a semantic point of view the stretches of behaviour that are relevant for interpretation as dialogue acts may be discontinuous, may overlap, and may even contain parts from more than one turn unit. 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 *functional segment* as a "minimal" stretch of communicative behaviour that has a communicative function (and possibly more than one).

Example (4) shows that a functional segment may be discontinuous:

(4) 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 indicates 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:

[4]Cf. Allwood, 2000, on defining units of communication.

(5) 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 (6) shows that a dialogue act unit 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.

(6) 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, B: and then a Garoeda flight at 11:45,

The complications of discontinuity, overlap, and spreading over multiple turns can be handled by taking the multidimensional view on communication that was described above, 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 interpretation 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. This means that a given stretch of bahaviour is segmented 'multidimensionally', in as many ways as there are dimensions in which parts of it have a communicative function. 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. |
| (7) | 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 (7) the second turn unit is segmented is three dimensions: (1) Task/Activity; (2) Feedback; and (3) Time Management. In the Task 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 (a task-related answer). 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.

**Automatic multidimensional segmentation**

We investigated experimentally how well functional segment boundary can be detected automatically using multidimensional segmentation. For this purpose we used human-human multi-party interactions in English (AMI-meetings with 4 participants)[5]. The AMI corpus contains manually produced orthographic transcriptions for each individual speaker, including word-level timings that have been derived using a speech recogniser in forced alignment mode. The meetings are video-recorded and each dialogue is also provided with sound files (for our analysis we used recordings made with close-talking microphones to eliminate noise). Three scenario-based[6] meetings were selected to constitute a training set of 3,897 functional segment instances, 17335 words. Thus, the average segment length is 4.4 words.

[5] Augmented Multi-party Interaction (`http://www.amiproject.org/`).

[6] Meeting participants play different roles in a fictitious design team that takes a new project from kick-off to completion over the course of a day.

The data was annotated according to designed taxonomy of dialogue acts DIT$^{++}$. Table 3 gives an overview of the percentage of instances for the ten most frequently occurring functional tags in the training set.

| Tag | Percentage |
|---|---|
| Time;STALLING | 20.7 |
| Auto-FB;POS.OVERAL | 18.7 |
| Turn;Turn Keeping | 7.5 |
| Task;INFORM | 6.8 |
| Task;INFORM Elaborate | 3.5 |
| Task;INF.Agreement | 2.5 |
| Task;Propositional-Question | 2.3 |
| Task;SUGGEST | 2.0 |
| Task;INFORM Justify | 2.0 |
| Task;CHECK-Question | 1.6 |

Table 3: *Percentage of instances for most frequent functional tags in the AMI corpus.*

In the selected dialogues, the majority of the functional segments address the Task dimension (33%), followed by Auto-Feedback (20%); Time Management (16.8%); Turn Management (15); Own Communication Management (8.7%); Allo-Feedback (0.7%); Discourse Structuring (2.2%); Partner Communication Management (0.3%);Social Obligation Management (0.3%); and Contact Management (0.1%).

The features and their selection play a very important role in supporting accurate recognition and classification of functional segments and their computational modelling may be expected to contribute to improved automatic dialogue processing. The features included in the data sets are those relating to *dialogue history*, *prosody* and *word occurrence*. Dialogue history consists of the tags of the 10 preceding utterances. Additionally, the tags of utterances to which the utterance in focus was a direct response to are included as features, as well as timing. Note that some segments are located inside other segments, for instance in case of backchannels and interruptions that do not cause turn shifting; the occurrence of these events is encoded in a feature.

Prosodic features that are included are minimum, maximum, mean, and standard deviation of *pitch* (F0 in Hz), *energy* (RMS), *voicing* (fraction of locally unvoiced frames and number of voice breaks), speaking rate (number of syllables per second) and *duration* both of the current and preceding segments. We examined both raw and normalized versions of these features. Speaker normalization techniques modify the spectral representation of incoming speech waveforms in an attempt to reduce variability between speakers, which is partly due to differences in speakers' anatomy. Speaker-normalized features were normalized by computing z-scores (z = (X- *mean*)/*standard deviation*) for the feature, where *mean* and *standard deviation* were calculated from all functional segments produced by the same speaker in the dialogues. We also used normalizations by the first speaker turn and by prior speaker turn.

Word occurrence is represented by a bag-of-words vector[7] indicating the presence or absence of words in the segment. As lexical features bi- and trigram models are constructed. Additionally we indicated the speaker (A, B, C, D) and the addressee (other participants individually or the group as a whole).

To be able to identify the segment boundaries, we assign to each token its communicative function label as well as indicated whether a token is starting a segment (B), is inside a segment (I), is ending a segment (E), outside a segment (O), or forms a functional segment on its own (BE) (see Table 4).

**Classifier**

We used rule induction algorithm, we chose *Ripper* (Cohen, 1995). The results of all experiments were obtained using 10-fold cross-validation[8]. Table in 5 shows the success scores on automatic segment boundaries identification task:

[7] With a size of 1,640 entries.

[8] In order to reduce the effect of imbalances in the data, it is partitioned ten times. Each time a different 10% of the data is used as test set and the remaining 90% as training set. The procedure is repeated ten times so that in the end, every instance

| Speaker | Token | Task | Auto-F. | Allo-F. | TurnM. | TimeM. | ContactM. | DS | OCM | PCM | SOM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| B | it | B | O | O | O | O | O | O | O | O | O |
| B | has | I | O | O | O | O | O | O | O | O | O |
| B | to | I | O | O | O | O | O | O | O | O | O |
| B | look | I | O | O | O | O | O | O | O | O | O |
| B | you | O | O | B | O | O | O | O | O | O | O |
| B | know | O | O | E | O | O | O | O | O | O | O |
| B | cool | I | O | O | O | O | O | O | O | O | O |
| D | mmhmm | O | BE | O | O | O | O | O | O | O | O |
| B | and | I | O | O | BE | O | O | O | O | O | O |
| B | gimmicky | E | O | O | O | O | O | O | O | O | O |

Table 4: *Segment boundaries encoding per dimension*

| Features | Accuracy (in %) | Precision | Recall | F-scores |
|---|---|---|---|---|
| Prosody | 79.7 | 0.6 | 0.4 | 0.55 |
| Prosody + Wording | 81.2 | 0.7 | 0.49 | 0.54 |
| Prosody + Wording + Speaker switch | 85.8 | 0.73 | 0.62 | 0.64 |
| Best selected features | 86.2 | 0.78 | 0.64 | 0.69 |

Table 5: *Success scores of the RIPPER classifier on the segment boundaries classification task.*

Baseline for this classification task was 72.6 (percentage of instances of the most frequent class, here token inside functional segment).

From this table we can conclude that prosodic features are important role for the identification of segment boundaries. The rules induced by Ripper show that segments which contains only one token are the longest (460 ms) and tokens inside segments have the shortest duration (240 ms). It is also observed that begin-of-segment tokens are almost twice as short as end-of-segment tokens (209ms and 419ms respectively). Thus, the speaker used to lengthening the words at the end of the utterance and to slow down their speech tempo (9.9 syllables per second when start to speak and 5.7 when finishing the utterance).

Further, one-token segments are less voiced (28% unvoiced frames). This can be explained by the fact that significant amount of one-token segments are positive auto-feedbacks at the level of execution used as Backchannels, where the speaker shows his evaluation of the previous utterance in the most silent way without actually claiming the turn, and Stallings, which naturally contain many unvoiced frames, because the speaker is hesitating. Interesting is also the fact that the tokens which start the segment are not highly voiced as well (33% unvoiced frames). Tokens inside segments are the most voiced ones (24% unvoiced frames).

As for pitch features, it is observed that one-token segments have the highest mean pitch detected (164Hz) and the lowest pitch was measured for inside-segment tokens (144Hz). There is however no significant differences between one-token segments and begin-of-segment tokens, they are both quite high pitched tokens. The speaker-normalized mean pitch helps to differentiate between inside-segment tokens and the others. The standard deviation in pitch is of particular importantce to detect end-of-segment tokens, because this value was significantly higher for the end of the uuterance than for the tokens in other positions. This also applicable to speaker-normalized maximum pitch which is considerably higher at the end of the speaker utterance (206Hz comparing with 176Hz inside the segment).

Using automatic feature selection algorithm, which ranks features measuring their information gain with respect to the class identified, token and token combinations (bigram) were selected as best features. Discourse markers (see Section 6) are particularly good indicators of segment boundaries, because they

---

has been used exactly once for testing (Witten, 2000) and the scores are averaged. The cross-validation was stratified, i.e. the 10 folds contained approximately the same proportions of instances with relevant tags as in the entire dataset.

so to speak bracket segment indicating its start, e.g. *and*, *well*, *so*, etc., or its ending, e.g. *anyway*.

The main conclusion which can be drawn here is that segment boundaries are well detectable using only low-level features and these multiple information sources need to be combined in order to detect functional segment boundaries reliably.

## 4 Multifunctionality

### 4.1 Forms of multifuncitoanlity

Allwood (1992) distinguished two forms of multifunctionality, called *sequential* and *simultaneous*, using the following example:

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

Allwood observes that A's contribution in this dialogue fragment "contains sequentially the functions *feedback giving, request, request, statement and response elicitation*. Furthermore, the statement 'Bill will be there' could simultaneously be a promise and thus illustrates simultaneous multifunctionality."

Segmenting a dialogue into utterances in the sense introduced in the previous section, rather than into turn units has the effect of eliminating sequential multifunctionality. Besides the simultaneous multifunctionality distinguished by Allwood, there are however still other, segmentation-related forms of multifunctionality that remain, namely *discontinuous, overlapping*, and *interleaved multifunctionality*. The first of these occurs when an utterance has more than one communicative function due to being interrupted by a smaller utterance which has a different communicative function, as in example (4) above. Example (5) illustrates the phenomenon of overlapping multifunctionality, which occurs when part of an utterance with a certain function forms a sub-utterance with another function. *Interleaved* multifunctionality occurs when two utterances with different functions are interleaved to form a complex utterance, and is illustrated by the following example.

(9) I think twenty five euros for a remote... *is that locally something like fifteen pounds?...* is too much money to buy an extra remote or a replacement one .. *or is it even more in pounds?*

Here we see the discontinuous statement *I think twenty five euros for a remote [...] is too much money to buy an extra remote or a replacement one* interleaved with the discontinuous question *is that locally something like fifteen pounds [...] or is it even more in pounds?*

These examples show that the segmentation of dialogue into utterances in the usual sense of a linear sequence of units whose concatenation is the dialogue as a whole does not lead to distinguishing the stretches of behaviour that form functional units. Instead, such units should be allowed to be discontinuous, to overlap, and to be interleaved. To avoid terminological confusion, we introduced the term *functional segment* for this purpose.

### 4.2 Semantic types of multifunctionality

The multifunctionality of dialogue utterances not only takes several forms, as noted above, but also comes in semantically different varieties. The following four types can be distinguished:

**independent:** a functional segment has more than one communicative function, due to having features expressing each of these functions (when used in certain contexts);

**entailed:** a functional segment has two (or more) communicative functions because one function logically entails another;

**implicated:** a functional segment has two (or more) communicative functions because one function is conversationally implicated by another function;

**indirect:** the segment constitutes an indirect dialogue act, i.e. it has another communicative function than it would appear at first sight, which can be inferred from its 'literal' function in the context in which it occurs.

We discuss each of these types of multifunctionality in turn.

| *example utterance* | *dialogue act type* | *information category* |
|---|---|---|
| *Can I change the contrast now?* | Task-related propositional question | task information |
| *Please press reset first* | Task-related request | task information |
| *Did you say Thursday?* | Feedback check question | own processing success |
| *Okay?* | Feedback elicitation | partner processing success |
| *Let me see,...* | Stalling | processing time estimates |
| *Just a minute* | Pause | processing time estimates |
| *Well,...* | Turn Accept | turn allocation |
| *Tom?* | Turn Assign | turn allocation |
| *Let's first discuss the agenda* | Dialogue structure suggestion | dialogue plan |
| *Can I help you?* | Dialogue structure offer | dialogue plan |
| *On june first I mean second* | Self-correction | own speech production |
| *.... you mean second* | Partner correction | partner speech production |
| *Hello?* | Contact check | presence and attention |
| *You're welcome* | Thanking downplayer | social pressure |

Table 6: Semantic information categories as related to dialogue act types, and example utterances.

#### 4.2.1 Independent multifunctionality

A functional segment may have several independent communicative functions, in different dimensions. Examples are:

1. ”Thank you”, spoken with markedly high pitch and cheerful intonation (like goodbyes often have), to signal goodbye in addition to gratitude;
2. “Yes”, said with in intonation that first falls and subsequently rises, expressing postive feedback (successful understanding etc.) and giving the turn back to the previous speaker;
3. Turn-initial Stalling and Turn Take (or Turn Accept);
4. Excessive turn-internal Stalling and elicitation of support (i.e., eliciting an utterance completion act in the Partner Communication Management dimension).

Semantically, the interpretation of an utterance which displays independent multifunctionality comes down to two (or more) independent update operations on different dimensions of an addressee's information state, one for each communicative function.

#### 4.2.2 Entailed functions

In the case of an entailment relation, a functional segment has a communicative function, $F_1$ expressed by utterance features, which is characterized by a set of preconditions which logically imply those of a dialogue act with the communicative function $F_2$.

Some examples of entailment relations between dialogue acts are:

1. Justification, Exemplification, Warning all entailing Inform; Agreement, Disagreement, Correction entailing Inform; Confirmation and Disconfirmation both entailing Propositional Answer; Check Question entailing Propositional Question;
2. Answer, Accept Offer, Reject Offer, Accept Suggestion, Reject Suggestion entailing positive feedback;
3. Responsive dialogue acts for social obligations management, such as Return Greeting and Accept Apology entailing positive feedback on the corresponding initiating acts;
4. Positive evaluative feedback entailing positive feedback on perception and understanding;
5. Negative feedback on perception entailing negative feedback on understanding (see below, Section 4.2.4).

An imortant case of entailed functions concerns communicative functions that differ in their level of specificity: more specific dialogue acts entail less specific ones with the same semantic content. Entailed functions used within the same dimension correspond to the context update operation representing the entailed interpretation being subsumed by the update operation of the entailing one. They are thus semantically vacuous, and it therefore does not seem to make much sense to consider such cases as multiple functions that can be assigned to a functional segment.

Entailments frequently occur between an act in a non-feedback dimension and a feedback act. An answer, for example, is semantically related to a preceding question. Relations such as the one between an answer and the corresponding question are called *functional dependency relations*[9], This type of relation is relevant for answers, responses to directive dialogue acts (such as Accept Request and Reject Offer), and more generally to those dialogue acts that have a 'backward-looking function' (Allwood, 2000; Allen & Core, 1997). The fact that a speaker responds to a previous dialogue act implies that the speaker has (or at least believes to have) successfully processed the utterance(s) expressing the previous dialogue act, and so the occurrence of a responsive dialogue act entails a positive feedback act.

### 4.2.3 Implicated functions

Implicated multifunctionality occurs when a functional segment has a certain communicative function by virtue of its observable features (in the given dialogue context), and also another communicative function due to the occurrence of a conversationally implicature. Like all conversational implicatures, this phenomenon is context-dependent, and the implicatures are intended to be understood by the addressee. Examples are:

1. an expression of thanks implicating positive feedback at all levels of processing the previous utterance(s) of the addressee;
2. positive feedback implied by shifting to a new topic, related to the previous one; more generally, by any relevant continuation of the dialogue;
3. negative feedback, implied by shifting to an unrelated topic; more generally, by any 'irrelevant' continuation of the dialogue.

Implicated functions are not expressed explicitly through the features of expressions, but can be inferred as being likely from the interpretation of the utterance features (as indicating a type of certain dialogue act) in a given context. Implicated functions correspond semantically to an additional context update operation, hence they are a true source of multifunctionality.

### 4.2.4 Entailed and implicated feedback functions

A speaker who provides feedback about his perception, understanding, or evaluation of previous utterances. may be specific about the level of processing that his feedback refers to. For instance, a literal repetition of what was said with a questioning intonation is typically a signal that the speaker is not sure he heard well, whereas a rephrasing of what was said is concerned with correct understanding. A signal of positive understanding implies that the speaker also perceived well; on the other hand, a signal of imperfect understanding implies good perception.

In DIT, five levels of processing are distinguished which have logical relationships that turn up as implications between feedback acts at different levels:

(10) attention $<$ perception $<$ understanding $<$ evaluation $<$ execution

'Evaluation' should be understood here in relation to the information-state update approach followed in DIT, and the requirement that information states at all times be internally consistent. For example, the recipient of an inform act with a semantic content $p$ knows, upon understanding the behaviour expressing this act, that the speaker wants him to insert the information $p$ in his information state. Before doing this, the recipient has to evaluate whether $p$ is consistent with his current state; if not; the update would be unacceptable. Successful execution would mean that the recipient inserts $p$ in his information state.

The implication relations between feedback at different levels are either entailments or implicatures. In the case of positive feedback, an act at level $L_i$ entails positive feedback at all levels $L_j$ where $i > j$; positive feedback at execution level therefore entails positive feedback at all other levels. By contrast,

[9] See also ISO (2009) for a discussion of these and other relations.

positive feedback at level $L_i$ *implicates* negative feedback at all levels $L_j$ where $i < j$; for instance, a signal of successful perception implicates that there is a problem with understanding, for why not signal successful understanding if that were the case? This is, however, not a logical necessity, but rather a pragmatic matter, hence an implicature rather than an entailment.

For negative feedback the entailment and implicature relations work in the opposite direction from positive feedback. For allo-feedback the same relations hold as for auto-feedback.

### 4.3 Indirect speech acts

The phenomenon known as 'indirect speech acts' is another potential source of multifunctionality. An utterance such as *Can you pass me the salt?* has been analysed as expressing both a question about the addressee's abilities and, indirectly, a request to pass the salt.

Using DIT or another semantic, ISU-based approach, such an analysis does not make much sense, however, since a request to do X is normally understood to carry the assumption (on the part of the speaker, S) that the addressee (A) is able to do X; hence the interpretation of the utterance as a request would lead to an update of the context to the effect that ( among other things) *A believes that S believes that A is able to pass the salt*. The interpretation as a question about the addressee's abilities would lead to an update including that *A believes that S wants to know whether A is able to pass the salt*. These two updates would be in logical conflict with each other, and would result in an inconsistent information state. Now a context model might be allowed to contain certain inconsistencies when human dialogue participants are modeled, but the inconsistency between believing that *S believes that A is able to do X* and believing that *S wants to know whether A is able to do X* is too blatant to be accptabl in any context model.

A popular alternative analysis of cases like *Can you pass me the salt?* and *Do you know what time it is?* is to interpret such utterances as just the indirectly expressed request *Please pass me the salt* and *Please tell me what time it is*, respectively. On this analysis, the indirectness is considered as a matter of politeness, and not having any semantic consequences.

The DIT analysis of such cases is as follows. S has a goal G that could be achieved by successful performance of a dialogue act with function $F_1$; however, $F_1$ has a precondition $p_1$ of which S does not know whether it is satisfied, and which S believes A knows whether it is satisfied (for instance, a property of A). S therefore asks A whether $p_1$. A understands this situation, and more specifically understands that S would like to perform the dialogue act with function $F_1$ if the condition $p_1$ is satisfied. In other words, S's utterance is understood as a conditional request: *If you are able to pass me the salt, please do so*. A similar analysis applies to a range of other types of indirect questions.

To the extent that a functional segment realizing an indirect dialogue act, such as a conditional act, can be analysed as 'indirect' but conventionalized expression of a single dialogue act, different from the type of act that the function indicators in the segment would suggest, rather than an expression of *both* a 'direct' and an 'indirect' dialogue act, the phenomenon of indirect dialogue acts is not a source of multifunctionality.

### 4.4 Empirical determination of multifunctionality

The multifunctionality of utterances in dialogue can be empirically investigated given a corpus of dialogues annotated with communicative functions. We investigated the multifunctionality that is observed in a corpus of dialogues annotated with the DIT$^{++}$ scheme, taking two variables into account:

(i) the segmentation method that is used, i.e. the choice of units in dialogue to which annotations are assigned; and

(ii) the annotation strategy that is used, reflecting alternative views on what counts as multifunctionality.

Two expert annotators marked up 17 dialogues in Dutch (around 725 utterances) using the DIT$^{++}$ scheme as part of an assessment of the usability of the annotation scheme. Several types of dialogue were included:

(1) dialogues over a microphone and head set with a WOZ-simulated helpdesk, providing assistance in the use of a fax machine (from the DIAMOND corpus[10]);

(2) human-human telephone dialogues with an information service at Amsterdam Airport;

(3) human-computer telephone dialogues about train schedules (from the OVIS corpus); [11]

(4) Dutch Map Task dialogues.

We compared two segmentation methods:

**a. turn-based:** the turn unit is taken as the unit which is annotated with communicative functions;

**b. functional-segment based:** functional segments, as defined in Section 3 above, are distinguished for each (possibly discontinuous) stretch of behaviour which has one or more communicative functions.

The dialogues were segmented into functional segments and annotated accordingly; from this segmentation and annotation we reconstructed the annotation that would correspond to turn-based segmentation.

The following 8 strategies were compared for dealing with the various possible sources of multifunctionality:

**a. strictly feature-based:** only communicative functions are marked which are recognizable from utterance features (lexical, syntactic, prosodic), given the context of the preceding dialogue. Only explicit feedback functions are marked, and Turn Management functions are marked only if they are explicitly indicated through lexical and/or prosodic features;

**b. + implicated functions:** implicated functions are are marked as well;

**c. + turn taking:** a turn-initial segment (i.e., a functional segment occurring at the start of a turn) is marked by default as having a Turn Take function if it does not already have a Turn Grab function (i.e., it forms an interruption) or a Turn Accept function (i.e., the speaker accepts the turn that was assigned to him by the previous speaker). In other words, starting to speak is by default annotated as an indication of the Turn Take function;

**d. + turn releasing:** similarly, ceasing to speak is by default annotated as a Turn Release act;

**e. + entailed feedback functions:** entailed feedback functions are also marked, such as the positive feedback on understanding that is entailed by answering a question or accepting an offer;

**f. + inherited functions:** entailed functions within a dimension, due to degrees of specificity are also marked, such as a Check Question also being a Propositional Question, and a Warning also being an Inform;

**g. + entailed feedback levels:** signals of positive feedback at some level of processing are also marked as positive feedback at lower levels, and negative feedback at a certain level is also marked as negative feedback at higher levels;

**g. + implicated feedback levels:** signals of positive feedback at some level of processing are also marked as (implicated) negative feedback at higher levels; signals of negative feedback at a certain level are also marked as positive feedback at lower levels;

**i. + indirect functions:** in the case of indirect speech acts, both the function of the direct interpretation and the one(s) of the intended indirect interpretation(s) are marked.

The dialogues were annotated using strategy b; the annotations according to the strategies c-i were reconstructed by adding the implied, indirect or default functions.

The results are summarized in Table 4.4. The figures in boldface in the second row represent the *minimal degree of multifunctionality* that it found, leaving aside all entailed functions, turn management function, indirect functions, and implied feedback levels. When fine-graned multidimensional segmentation is applied, using functional segments, then all sequential, overlapping, discontinuous and interleaved multifunctionality is eliminated and only purely simultaneous multifunctionality remains. The minimal

Table 7: Multifunctionality for various annotation strategies and segmentation methods.

| *segmentation method:* annotation strategy: | *turn* unit | *functional segment* |
|---|---|---|
| a. strictly feature-based | 2.5 | 1.3 |
| b. + implicated functions | **3.1** | **1.6** |
| c. + turn taking | 4.0 | 2.1 |
| d. + turn releasing | 4.8 | 2.6 |
| e. + entailed feedback | 5.2 | 2.8 |
| f. + inherited functions | 5.6 | 3.0 |
| g. + implic. feedb. levels | 6.3 | 3.2 |
| h + entailed feedb. levels | 6.6 | 3.4 |
| i. + indirect functions | 6.7 | 3.6 |

multifunctionality turns to be 1.6 in our data, i.e. a functional segment on average has 1.6. communicative functions.

The figures in the column for turn-based segmentation have to be taken with a big grain of salt, as they depend on the complexity of the turns in the dialogues that are considered, and this may vary a lot depending on the kind of dialogue. By contrast, the concept of a functional segment is very much independent of dialogue genre or communciative setting, hence these figures may be assumed to have a much broader significance.

## 5 Dialogue act annotation

Dialogue act annotation is the activity of marking up stretches of dialogue behaviour (linguistic, nonverbal, or multimodal) with information about the dialogue acts which are performed in these stretches of behaviour, and is often limited to marking up their 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'.

The DIT$^{++}$ schema takes 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 under-

[10]See `http://ls0143.uvt.nl/diamond`

[11]`http://www.let.rug.nl/~vannoord/Ovis`

stood.

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 utterance surface characteristics in combination with 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).

### 5.1 Annotation schemas

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. It has been noted, however, that one-dimensional annotation schemas also have serious disadvantages (see e.g. e.g. Klein et al., 1999; 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,

Multidimensional annotation using the DIT$^{+}+$ annotation schema leads to high-precision annotation due to:

- the relatively large number of communicative functions, DIT$^{+}+$ contains 96 defined communicative functions plus open categories for rhetorical and emotional
- variations, and possible function 'qualifiers';
- motivated careful choice of dimension; 10-dimensional structure giving optimal support to the assignment of multiple communicative functions to utterances;
- fine-grained multidimensional segmentation into functional units;
- well-worked out notion of multifunctionality, taking logical and pragmatic implication relations between functions into account.

### 5.2 Evaluation

For obtaining sound theoretical insights into dialogue phenomena, or for obtaining training data for automatic dialogue act tagging, dialogue act annotations with high reliability are a prerequisite. A dialogue act scheme can be applied reliably if the assignment of the categories in the scheme does not depend on individual judgment, but on a shared understanding of what the categories mean and how they are to be used. A frequently used way of evaluating human dialogue act annotations is measuring the inter-annotator agreement using Cohen's kappa statistic (Carletta, 1996), (Cohen, 1960) which is given by:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where $p_o$ is observed agreement and $p_e$ chance agreement.

Several annotation and evaluation experiments of the DIT$^{+}$+ annotation scheme were performed. Two groups of three undergraduate students annotated the dialogue material in two sessions per group. They had been introduced to the DIT$^{+}$+ annotation scheme and the underlying theory while participating in a course on pragmatics. It was noticed that hen considering inter-annotator agreement for the use of multidimensional tags from complex hierarchically structured tagsets, standard kappa statistics is not an appropriate measure because they cannot deal with the possibility of *partial* agreement (see (Geertzen & Bunt, 2006)). Instead, a weighted kappa statistic was adopted which can take into account a probability distribution typical for each annotator, generalize it to the case for multiple observers by taking the average over the scores of annotator pairs, and which uses a distance metric taking the structure of the taxonomy of tags into account. Geertzen & Bunt (2006) proposed the weighting in Cohens $\kappa_w$ and came to a coefficient which we will call *taxonomically weighted kappa*, denoted by $\kappa_w t$:

$$\kappa = 1 - \frac{\sum(1 - \delta(i,j)).p_o ij}{\sum(1 - \delta(i,j)).p_e ij}$$

where $\delta$ is a distance metric that measures disagreement and is a real number normalized in the range between 0 (not related functions) and 1 (identical functions). Table 8 shows the results, and compares standard and taxonomically weighted kappa scores.

| Dimension | standard | | | weighted | | |
|---|---|---|---|---|---|---|
| | $p_o$ | $p_e$ | $\kappa$ | $p_o$ | $p_e$ | $\kappa$ |
| Task | 0.52 | 0.09 | 0.47 | 0.76 | 0.17 | 0.71 |
| Auto-Feedback | 0.32 | 0.14 | 0.21 | 0.87 | 0.69 | 0.57 |
| Allo-Feedback | 0.53 | 0.19 | 0.42 | 0.79 | 0.50 | 0.58 |
| Turn Management | 0.90 | 0.42 | 0.82 | 0.90 | 0.42 | 0.82 |
| Time Management | 0.91 | 0.79 | 0.58 | 0.91 | 0.79 | 0.58 |
| Contact Management | 1.00 | 0.53 | 1.00 | 1.00 | 0.53 | 1.00 |
| Own Communication Management | 1.00 | 0.50 | 1.00 | 1.00 | 0.95 | 1.00 |
| Partner Communication Management | 1.00 | 1.00 | nav | 1.00 | 1.00 | nav |
| Dialogue structuring | 0.87 | 0.48 | 0.74 | 0.87 | 0.48 | 0.74 |
| Social Obligation Management | 1.00 | 0.19 | 1.00 | 1.00 | 0.19 | 1.00 |

Table 8: *Standard and taxonomically weighted kappa statistics per dimension drawn from the set of all annotation pairs (adopted from Geertzen & Bunt (2006)*

Annotation results were evaluated qualitatively as well during post-experimental sessions where annotators were interviewed about the difficulties they experienced while performing annotations as well as consulting annotation guidelines. It was found that annotators often have difficulties to determine a specific level of feedback (attention, perception, understanding, evaluation or execution), and that for Turn Management the annotation guidelines were unclear. These and other more detailed findings were used to design a revised tag set and improve the annotation guidelines within the European project LIRICS (see (Bunt et al., 2007)). Within this project, the dialogue act annotation scheme was developed

| Function class | English | Dutch | average |
|---|---|---|---|
| Information-seeking | 0.96 | 0.98 | 0.97 |
| Information-providing | 0.98 | 0.99 | 0.98 |
| Feedback | 0.98 | 0.99 | 0.99 |
| Interaction management | 0.92 | 0.96 | 0.94 |
| Social obligations management | 0.94 | 0.94 | 0.94 |

Table 9: Inter-annotator agreement for LIRICS communicative functions

Table 10: *Communicative function recognition scores per dimension in terms of accuracy (in %) comparing to baseline scores (BL) for each dimension and data set.*

| | DIAMOND data | | AMI data | | OVIS data | |
|---|---|---|---|---|---|---|
| Dimension | BL | Accuracy (%) | BL | Accuracy (%) | BL | Accuracy (%) |
| Task | 64.9 | 70.5 | 66.8 | 72.3 | 60.8 | 73.5 |
| Auto-Feedback | 71.1 | 85.1 | 77.9 | 89.7 | 66.1 | 75.9 |
| Allo-Feedback | 86.9 | 96.6 | 96.7 | 99.3 | 52.5 | 80.1 |
| Turn Management | 69.5 | 90.0 | 59.0 | 93.0 | 89.8 | 99.2 |
| Time Management | 65.6 | 82.2 | 69.7 | 99.4 | 95.5 | 99.4 |
| Discourse Structuring | 59.0 | 67.9 | 98.0 | 92.5 | 76.3 | 89.4 |
| Contact Management | 88.0 | 95.2 | 99.8 | 99.8 | 87.7 | 98.5 |
| Own Communication Management | 77.4 | 83.1 | 89.6 | 94.1 | 99.7 | 99.7 |
| Partner Communication Management | 45.4 | 62.6 | 99.7 | 99.7 | 99.8 | 99.8 |
| Social Obligation Management | 80.3 | 92.2 | 99.6 | 99.6 | 96.2 | 98.4 |

which is slightly simplified version of DIT$^{+}+$ taxonomy, e.g. the feedback levels were eliminated by keeping only underspecified feedback functions such as Positive and Negative Auto-Feedback, Positive and Negative Allo-Feedback and Feedback Elicitation. Communicative functions defined as ISO data categories, following ISO standard 12620. In April-June 2007 dialogue act annotation was performed for three languages: English, Dutch, and Italian.

For English selected dialogues from two dialogue corpora were annotated: TRAINS[12] (5 dialogues; 349 utterances) and MapTask[13] (2 dialogues; 386 utterances). For Dutch selected dialogues from two dialogue corpora were annotated: DIAMOND ((one extended dialogue, 301 utterances) and Schiphol (Amsterdam Airport) Information Office (6 dialogues; 202 utterances). For Italian 6 selected dialogues (393 utterances) from the SITAL corpus were annotated. For English and Dutch the test suite dialogues were all annotated by two expert annotators. An analysis of the agreement between their annotations reveals that in all of the frequently addressed dimensions a very high agreement was reached; see Table 9.

While human annotators are quite successful in communicative recognition, the question arises whether comparable scores can be obtained in machine recognition. We used the rule induction algorithm *Ripper*. The advantage of such an algorithm is that the regularities discovered in the data are represented as human-readable rules. It is also shown in (Geertzen et al., 2007) that Ripper performed best on our data compared to statistical learners (e.g. Naive-Bayes classifiers) and memory-based learners (e.g. IB1).

Table 10 presents the resulting scores using the Ripper classifier obtained in 10-fold cross-validation experiments. The features which are used in classification experiments are described in Section 3. Quantitative and qualitative analysis measuring annotating consistency, reliability and accuracy showed that DIT$^{+}+$ scheme and its simplified variant LIRICS scheme can be considered as usable dialogue act

[12]For more information about the TRAINS corpus please visit `http://www.cs.rochester.edu/research/speech/trains.html`

[13]Detailed information about the MapTask project can be found at `http://www.hcrc.ed.ac.uk/maptask/`

schemes in terms of clarity and accessibility of the tag set, and the use of fundamental concepts. As our results show, communicative functions defined in the DIT$^{+}+$ and LIRICS tag sets are recognizable with fairly good precision not only by human annotators but also by machine.

### 5.3 Towards an ISO standard for dialogue act annotation

As part of a larger effort to promote the building of interoperable language resources, the ISO organization has launched a series of projects that aim at defining annotation and representation standards for language resources. In particular, in the are of semantic annotation the project “Semantic annotation framework” (ISO 24617) has been launched, which consists of several sub-projects that each focus on a particular type of semantic information. The (sub-)project ”24617-2 Part 2: Dialogue acts” is focused on defining a standard for annotating spoken and multimodal dialogues with dialogue act information.

In this project, a multidimensional approach has been chosen in the style of the framework outlined in Section 2. Nine of ten dimensions of the DIT$^{+}+$ schema form the ‘core dimensions’ of the proposal in its current state (see ISO 24617-2, 2009). The DIT$^{+}+$ dimension of contact management is considered as ‘optional’, since it plays a marginal role in certain kinds of dialogue.

The ISO standard proposal distinguishes 30 core dimension-specific and 19 core general-purpose functions, taken from the DIT$^{+}+$ taxonomy, using the following criteria:

(11) Every communicative function is required to be:

1. empirically observed;
2. theoretically validated;
3. relevant for obtaining a good coverage of the phenomena in a given dimension;
4. recognizable with acceptable precision by human annotators and by automatic interpretation systems;
5. present in a significant number of annotation schemes.

The ISO proposal includes the definition of Dialogue Act Markup Language (DiAML) compliant with the ISO Linguistic Annotation Framework (LAF, Ide et al., 2004), which distinguishes between *annotations* and *representations*. The term ‘annotation’ refers to the information that is added to segments of language data; 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. According to LAF, annotations rather than representations are the proper level of standardization. This distinction is reflected in the definition of DiAML, 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. The concrete syntax specifies a way to represent annotation structures in XML.

A fundamental requirement on semantic annotation is that semantic markups should have a well-defined semantics (Bunt & Romary, 2002). The DiAML language has a formal model-theoretic semantics associated with its abstract syntax, that rests on the assignment of information-state update schemas to communicative functions, which can be instantiated with a given semantic content.

A concrete example is shown in 12. P2’s utterance is segmented into two overlapping functional segments: one in the Auto-Feedback dimension and one in the Task/Activity dimension, with value ‘answer’ qualified as ‘uncertain’. Annotations may be attached directly to primary data like stretches of speech, defined by temporal begin and end points, but more often they will be attached to structure at other levels of analysis, such as the output of a tokenizer. Here we assume the TEI-ISO standard ISO 24610-1 is followed for attaching information to digital documents. In the example, we assume that the dialogue participants are identified in the metadata of the primary data as ”p1” and ”p2”, and that their utterances are segmented multidimensionally into the functional segments ”fs1”, ”fs2”, and ”fs3”.

(12) a.

| | |
|---|---|
| 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] |

b.
```
<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>
```

# 6 The semantics of discourse markers

## 6.1 Discourse markers

Discourse markers are key indicators of discourse structure, and have been proved to be useful devices for (a) segmenting discourse into meaningful units, and (b) identifying relations between these units. The determination of the meanings of discourse markers is often crucial for understanding the message that is communicated.

Discourse markers have been studied for their role in the organization of discourse structure in larger texts (Mann & Thompson, 1988), (Sporleder & Lascarides, 2008), in argumentative dialogues (Cohen, 1984), in interviews (Schiffrin, 1987) and (Hirschberg & Litman, 1993) and in dialogues that are highly interactive in nature and are characterized by rapid turn switching among participants, such as task-oriented dialogues (Heeman & Allen, 1999) or meeting conversations (Popescu & Zufferey, 2006).

We have seen in Section 4 that the assignment of communicative functions to utterances in multiple dimensions helps to represent the meaning of dialogue contributions in a more accurate way than is possible when only a single function is considered. We observed that even such small dialogue segments as discourse markers may have various communicative purposes relating to the underlying task or activity, attention, topic management, turn taking, etc. They may also have various communicative functions simultaneously. Consider the following dialogue fragment:

(13) *A1: that's why i think the option of the kinetic thing*
*A2: which basically means as long as you shake it like an automatic watch*
*D1: -1.781* ***but***
*D2: are people gonna wanna shake their movie controller?*

Utterance D1 is multifunctional since it indicates that (1) the speaker wants to have the turn by interrupting the previous speaker (signalled by 'but' overlapping A3); (2) the speaker interpreted and evaluated the utterances A1 and A2 successfully; and (3) the speaker encountered a problem in applying the information from the previous utterances (due to the adversative meaning of 'but'): he probably does not agree with the previous claim or needs some clarification, which is indeed expressed in D2. Thus, as the example shows, the various functions of 'but' are related to different dimensions of the interaction, such as the allocation of the speaker role and the processing of previous utterances.

One aspect of the meaning of discourse markers is that they may not only have a variety of semantic functions, but that they may also have several functions simultaneously – they may be *multifunctional*. The multifunctionality of discourse markers has been described first by Schiffrin in (Schiffrin, 1987) and by Hovy (Hovy, 1995). They argue that there are several parallel simultaneous structures that underlie coherent discourse and argue that an adequate description of discourse requires at least four distinct structural analyses: semantic, interpersonal/goal-oriented, attentional/thematic, and rhetorical. They noticed that discourse markers may simultaneously have roles within each of these structures, e.g. the discourse marker *and* may 'coordinate ideas' and 'continue a speaker's action'.

Although these approaches seem to apply very well to the analysis of the meaning of discourse markers in dialogue, they escaped extensive and formal description. The multidimensional framework allows to compute a 'multidimensional' semantics of discourse markers by relating multiple context update operators (dialogues acts performed by discourse markers) to different compartments of structured context models which include, besides information states of the usual kind (beliefs and goals related to a task domain), also a dialogue history, information about the agent's processing state, beliefs about dialogue partners' processing states, information and goals concerning the allocation of turns, and so on, relating to the various 'dimensions' that dialogue acts belong to.

We analysed the multifunctionality of discourse markers empirically. For this purpose, we selected three meetings containing 17,335 words which form 3,897 functional segments with an average length of 4.4 words from the AMI corpus[14].

### 6.2 The multifunctionality of discourse markers

We found that discourse markers are used (i) as 'preface' of a range of general-purpose functions, in particular of Informs of various rhetorical kinds; (ii) as indicators of dialogue acts with a dimension-specific function, e.g. of topic shifts; and (iii) as full-blown dialogue acts (without explicit semantic content), e.g. as a Turn Take act. This means that discourse markers can have two kinds of meanings: as a dialogue act, i.e. as a context update operator, and as an element that contributes to the determination of the communicative function of a dialogue act with either a general-purpose or a discourse-specific function.

Table 11 lists some discourse markers (DMs) with their absolute frequency in our corpus, gives an overview of their observed mutlifunctionality by indicating the average number of communicative functions, and lists the observed communicative functions.

| DM | Occurrence | Multifunctionality | GP | Auto-F. | Allo-F. | TurnM. | TimeM. | DS | ContactM. | OCM |
|---|---|---|---|---|---|---|---|---|---|---|
| and | 214 | 2.6 | elaborate<br>suggest<br>exemplify<br>explain | pos.<br>evaluate<br>pos.<br>execute | pos.<br>execute | take<br>grab<br>keep | stall | topic<br>shift | | |
| because | 37 | 1.2 | justify<br>explain | | | keep | stall | | | |
| but | 87 | 1.9 | disagree<br>correct<br>explain<br>warning | pos.<br>evaluate<br>neg.<br>execute | neg.<br>execute | take<br>grab<br>keep | | | | |
| I mean | 41 | 2 | elaborate | | | keep | | | | retract |
| like<br>such as | 70 | 1.7 | exemplify | | | keep | stall | | | |
| oh | 31 | 2 | answer | pos/neg.<br>execute | | grab<br>accept | pause | topic<br>shift | | error<br>signal |
| so | 226 | 2 | conclude<br>suggest<br>elaborate | pos.<br>execute | | take<br>grab<br>keep<br>give | stall | topic<br>shift<br>open<br>preclose | indi-<br>cate | retract |
| then | 45 | 1.9 | instruct<br>elaborate<br>suggest | pos.<br>execute | | take<br>grab<br>keep | stall | topic<br>shift | | |
| well | 63 | 2.1 | disagree<br>correct<br>explain | pos.<br>execute | neg.<br>execute | grab<br>take<br>accept<br>keep | stall | topic<br>shift | | retract |
| you<br>know | 84 | 2.3 | | | check<br>elicit | give<br>keep | stall | | check | retract |

Table 11: *Distribution and observed multifunctionality of discourse markers.*

Note that all DMs serve more than one communicative function. *And* is the most multifunctional discourse marker in our corpus and *because* the least multifunctional one. *Because* mostly prefaces Informs

[14] Augmented Multi-party Interaction (`http://www.amiproject.org/`)

with the rhetorical functions Justify or Explain, and only in 2.4% of all cases is used to simultaneously perform Turn Keeping and Stalling acts. All discourse markers, except *'you know'*, preface general-purpose functions used in the Task/Activity dimension or in an other dimension (often in Discourse Structuring or Feedback) and may perform dialogue acts addressing other dimensions simultaneously. This pattern is observed for 50.7% of all studied DMs. A discourse marker may also perform full-fledged dialogue acts addressing more than one dimension simultaneously. This is often the case for Turn Management in combination with Feedback, Time Management, Discourse Structuring and Own Communication Management (27.7% of all discourse markers are observed to be used in this way). It was noticed that at most 3 dialogue acts are performed by one discourse marker in a given context, e.g. feedback, turn and time management acts. A third pattern of DM use, which was observed in 18.2% of cases, is as a single dialogue act, e.g. a turn taking act or a feedback act. In the rest (3.4%) discourse markers are part of general purpose functions and do not perform a dialogue act on their own.

Different uses of discourse markers require recognition based on observable features like differences in prosody, such as *pitch* (F0 in Hz), *energy* (RMS), *voicing* (fraction of locally unvoiced frames and number of voice breaks), *speaking rate* (number of syllables per second) and segment *duration*, and features from surrounding lexical material such as frequent collocations. We observed significant mean differences for both raw and speaker-normalized features in terms of duration (DMs are almost twice as long as non-DMs:[15] 327ms and 217ms respectively, and Stallings and Keepings acts are even more longer: 585ms); initial pause (there is no or a negligible small pause before non-DMs, and initial pauses before DMs range between 59 and 228ms); mean pitch (*and* as DM has higher mean pitch: $>$ 12Hz). Preceding and following tokens as features also have high information gain. For example, Inform Elaborate is often signalled by focusing adverbs like *especially*, *mainly*, *additionally*, etc, or contains relative pronouns like *who*, *whose*, *which* and *that*. The presence of some expressions of comparision was noted in Exemplifications, such as *one of*, *rather than*, *like*, *comparing*, *by contrast*, *similar*, etc. The most frequent words that occurred in Suggestions are *maybe*, *might*, *better*, *should*, *could/can*, *probably* and *let's*; and Discourse Structuring functions are marked by *next*, *then*, *other* and *further*.

To summarize we can conclude that discourse markers are truly multifunctional elements in dialogue. Discourse markers are good indicators of (plentiful) general-purpose communicative functions, such as informs, elaborations of various kinds, suggestions, warnings, disagreements, etc., mostly in relation to the task or activity that underlies the dialogue, but they are also frequently used to create or maintain the conditions for successful interaction. The analysis of discourse markers as important instruments for the understanding of dialogue and its computational modelling can only benefit from a multidimensional approach to the exploration of their complex meaning.

## 7 Generation of multifunctional utterances

We have seen that functional segments mostly have more than one communicative function. Those segments which are neither turn-initial nor turn-final, and hence do not have a Turn Take or Turn Release function, have an average multifunctionality of 1.6 due to their function indicating features and conversationally implicated functions (leaving implied feedback functions and indirect functions out of consideration). This conclusion does not only have consequences for the semantic interpretation of dialogue utterances, but also for their generation by spoken dialogue systems. When a dialogue system, in interaction with a user, has the turn and should produce a contribution that is appropriate for continuing the dialogue, how should it take multifunctionality into account?

The stretch of speech, possibly combined with visual elements, which a dialogue system produces when it has the turn, is by definition a stretch of the kind that we have called a *turn unit*. Table (4.4) shows that in human-human dialogue a turn unit has on average a multifunctionality that is twice that of a functional segment. Leaving turn management functions, feedback functions, and indirect function aside, a turn unit on average has 3 communicative functions where a functional segment has 1.6, which confirms the experience that turn units usually contain more than one functional segment. Existing

[15] By 'non-DM' is meant that the same lexical item is used not as a discourse marker but in another way, which is mostly that of a propositional connective.

dialogue systems such as TRIPS (Allen et al. 2001), GoDIS (Larsson, 2001) generate an output by constructing an utterance which, when interpreted successfully by the user, satisfies a particular goal or step in a plan. Since a natural language utterance almost inevitably has multiple communicative functions, however, the user may be expected to interpreted the utterance as indeed having all these functions, whereas the dialogue system produced the utterance in order to have just *one* of them. This is likely to lead to misunderstandings. An example of such a misunderstanding is the following.

The experimental OVIS spoken (telephone) dialogue system, which provides information about train travel in the Netherlands, and which was put in practice for some time to gather information about its potential usability, would ask the user after successfully producing a itinerary whether the user is interested in an alternative possible itinerary. For example, in the following fragment the user has requested an itinerary for traveling by train from Heeze to Purmerend on Friday morning, arriving before 09:30. After the system has provided one, departing from Heeze at 06:38, with stopovers in Eindhoven, Utrecht, and Amsterdam, the system continues as follows:

| | | | |
|---|---|---|---|
| (14) | 1. | S: | Would you like to see an alternative itinerary? |
| | 2. | U: | Yes please. |
| | 3. | S: | Same day? |
| | 4. | U: | Yes, same day. |
| | 5. | S: | Same destination? |
| | 6. | U: | Yes, same destination. |
| | 7. | S: | Same place of departure? |
| | 8. | U: | Yes. |
| | 9. | S: | Departing earlier or later? |
| | 10. | U: | Earlier please. |
| | 11. | S: | There is no earlier possibility.<br>Would you like to see a later possibility? |
| | 12. | U: | No thank you. |

The source of the problem that we see here is that the OVIS system designers, unaware of the multiple dimensions that dialogue utterances may address, considered an utterance like (14.1) as expressing just a question. However, the utterance is interpreted by the user as (also) expressing an *offer*, and this leads to a rather comical situation, comparable to the one that arises in the following case:

| | | | |
|---|---|---|---|
| (15) | 1. | A: | Would you like to have some coffee? |
| | 2. | B: | Yes, that would be great. |
| | 3. | A: | I have no coffee. |

Can this sort of misunderstandings be avoided? This question has not been investigated systematically, but it seems certain that, if spoken dialogue system designers are unaware of the natural multifunctionality of the spoken utterances that their systems generate, then this issue will not be resolved. What could be done for a start is to feed the spoken utterances which a dialogue system generates back into its own (input) interpretation component which, if designed to recognize multiple communicative functions, will make exlicit which functions are likely to be assigned to this utterance in addition to the intended one. The system can then check if one or more of these additional functions would be inappropriate in the given context, and if so, modify its output accordingly.

More generally, designing an output generation component to be aware of the potential multifunctionality of its results and to take multiple communicative functions into account, opens the perspective of generating utterances which are multifunctional *by design*, rather than by accident. First steps in this direction have been taken in the PARADIME dialogue manager Keizer and Bunt, 2007).

## 8 Conclusions

In this paper we have argued that the study of spoken dialogue can benefit in many ways from the use of a multidimensional theoretical framework, based on (1) a well-defined notion of dimension, and (2) a notion of dialogue act, formalized by means of an information state update semantics.

We have shown that such an approach leads to a new view on dialogue segmentation, which abandons the traditional idea of segmentation in the sense of cutting up the speech stream into a linear sequences of units. A more meaningful way of segmenting dialogue is in terms of 'functional segments' which can overlap, can include other segments, can be discontinuous, and can spread over multiple turns. This notion of functional segment allows a more accurate determination of functionally meaningful parts in a dialogue. We have showed that the boundaries of functional segments can quite well be detected automatically.

We have also shown that the combination of a fine-grained conceptual framework, based on the view that functional segments may express multiple dialogue acts in different dimensions, is useful for a conceptual qualitative as well as an empirical quantitative analysis of the multifunctionality of dialogue utterances. The question why dialogue utterances are multifunctional has been answered by treating participation in a dialogue as involving multiple activities at the same time, such as making progress in a given task or activity; monitoring attention and understanding; taking turns; and managing time and social obligations. This approach has been backed up by empirical data, which show that functional segments display both what we called 'independent' multifunctionality and implicated multifunctionality. Entailment relations between dialogue acts, and default and indirect functions add further to the mulltifunctionality that can be observed.

In dialogue act annotation is is not uncommon to take a multidimensional approach. We have shown that such an approach provides maximal support for dialogue annotators when it is based on a swell-defined notion of dimension. Although the resulting 10-dimensional annotation schema may seem complex, we have established that human annotators can work very well with it, and that machines can learn to apply it with acceptable accuracy. We briefly indicated how these results find their way into an effort of the ISO organization that aims at developing an international standard for dialogue annotation.

We have shown that discourse markers, which play an important part in the structuring and the semantic interpretation of discourse, tend to display a multifunctionality which can be captured surprisingly well in terms of the dimensions and communicative functions of the DIT$^{++}$ framework. Discourse markers were incidentally found to be very important in automatic detection of functional segment boundaries.

We finally provided some, admittedly rather speculative, discussion of the potential benefits of a multidimensional approach to the automatic generation of spoken dialogue utterances. Although we have only a very preliminary implementation of this idea to support the discussion, it seems clear that the multifunctionality that dialogue utterances naturally have, but which is usually considered a problem in computational approaches and swept under the rug, cannot be avoided, and should be *exploited* rather than ignored. A multidimensional approach, taking into account insights that have been obtained in the nature and forms of multifunctionality, opens the perspective of a spoken dialogue system that generates utterances which are multifunctional by design, rather than by accident.

**Acknowledgements**

We would like to thank the members of the Tilburg Dialogue Club who have over the years been involved in the development of DIT, in particular Jeroen Geertzen, Hans van Dam, Ielka van der Sluis, Simon Keizer, Roser Morante, and Amanda Schiffrin. We also thank the colleagues with whom we collaborate in the ISO project 24617-2 "Semantic annotation framework, Part 2: Dialogue acts": Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Fang, Koiti Hasida, Kiyong Lee, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum.

## References

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.

Allen, J., L. Schubert, G. Ferguson, P. Heeman, C.H. Hwang, T. Kato, M. Light, N. Martin, B. Miller, M. Poesio, D. Traum. 1994. *The TRAINS project: A case study in defining a conversational planning agent.* Technical Report 532, Computer Science Department, University of Rochester.

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.

Allen, J., Ferguson, G., and Stent, A. 2001. *An architecture for more realistic conversational systems*. Proceedings of the 6th International Conference on Intelligent User Interfaces (IUI'01), Santa Fe, New Mexico, USA, pp. 1-8.

Allwood, J. 1992. *On dialogue cohesion.* Gothenburg Papers in Theoretical Linguistics 65. Gothenburg Unviersity, Department of Linguistics.

Allwood, J. 1984. Obligations and options in dialogue. *THINK Quarterly* 3 (1), 9–18.

Allwood, J. 2000. An activity-based approach to pragmatics. In H. Bunt and W. Black, editors, *Abduction, Belief and Context in Dialogue*, John Benjamins, Amsterdam, pp. 47–80.

Bunt, H. 2000. *Dialogue pragmatics and context specification*. H. Bunt and W. Black (eds) Abduction, Belief and Context in Dialogue. Amsterdam: Benjamins, 81-150.

Bunt, H. 2005. *A Framework for Dialogue Act Specification*. ISO-SIGSEM workshop, Tilburg. `http://let.uvt.nl/research/TI/sigsem/wg/meeting4/fdas-orig.pdf`

Bunt, H. 2006. Dimensions in dialogue annotation. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006).*

Bunt, H., Petukhova, V., and Schiffrin, A. 2007. *Multilingual test suites for semantically annotated data.* LIRICS Deliverable D4.4

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

Carletta, J. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249-254.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, 20:37–46.

Cohen, R. 1984. *A computational theory of the function of clue words in argument understanding*. Coling-ACL 1984, Standford, pp. 251–258.

William W. Cohen. 1995. Fast effective rule induction. In *Proceedings of the 12th International Conference on Machine Learning (ICML'95)*, pages 115–123.

Geertzen, J., and Bunt, H. 2006. *Measuring annotator agreement in a complex hierarchical dialogue act annotation scheme.* Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue, Sydney, Australia, pp. 126–133.

Geertzen, J., V. Petukhova, and H. Bunt. 2007. A Multidimensional Approach to Utterance Segmentation and Dialogue Act Classification. *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, Antwerp, pp. 140–149.

Geertzen, J., Petukhova, V. and Bunt, H. 2008. *Evaluating dialogue act tagging with naive and expert annotators.* Proceedings of the Sixth International Conference on Language Resources and Evaluation, Marrakech, Morocco.

Geertzen, J. 2009. *The automatic recognition and prediction of dialogue acts.* Ph.D. Thesis, Tilburg Universsity, February 2009.

Heeman, P.A. and Allen, J.F. 1999. Speech repairs, intonational phrases and discourse markers: Modelling speakers utterances in spoken dialogue. *Computational Linguistics*, 12(3): 1–45.

Hirschberg, J. and Litman, D. 1993. Empirical studies on the disambiguqtion of cue phrases. *Computational Linguistics*, 25(4): 501–530.

Hovy, E.H. 1995. *The multifunctionality of discourse markers*. Proceedings of the Workshop on Discourse Markers, Egmond-aan-Zee, The Northerlands.

ISO. 2009. *Semantic annotation framework, Part 2: Dialogue acts.* ISO TC 37/SC 4/N442 rev03, ISO, Geneva, April 2009.

Keizer, S. and Bunt, H. 2006. *Multidimensional dialogue management*. In Proceedings of the 7th Workshop on Discourse and Dialogue, *SIGDIAL 2006*, Sydney.

Keizer, S. and Bunt, H. 2007. *Evaluating combinations of dialogue acts*. In Proceedings of the 8th Workshop on Discourse and Dialogue, Antwerp.

Klein, M. 1999. *Standardization efforts on the level of dialogue act in the MATE project*. Available at `http://acl.ldc.upenn.edu/W/W99/W99-0305.pdf`

Larsson, S. 1998. *Coding Schemas for Dialogue Moves*. Technical report from the S-DIME project.

Larsson, S., Ljunglóf, P., Cooper, R., Engdahl, E., and Ericsson, S. 2000. *GoDis: an accommodating dialogue system.* Proceedings ANLP/NAACL 2000 Workshop on Conversational Systems, Seattle, Washington, pp. 7-10.

Mann, W. and Thompson, S. 1988. *Rhetorical structure theory: toward a functional theory of text organisation*. The MIT Press, Cambridge, MA.

Morante, R. 2007. *Computing Meaning in Interaction*. PhD Thesis, Tilburg University.

Nöth, E., A. Batliner, V. Warnke, J.-P. Haas, M. Boros, J. Buckow, R. Huber, F. Gallwitz, M. Nutt, and H. Niemann. (2002) On the use of prosody in automatic dialogue understanding. *Speech Communication*, 36(1-2):45–62.

Petukhova, V. and Bunt, H. 2009. *Dimensions of communication*. Technical Report TiCC TR 2009-003, Tilburg Centre for Creative Computing, Tilburg University.

Petukhova, V. and Bunt, H. 2009. *The independence of dimensions in multidimensional dialogue act annotation*. Proceedings NAACL HLT Conference, Boulder, Colorado, June 2009.

Popescu-Belis, A. 2005. *Dialogue Acts: One or More Dimensions?* ISSCO Working Paper 62, ISSCO, Geneva.

Popescu-Belis, A. and Zufferey, S. 2006. *Automatic identification of discourse markers in multiparty dialogues*. Working paper 65, ISSCO, University of Geneva.

Schiffrin, D. 1987. *Discourse Markers*. Cambridge: University Press.

Sporleder, C. and Lascarides, A. 2008. Using Automatically Labelled Examples to Classify Rhetorical Relations: An Assessment. *Natural Language Engineering*, 14.03: 369–416.

Shriberg, E., R. Bates, A. Stolcke, P., Taylor, D. Jurafsky, K. Ries, N. Coccaro, R. Martin, M. Meteer, and C. Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech*, 41(3-4):439–487.

Stolcke, A., K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. Van Ess-Dykema, and M. Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–373.

Traum, D., Larsson, S. 2003. The information state approach to dialogue management. In J. van Kuppevelt & R. Smith 9eds.) *Current and New Directions in Discourse and Dialogue*, pp. 325-345. Kluwer, Dordrecht.

Ian H. Witten and Eibe Frank. 2000. *Data mining: practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann Publishers, San Francisco:CA, USA.