# Multifunctionality in dialogue

Harry Bunt

▶ **To cite this version:**

Harry Bunt. Multifunctionality in dialogue. Computer Speech and Language, 2010, 25 (2), pp.222. 10.1016/j.csl.2010.04.006 . hal-00692189

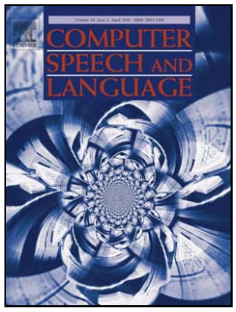
**HAL Id: hal-00692189**
**https://hal.science/hal-00692189**

Submitted on 29 Apr 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript




Title: Multifunctionality in dialogue

Author: Harry Bunt

PII: S0885-2308(10)00037-9
DOI: doi:10.1016/j.csl.2010.04.006
Reference: YCSLA 453

To appear in:

Received date: 8-7-2009
Revised date: 23-12-2009
Accepted date: 1-4-2010

Please cite this article as: Bunt, H., Multifunctionality in dialogue, *Computer Speech & Language* (2008), doi:10.1016/j.csl.2010.04.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

# Multifunctionality in dialogue

Harry Bunt,
TiCC- Tilburg center for Creative Computing
Department of Communication and Information Sciences
Tilburg University

**Abstract**

This paper studies the multifunctionality of dialogue utterances, i.e. the phenomenon that utterances in dialogue often have more than one communicative function. It is argued that this phenomenon can be explained by analyzing the participation in dialogue as involving the performance of several types of activity in parallel, relating to different dimensions of communication. The multifunctionality of dialogue utterances is studied by (1) redefining the notion of 'utterance' in a rigorous manner (calling the revised notion 'functional segment'), and (2) empirically investigating the multifunctionality of functional segments in a corpus of dialogues, annotated with a rich, multidimensional annotation schema. It is shown that, when communicative functions are assigned to functional segments, thereby eliminating every form of segmentation-related multifunctionality, an average multifunctionality is found between 1.8 and 3.6, depending on what is considered to count as a segment's communicative function. Moreover, a good understanding of the nature of the relations among the various multiple functions that a segment may have, and how functional segments relate to other units in dialogue segmentation, opens the way for defining a multidimensional computational update semantics for dialogue interpretation.

## 1 Introduction

Utterances in dialogue are often multifunctional, i.e., they serve more than one communicative function. While this has been widely recognized (see e.g. Allwood, 1992; Bunt, 1994; 2000; Popescu-Belis, 2005, Traum & Hinkelman, 1992; Traum, 2000), the consequences of this observation for the semantics of dialogue utterances and the design of dialogue systems have hardly been explored. Computational approaches to dialogue generally see multifunctionality as a problem (Traum, 2000), and tend to ignore the phenomenon in actual implementations.

One computationally oriented area of dialogue research that has tried to take the multifunctionality of utterances into account is that of the functional annotation of dialogue, usually referred to as *dialogue act annotation*. When annotating a dialogue utterance with its communicative function, one simply cannot avoid acknowledging that some utterances have more than one function. This has led to the development of 'multidimensional' dialogue act annotation schemas, where an utterance can be marked up

ACCEPTED MANUSCRIPT

with more than one functional tag. Until very recently, the multidimensional annotation schemas that have been proposed, such as DAMSL (Allen & Core, 1997); COCONUT (Di Eugenio et al., 1998) MRDA (Dhillon et al., 2004); do not base their 'multidimensionality' on a clear notion of dimension; rather they use 'dimension' as a label for a group of mutually exclusive tags. Also, these schemas follow the tradition in annotation work of conceiving text markup as a purely descriptive *labelling* activity, and as such cannot be expected to provide an explanatory account of multifunctionality.

The study of multifunctionality reported in this paper has been inspired by recent work in semantic annotation, which approaches the task of annotating linguistic (or multimodal, linguistic + nonverbal) behaviour with descriptors that have a formal semantics. On this approach, assigning a communicative function to an utterance is a step in its semantic interpretation. This seems appropriate, since assigning to an utterance like *Why don't you start?* the communicative function of an encouragement to start, rather than a question about the motives for not starting, is a matter of assigning a certain type of meaning to the utterance. We will argue that, by using a semantically well-defined set of communicative functions organized into conceptually clear dimensions, we obtain an approach to the annotation and interpretation of dialogue utterances which provides an explanatory account of the multifunctionality that is empirically found in natural dialogue.

Concerning the semantic interpretation of dialogue utterances, it may be noted that traditional approaches to the analysis of sentence meaning notoriously fail. This is partly because these approaches are rooted in the truth-conditional view of meaning, while dialogue utterances like *Good morning*, *Yes okay* and *Let me see...* have meanings that cannot be captured in terms of the truth or falsity of propositions.

Alternatively, the semantics of dialogue utterances has been studied in terms of 'information-state update' or 'context-change' approaches (Poesio & Traum, 1998; Bunt, 2000; Traum & Larsson, 2003; Bos et al., 2003), which view utterance meanings in terms of changes in the 'information states' or 'context models' of the dialogue participants. These approaches closely relate to the ideas of speech act theory, which views the use of language as the performance of communicative actions.

Action-based approaches to utterance meaning have to face the fact that, contrary to what traditional speech act theory tells us, dialogue utterances are often used to perform multiple communicative acts, such as answering a question, providing feedback about the understanding of the question, and taking the turn. The following example illustrates this.

(1)
1. A: What time is the next train to Amersfoort?
2. B: Let me see.... The next one will be at eleven twenty-five.
3. A: Eleven twenty-five. There's no earlier train to Amersfoort than that?
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. So apparently B's utterance 3 carried the additional function of providing positive feedback on B's understanding of the previous question.

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

ACCEPTED MANUSCRIPT

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 performed simultaneously. Moreover, some of these speech act types, such as feedback acts and turn-taking acts, have hardly if at all 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 context models have therefore in fact not related closely to speech act theory, but rather to modern, data-driven versions of 'dialogue act' theory, such as DIT (see Section 3.2). In a dialogue-act based framework, interpreting an utterance as expressing a certain type of dialogue act is implemented by associating a particular type of information state update with each type of dialogue act. Since a multifunctional dialogue utterance corresponds semantically to the performance of more than one dialogue act, such an utterance should be treated semantically as a combination of the updates corresponding to each of the dialogue act types expressed.

One of the reasons why dialogue utterances often have multiple communicative functions is that, in addition to the functions which are signaled through observable utterance features (choice of words, word order, intonation, accompanying gestures,...), other functions are often *implied* by what is signaled. Example 1 illustrates this: in the first part of B's utterance 2 the speaker signals that he is stalling for time, slowing down and using the expression *Let me see*; by implication this part of the utterance also constitutes a turn-taking act. The second part constitutes an answer due to its form and content plus the fact that it follows a question; by implication it also gives the feedback information that A's question was well understood. In Section 4 we will discuss the phenomenon of implied functions in some detail, and also consider other reasons why dialogue utterances may have multiple functions.

Besides the funcamental question *why* utterances are often multifunctional, a practical question is *how many* functions a dialogue utterance typically has, and which factors determine this. One of the obvious factors is the way dialogue is segmented into functional units. The relation between multifunctionality and segmentation will be discussed in Section 2. The quantitative aspect of utterance multifunctionality will be analyzed in Section 5, by investigating a dialogue corpus annotated with communicative functions, using various segmentation methods and annotation strategies.

Section 3 will introduce the notion of *dimensions* in dialogue analysis, and the theoretical framework that we will use in the rest of the paper, that of Dynamic Interpretation Theory (DIT), in particular the DIT$^{++}$ taxonomy of communicative functions. The paper ends with conclusions in Section 6.

[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.*

ACCEPTED MANUSCRIPT

## 2 Dialogue segmentation and the syntax of multifunctionality

Multifunctionality comes in a variety of forms. Allwood (1992) distinguishes two forms of multifunctionality, called *sequential* and *simultaneous*, and gives the following example to illustrate the distinction:

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

Sequential multifunctionality occurs when an utterance has several parts which each have a different communicative function. As Allwood observes: "A's utterance in the example 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." It should be noted that the term 'utterance' is used here in the sense of[2]

(3) "*unit in spoken dialogue which corresponds to a stretch of speech from one speaker, bounded by lack of activity or another communicator's activity.*" (Allwood, 1992)

Utterances in this sense, also called *turns*, can be quite complex, and so it is no wonder that they are often sequentially multifunctional. For this reason, smaller functional units within turns are often considered, alternatively also referred to as 'utterances'.

Precise definitions of 'utterance' are hard to come by, but utterances are commonly understood to be either single words or syntactically well-formed (contiguous) phrases or clauses (Allwood speaks of 'grammatical units') which can be seen as the linguistic realization of a speech act. Segmenting a dialogue into utterances has the advantage of being more fine-grained than a segmentation into turns, and thus allowing a more precise functional markup; on the other hand, the determination of utterance boundaries is a highly nontrivial task. Syntactic and prosodic features are often used as indicators of utterance endings (e.g. Shriberg et al., 1998; Stolcke et al., 2000; Nöth et al., 2002), but are in general not very reliable. In the case of nonverbal or multimodal communication, the notion of an utterance as a linguistically defined unit is even less clear.

The occurrence of sequential multifunctionality depends on the way in which a dialogue is segmented, and disappears if sufficiently small segments of dialogue are considered. Example (2) illustrates this: if A's first utterance is segmented into the following five utterances:

(4) 1. Yes!
2.. Come tomorrow.

[2] A problematic aspect of this definition is that it presupposes the occurrence of periods of inactivity; in natural multimodal dialogue, however, where the participants do not only speak but also use facial expressions, gestures, body posture, and gaze direction for communicative purposes, it can be argued that there are no such periods - for instance, one is always looking somewhere and one always has some facial expression. This can be remedied by modifying the definition so that it speaks of 'lack of *speech activity*' rather than 'lack of activity'.

ACCEPTED MANUSCRIPT

3. Go to the church.
4. Bill will be there.
5. OK?

then each of the smaller segments no longer has any sequential multifunctionality.

While segmenting a dialogue into utterances rather than turns has the effect of eliminating sequential multifunctionality, simultaneous multifunctionality may still occur, not at the level of turns but at that of utterances. The example of turn-initial stalling by *Let me see...* in dialogue (1), while also indicating a turn-taking act, is one illustration; another, mentioned by Allwood (1992), is that of a warning which is also an inform.

There are still other forms of multifunctionality that do not fit into the sequential-simultaneous dichotomy, namely *discontinuous*, *overlapping*, and *interleaved multifunctionality*. The first of these occurs when an utterance embeds a smaller utterance which has a different communicative function. The following example illustrates this.

(5) 1. C: What time is the first train to the airport on Sunday?
2. I. The first train to the airport on Sunday is at... *let me see...* 5.32.

Here we see a discontinuous answer *The first train to the airport on Sunday is at [.....] 5.32* to the preceding question. Segmentation by splitting up the turn into utterances in the sense discussed above, is problematic in such a case. Breaking up the turn into the parts (1) *The first train to the airport on Sunday is at*; (2) .. *let me see...* ; and (3) *5.32* would work well for the second and third part, which satisfy both the requirement of grammatical well-formedness and which can both be seen as realizations of speech acts, but the first part violates both requirements.

Example (5) also illustrates the phenomenon of *overlapping* multifunctionality, which occurs when part of an utterance with a certain function forms a sub-utterance with another function. In the example, the sub-utterance *The first train to the airport on Sunday* has the function of providing positive feedback on the understanding of the question, while the utterance as a whole answers the question. Again, splitting up the turn into a sequence of functionally relevant grammatical units would not work, since we would now be stuck with the chunk *is at*.

*Interleaved* multifunctionality occurs when two utterances with different functions are interleaved to form a complex utterance, and is illustrated by the following example.

(6) 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?* Splitting up the turn into a sequence of utterances again would not work since we would get some strange chunks, such as *I think twenty five euros for a remote*.

These examples show that the segmentation of dialogue into utterances in the usual sense does not always lead to distinguishing the stretches of behaviour that form functional units. Instead, such units should be allowed to be discontinuous, to overlap, and

ACCEPTED MANUSCRIPT

to be interleaved. To avoid terminological confusion, we introduce the term *functional segment* for this purpose, defined as follows:

(7) **Definition**. A functional segment is a minimal stretch of communicative behaviour that has a communicative function. Such stretches do not need to be grammatically well-formed or contiguous, and may have more than one communicative function.

Note that this definition contains a condition of 'minimality'; this is because whenever a certain stretch of dialogue has a certain communicative function, then an extension of that stretch also has that function; this condition ensures that functional segments are not longer than necessary.

Since a functional segment does not need to be contiguous, segmenting a dialogue into functional segments not only allows segments to be discontinuous parts of a turn and to be interleaved, but also to be spread over multiple turns. The following example shows an occurrence of this phenomenon:

(8)
A; Could you tell me what departure times there are flights to Frankfurt on Saturday?
B: Yes, let me have a look. OK, There's a Lufthansa flight leaving at 07:45,
A: yes,
B: and a KLM flight at 08:15,
A: yes,
B: and then there's a flight by Philippine airlines...
A: .....

In this example the answer to A's question to consists of a list of items which B communicates one by one in separate turns (of which the example only shows the first three items), in order not to overload A. The functional segment corresponding to the answer is thus *There is a Lufthansa flight at 07:45, and a KLM flight at 08:15, and then there's a flight by Phillippine airlines...*, spread over as many turns as there are items in the list. Note that, if functional segments were required to always be internal to a single turn, we would have the problem of saying what is the communicative function of a segment like *and a KLM flight at 08:15*. This would require the introduction of additional communicative functions like 'part of an answer', which would complicate the design of a taxonomy of dialogue acts considerably.

Note that functional segments can be very small, like *em* or *yes*, and quite long, like *Could you tell me what departure times there are flights to Frankfurt on Saturday* and the multi-turn answer to that in (8). The use of functional segments resolves various segmentation-related analysis problems, such as what to do in case a functional unit contains a part which has another (or the same) communicative function, or in case a functional unit is interrupted by a unit that has another function. For the study of multifunctionality in dialogue, the most important effect of using functional segments as the unit of analysis is that sequential multifunctionality is eliminated and that only simultaneous multifunctionality remains at the level of functional segments. This is a

ACCEPTED MANUSCRIPT

truly minimal approach to multifunctionality that avoids all artifacts that may be due to segmentation.

One of the main claims of this paper is that the phenomenon of simultaneous multifunctionality in dialogue is best understood in terms of a multidimensional approach to dialogue analysis. The next section will introduce the idea of multiple dimensions in dialogue analysis, which will be applied in the rest of the paper for studying the multifunctionality in dialogue both qualitatively and quantitatively.

# 3 Dimensions in dialogue analysis

## 3.1 Function clusters and dimensions

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

Multidimensional schemas support dialogue utterances to be coded with multiple tags, and make the multifunctionality of dialogue utterances explicit. Such schemas typically have a relatively large tag set, and a structured organisation of such a tag set has several advantages:

- Clustering semantically related tags improves the transparency of the tag set, as the clusters indicate the kind of semantic information that is considered. 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.

- The introduction of dimensions in a tag set naturally leads to a hierarchical or lattice structure in the set, which 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 of tags, and subsequently more fine-grained distinctions may be tested to decide on a tag within the cluster. 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 either mutually exclusive (such as 'signal understanding' and 'signal non-understanding'), or are are related by an entailment relation (such as a 'confirmation' also being an 'answer'). In both cases, an annotator should choose only one tag from the cluster (in the latter case the most specific tag for which there is sufficient evidence). In this way the organisation of the tag set supports annotators in avoiding the consideration of inconsistent or irrelevant combinations of tags.

ACCEPTED MANUSCRIPT

- 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. The choice of a particular cluster of tags can typically be made on the basis of less information than that of a particular tag inside the cluster. The same is true for choosing a more general tag from the cluster versus a more specific tag (e.g. 'answer' versus 'disconfirmation'). Human annotation is often more detailed than automatic annotation because of the difference in semantic information that is effectively available. This means that automatic and human annotation are often not identical but still may be compatible, which can be expressed and measured precisely by taking the semantic relations within a cluster into account for computing annotator-agreement scores (Geertzen & Bunt, 2006).

An annotation schema is usually not just a tag set, but comes with instructions or guidelines for using the tags. A multidimensional approach to annotation can therefore be reflected in a schema in two ways: in the tag set, by structuring it into clusters of tags, and allowing annotators to pick one tag from each cluster; or in the guidelines, instructing annotators how to choose and apply multiple tags from a flat list of tags, and specifying the restrictions that apply for tag combinations. If the tag set is fairly extended and does not have any structure, then it is next to impossible to formulate good instructions for how to use the tags in multiple tagging, since there is no easy way to refer to groups of tags. Therefore, the recognition of the claim that utterances in dialogue tend to be multifunctional leads almost directly to the introduction of structure in the tag set.

The well-known DAMSL schema (DAMSL = Dialogue Act Markup using Several Layers) is organized into 'layers' and 'dimensions'. Four layers are distinguished: Communicative Status, Information Level, and Forward and Backward Communicative Functions (FLF and BLF); the latter two are indeed clusters of communicative functions (the tags in the other layers are concerned with other kinds of information). The FLF cluster is subdivided into five clusters, including (roughly) the classes of commissive and directive functions, well known from speech act theory. The BLF cluster has 4 subclasses: Agreement, Understanding, Answer, and Information Relation. Core & Allen (1997) refer to these eleven subclasses as 'dimensions'. While the DAMSL documentation does not discuss or motivate the choice of layers and dimensions, these are clearly useful for structuring the tag set in a way that can help annotators to make their choices, supported by annotation guidelines which make use of this structure in the decision-making process. The DAMSL dimensions within the FLF and BLF clusters form mutually exlcusive sets of tags, and make DAMSL a nine-dimensional schema.

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

Bunt (2005; 2006) suggests to structure a tag set for multidimensional annotation by using a well-founded notion of *dimension*, based on the observation that participa-

ACCEPTED MANUSCRIPT

tion in a dialogue involves several activities beyond those strictly related to performing the task or activity for which the dialogue is instrumental. In natural conversation, among other things, dialogue participants constantly "evaluate whether and how they can (and/or wish to) continue, perceive, understand and react to each other's intentions" (Allwood, 2000). They share information about the processing of each other's messages, elicit feedback, and manage the use of time and turn allocation, of contact and attention, and of various other aspects. We propose to define dimensions as corresponding to such aspects of communication, in sofar as they are identifiable and separable in communicative behaviour.

These considerations lead to the following definition of dimension:

(9) **Definition.** A dimension is an aspect of participating in dialogue which:

1. dialogue participants can address by means of dialogue acts;
2. can be addressed independently of the other aspects of participating in dialogue which are distinguished.

The first of these conditions means that only aspects of communication are considered that can be distinguished according to empirically observable behaviour in dialogue. The second condition requires dimensions to be independent (or 'orthogonal'). What is meant here by the independence of dimensions is that the communicative functions which a segment can have in one dimension are not, or at least not fully, determined by its functions in other dimensions. 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 at most as many tags as there are dimensions, one function for each dimension (namely the most specific function for which there is sufficient evidence to mark it).

Note that each of the dimensions as targeted in definition (9) constitutes a particular type of communicative activity, and the dialogue acts involved in these types of activity are concerned with different types of information: feedback acts are concerned with the success of processing previous utterances; turn management acts are concerned with the allocation of the spekaer role, task-related acts are concerned with performing the underlying task or activity; and so on. Each dimension therefore corresponds with a particular type of semantic content of the dialogue acts involved.

A multidimensional annotation schema has a set of functions for each of its dimensions, and these sets are often conceived to consist of mutually exclusive tags. This means that, once a certain tag has been assigned to a segment, all other tags in the same dimension do not need to be considered. In cases where one communicative function is a specialization of another, such as a warning being a special kind of inform and a check being a special kind of question, one would like to include both the general and the more specific function in the same dimension. These tags are not mutually exclusive; on the contrary: every segment that is encoded with the more specialized tag could also be encoded with the less specialized one.

ACCEPTED MANUSCRIPT

Multiple markup within a dimension can be avoided, however, by stipulating in the annotation guidelines that in the case of related functions with different degrees of generality/speciality the most specific functions should always be encoded for which there is sufficient evidence. To make this possible, we formulate the following requirement for the set of functions that can be used in a given dimension:

(10) **Functional dependence principle:** Any two communicative functions that can be used for addressing a given dimension are either mutually exlcusive, i.e. if one of them applies then the other one does not; or one is a specialization of the other.

A set of dimensions that satisfies the requirements (9) and (10) explains and optimally supports the assignment of multiple tags to functional segments, as it allows each segment to have as many functions as there are dimensions: one function at most for each aspect of communication. Sometimes it is possible to infer from these functions, in a given dialogue context, that a segment has additional comunicative functions. This phenomenon is analysed in Section 4.

An annotation schema with dimensions that each have a clear relation to a particular aspect of communication has many advantages. First, annotators have a conceptual interpretation of the dimensions of the annotation task in terms of aspects of communication. Second, an annotator or analyst who is specifically interested in certain aspects of communication can choose to use only the corresponding dimensions of the schema. Third, an annotator or analyst who is especially interested in an aspect that is not covered by the schema can add a dimension, provided that it satisfies the condition of orthogonality with respect to the other dimensions.

Petukhova & Bunt (2009a; 2009b) provide an up to date survey and analysis of the dimensions that occur in 18 existing annotation schemas. Three criteria are formulated that a proposed dimension should satisfy in order to meet the requirements of definition (9):

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

1. theoretically justified, in the sense that it is a well-established and well-studied aspect of communication;
2. empirically observed in the functions of dialogue utterances;
3. addressable independently of the other dimensions.

The second an third criterion are applied in a corpus-based examination of dimensions that have been used in these annotation schemas. The results of this study provide support for distinguishing the 10 dimensions of the DIT$^{++}$ annotation schema, that is presented in the next section.

## 3.2 Dynamic Interpretation Theory and DIT++

The semantic framework of Dynamic Interpretation Theory (DIT, see Bunt, 1989; 1990; 1994; 2000; 2009) ) takes a multidimensional approach to dialogue in the sense,





Accepted Manuscript

ACCEPTED MANUSCRIPT

mentioned above, that participation in a dialogue is viewed as performing several activities in parallel, such as pursuing a 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 consist of the performance of dialogue acts, which are formally interpreted as update operations on certain aspects of the context models (or 'information states')[3]; 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 context model with the semantic content when he understands the corresponding aspect of the meaning of a dialogue utterance.

DIT distinguishes the following 10 dimensions:

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.

Petukhova and Bunt (2009a) performed a corpus-based investigation to determine the independent addressability of these dimensions, as required in definition (9) and in (11). Three spoken dialogue corpora were used in the study, namely the AMI corpus of multimodal multi-party dialogues[4] the OVIS corpus of human-computer telephone dialogues about train schedules (see Section 5.1), and the DIAMOND corpus of simulated

[3] 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'. 'Information state', the most commonly used term in the literature, is not an entirely satisfactory description of the state of a dialogue participant, since such a state includes besides the participant's beliefs also goals, plans, expectations, attitudes like preferences and fears, social pressures, and a record of the dialogue history.

[4] **A**ugmented **M**ulti-party **I**nteraction (`http://www.amiproject.org/`)

ACCEPTED MANUSCRIPT

Table 1: *Overview of dimensions being addressed without any other dimension also being addressed in AMI, OVIS and DIAMOND data, expressed in relative frequency in percentages.*

human-helpdesk dialogues (see Section 5.1). The results of these study are given in Table 1, and show that each of the DIT$^{++}$ dimensions is indeed addressable independent of the other dimensions.

| |
|---|
| *Information Transfer Functions* |
| *Information-seeking functions* |
| *Questions* |
| propositional question, set question, |
| alternatives question, check question, etc. |
| *Conditional questions* |
| indirect propositional question, set question, |
| alternatives question |
| *Information-providing functions:* |
| *Informing functions:* |
| inform, agreement, disagreement, correction; |
| *Informs with rhetorical functions such as:* |
| elaboration, explanation, justification,... |
| *Answer functions:* |
| propositional answer, set answer, confirmation, |
| disconfirmation |
| *Action Discussion Functions* |
| *Commissives* |
| offer, promise, address request, address suggestion |
| other commissives, expressable by means of performative verbs |
| *Directive functions:* |
| instruct, request, address offer, conditional request, suggestion |
| other directives expressable by means of performative verbs, |
| such as advice, permission,encouragement, urge,..., |

Table 2: *Classes of general-purpose communicative functions in DIT$^{++}$.*

DIT has been developed as a framework for the analysis of human dialogue (see e.g. Bunt et al., 2007; Petukhova & Bunt, 2009c), in particular for providing a formal and computational semantics for dialogue utterances (see Bunt 1995; Morante, 2007; Petukhova & Bunt 2008), and for designing dialogue systems (Bunt, 1996; Keizer & Bunt, 2006; 2007). Concerning its implementation in interactive systems, DIT has been the basis of designing the PARADIME dialogue manager of a multimodal dialogue system for information extraction (Op den Akker et al., 2005). The DISCUS system has been implemented to explore the context update mechanisms defined for dialogue acts in the DIT$^{++}$ taxonomy (Morante & Keizer, 2006; Morante, Keizer and Bunt, 2007). These update mechanisms form the basis of a computational model of grounding (Bunt, Morante and Keizer, 2007). Finally, the DIT$^{++}$ taxonomy plays a role in the design of

ACCEPTED MANUSCRIPT

| *Dimension* | *Dimension-specific functions* | *Representative expressions* |
|---|---|---|
| Task/Activity | OpenMeeting, CloseMeeting; | domain-specific |
| | Appoint, Hire, Fire | 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 utterance |
| Discourse Structure Man.t | DialogueActAnnouncement | *Question.* |
| | TopicShiftAnnouncement | *Something else.* |
| Social Obligations Man. | Apology | *I'm sorry.* |
| | Greeting | *Hello!, Good morning.* |
| | Thanking | *Thanks.* |

Table 3: *Examples of dimension-specific communicative functions and representative expressions for each dimension.*

a functional markup language for embodied conversational agents ; see Bunt (2009).

One of the products of DIT is a taxonomy of communicative functions, called the DIT$^{++}$ taxonomy, that was designed for the purpose of dialogue act annotation and dialogue system design,[5] and which integrates elements from various other annotation schema, such as the DAMSL, TRAINS, and Verbmobil taxonomies (Allen & Core, 1997; Allen et al., 1994; Alexandersson et al., 1998) into the DIT schema (see Bunt, 1994; 2000; 2009).

Since a dimension in dialogue analysis, as defined in the previous section, is concerned with a particular aspect of interacting, or in other words with a particular type of information concerning the interaction, like the 10 dimensions listed above, it may be observed that communicative functions like Question and Answer do not belong to any dimension. This is because one can ask questions not only about something in the task, but also about agreeing to close a topic, or about whose turn it is to say something, or about any other aspect of interacting, so a dialogue act with the function Question can belong to *any* these dimensions, depending on the type of information that it has as its semantic content. Of course every occurrence of a dialogue act with a Question function falls within one of the dimensions, depending on the type of information that is asked. Similarly for answers, statements, requests, offers, agreements, (dis-)confirmation, and so on. Clusters of such general types of dialogue acts, which belong to what in speech act theory is sometimes called *core speech acts*, therefore do not constitute a dimension (as is the case in DAMSL) and do not belong to any dimen-

[5]For more detail see `http://dit.uvt.nl`.

ACCEPTED MANUSCRIPT

sion, 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 Release, Stalling, Introduce Topic, and Apology. The DIT$^{++}$ taxonomy therefore consists of two parts: (1) a taxonomy of *general-purpose functions*; and (2) a taxonomy of *dimension-specific functions*.

The general-purpose functions are divided into four categories: information-seeking functions, which include a range of question types; information-providing functions, which include plain statements, statements with a rhetorical function, and a range of answer types; commissives, and directives. The functional dependence principle (10) that holds for dimensions also holds for the class of general-purpose functions as a whole (the four main categories being mutually exclusive) and for each of the four categories separately. For example, within the category of information-providing functions Agreement and Disagreement are mutually exclusive specializations of Inform, whereas Correction is the single specialization of Disagreement.

Table 2 shows the classes 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. This means that a context model should not only contain information about the task domain and the current state of the task that underlies the dialogue (and beliefs about the sharing of such information with other dialogue participants), but also information about the allocation of the speaker role and other participant roles, about the success of processing previous utterances, and about social obligations or pressures. Table 3 lists these information types, which have been studied in more detail in Bunt (1996; 1999; 2000); and illustrates their use by dialogue utterances whose update semantics involves these types of information.

In the European project LIRICS[6] a subset containing 60 of the most important and widely used communicative functions from the DIT$^{++}$ taxonomy was cast in the form of definitions following ISO standard 12620 (for 'data categories'). This tag set (see LIRICS, 2007a) was endorsed by an ISO expert group (TC 37/SC 4/TDG 3) and was evaluated by annotating test suites in English and Dutch by multiple annotators. The oucome of this evaluation was that an inter-annotator agreement was found, expressed in terms of the standard kappa metric, of 0.93 without significant differences between English and Dutch (see LIRICS, 2007b). This means that these communicative functions can be used by human annotators with near-perfect agreement. Geertzen (2009) reports encouraging results of machine learning experiments aimed at applying the same tag set in automatic annotation.

## 4 Semantic types of multifunctionality

We noted above that the multifunctionality of stretches of dialogue is reduced to a minimum when we take functional segments, as defined in (7), as the unit of analysis.

[6] See `http://lirics.loria.fr`

ACCEPTED MANUSCRIPT

| *example utterance* | *dialogue act type* | *information category* |
|---|---|---|
| *Can I change the contrast?* | 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 4: *Semantic information categories as related to dialogue act types, and example utterances.*

A certain stretch of communicative behaviour can have a communicative function for any of two reasons:

1. It has properties which, in the context in which it occurs, are indicators of that function. Such properties may be lexical/idiomatic, syntactic, prosodic, or non-verbal, in the case of multimodal interaction;

2. Some functions imply other functions; therefore a stretch of behaviour which has one communicative function also has another.

The following examples illustrates the two cases:

(12) a. A: 1. Is this the intercity to Eindhoven?
B: 2. Yes it is.

b. C: 1. Do you know where Louise is?
D: 2. The secretaries have their annual outing today.

In (12a) the utterance by B contains lexical indicators of the function of an answer to A's question. Now one of the semantic characteristics of an answer is that the speaker provides certain information $p$ in response to a wish to have that information (or its negation, in the case of a propositional question) - a wish that was understood from a previous question. An answer therefore always implies that the speaker believes to have understood the corresponding question - so an answer implies positive feedback. This feedback function can thus be associated with the answer utterance by implication, rather than as an effect of the properties that the utterance has.

In the case of (12a) one might object to this analysis and claim that the feedback function of the utterance can be derived directly from its properties. We have seen that



Accepted Manuscript

ACCEPTED MANUSCRIPT

*yes* is often an indicator of positive feedback - however, in the present example *yes* has the function of an answer to a 'yes-no' question, for which it is also an indicator. Example (12b), moreover, shows that an answer utterance does not need to have any properties which indicate its feedback function. D's utterance 2 constitutes an answer to C's question if Louise is a secretary and D believes this knowledge to be shared with C. D's utterance in this case doesn't have any properties that would indicate its feedback function. In general, functional segments with an answer function have a positive feedback function which can be inferred from the answer function, not from the properties of the segment.

Notice that, while a functional segment may have communicative functions for two reasons, it has always at least one these functions because of its function indicators. It may have multiple functions because it contains indicators for more than one function and/or because an indicated function implies another one. In the first case we speak of *independent multifunctionality*, in the second case of *implied multifunctionality*. The latter can be subdivided into three categories, depending on the kind of implicative relation into *logically entailed*, *conversationally implicated*, and *indirect*. We discuss each of these types of multifunctionality in turn in the next subsections.

## 4.1 Independent multifunctionality

Independent multifunctionality occurs when a functional segment has several, logically independent communicative functions in different dimensions, due to containing indicators of more than one such function.

One may expect this type of multifunctionality not to occur very frequently; many combinations of communicative functions cannot be encoded in a single functional segment, since their respective indicators are mutually exclusive. For example, the interrogative syntactic mood is an indicator of all information-seeking functions, whereas the imperative mood is a prime indicator of all directive functions - it is therefore next to impossible to encode an information-seeking and a directive function in one segment. In Section 5 we will see if the expectation is empirically confirmed that independent multifunctionality is not a frequent phenomenon. Implied multifunctionality, by contrast, does not require encoding, and may therefore be expected to occur abundantly.

Examples of independent multifunctionallity are the following:

1. *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; example (8) illustrates this.
2. *Thank you*, spoken with markedly high pitch and cheerful intonation (like goodbyes often have), to signal goodbye in addition to gratitude. (—it Thank you may additionally also signal positive feedback as well as pre-closing of the dialogue - see below, Section 4.2.2.)
3. Turn-initial Stalling and Turn Take (or Turn Accept), as illustrated by turn 2 in dialogue fragment (1); turn-internal Stalling and Turn Keep, as illustrated by turn 2 in example (5).

ACCEPTED MANUSCRIPT

4. An excessive amount of turn-internal Stalling and elicitation of support (i.e., eliciting an utterance completion act in the Partner Communication Management dimension), as in the following example:
   A: *The code that we need to specify here ...ehm .. ehm .. that's ... ehm ..*
   B: *that's the group code, CW183.*

Semantically, the interpretation of a segment which displays independent multi-functionality 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 Implied communicative functions

So far, we have relied on the reader's intuitions for interpreting what was meant by 'implied communicative functions'. Before going into the various kinds of implication relations among communicative functions, we first give a proper definition of functional implication.

Implications among communicative functions have their basis in implication relations among *dialogue acts*. An implication relation holds between two dialogue acts $A_1$ and $A_2$ if the context update operation that forms the interpretation of $A_1$ subsumes the update operation that forms the interpretation of $A_2$. What is meant by subsumption here is the following. Given a context model $M$, the semantic interpretation of dialogue act $A_1$ is a set $|A_1|_M$ of changes to be made to $M$. Similarly for $A_2$. In the case where $F_1$ = Warning and $F_2$ = Inform, and in all similar cases where $F_1$ is a specialization of $F_2$, the set of changes $|A_2|_M$ is simply a subset of the set of changes $|A_1|_M$, but there are also cases where one of the changes in $|A_1|_M$ is not *additional* to the set of changes $|A_2|_M$, but where it is logically stronger than a corresponding change in $|A_2|_M$. For instance, a request by $S$ to $A$ to do some action $\alpha$, if understood by $A$, leads to the change (1) *A believes that S wants A to perform* $\alpha$; by comparison, the corresponding indirect request leads to the change (2) *A believes that S wants A to perform* $\alpha$ *if A agrees*. Clearly, (1) is stronger than (2). In general, what is meant by set of changes $\Sigma_1$ subsuming $\Sigma_2$ is that every element of $\Sigma_2$ is a logical consequence of the set of premises $\Sigma_1$. Formally:

(13) **Definition.** A dialogue act $A_1$ implies a dialogue act $A_2$ iff for any context model $M$: $|A_1|_M \supset |A_2|_M$, i.e., every $p \in |A_2|_M$ is deducible from $|A_1|_M$.

With this definition, we can now say that a communicative function $F_1$ implies a function $F_2$ if the dialogue act formed by applying $F_1$ to a certain semantic content implies the act constructed by applying $F_2$ to the same or a closely related semantic content (what is meant by 'closely related' will be made clear below). Formally:

(14) **Definition.** A communicative function $F_1$ implies a communicative function $F_2$ iff for any semantic content $\gamma$, $F_1(\gamma)$ implies the dialogue act $F_2(\gamma)$.

In these definitions we have on purpose used the term 'implies' without being formal about what is meant by this term. We in fact distinguish three semantically different types of implication, as already indicated above: (1) logical entailment, which is

ACCEPTED MANUSCRIPT

a consequence of the meaning of the types of communicative functions involved; (2) implicatures, which correspond to default inferences that can be made unless the communicative functions involved occur in an (unusal) context where the implicature is canceled; (3) indirectness, in the sense of indirect speech acts, which require reasoning with the semantic content of a dialogue act and elements from the context. The distinction between these kinds if implications is semantically important, since they have different consequences for the update operations for interpreting functional segments that display implied multifunctionality.

#### 4.2.1 Entailed functions

In the case of an entailment relation, a functional segment has a communicative function $F_1$ due to its features, and this function implies the communicative function $F_2$ in the sense of definition (14), with "implies" interpreted as logical consequence. However, two types of entailed communicative functions have to be distinguished, depending on whether the $F_1$ and $F_2$ dialogue acts share the same triggering condition. We first consider the case where they do.

Entailment relations often exist between dialogue acts within the same dimension which have the same semantic content but communicative functions that differ in their level of specificity, more specific dialogue acts entailing less specific ones. Consider the following example:

(15) It's nearly five o'clock.

In response to the question what time it is (a Set Question, also known as WH-Question), this is an answer (more precisely: a Set Answer). Alternatively, it might be a Warning, as when the speaker knows that the addressee should leave no later than five o'clock because of another appointment; as yet another alternative, it could be a Justification for someone who is leaving a meeting before it's over, because he has a train to catch. Answer, Warning and Justification are three of the many specializations of the Inform function, and every segment that has one of the specialized functions by entailment also has the Inform function. This type of intra-dimension entailment relation has also been called *functional subsumption*. When two communicative functions which have a functional subsumption relation are applied to the same semantic content, then an entailment relation exists between the resulting dialogue acts. A warning that it's raining hard is also an inform that it's raining hard. Note that the entailing and the entailed act share the same triggering condition, namely that the speaker wants the addressee to know that it's raining hard.

Dialogue acts in different dimensions are concerned with different aspects of the interaction, and therefore with different types of information. There is therefore mostly no entailment relation or other semantic relation between them, but an exception is the relation between acts in non-feedback dimensions on the one hand and auto- and allo-feedback acts on the other. This type of entailment occurs for dialogue acts which respond to another dialogue act from another participant, such as accepting or rejecting an offer, a suggestion, an invitation, or a request; answering a question; responding to a greeting, accepting an apology, and so on. The relation between a responsive

ACCEPTED MANUSCRIPT

dialogue act (or act with a 'backward-looking function' (Allwood, 2000; Allen & Core, 1997) and the act that it responds to is called a *functional dependency relation*, and is part of the annotations that we will consider in Section 5. This relation is of obvious importance for determining the semantic content of the responding act. Moreover, 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 dialogue act that he responds to, and so the occurrence of a responsive dialogue act entails a positive (auto-)feedback act. This is the second type of entailment relation mentioned above, where the entailing and entailed acts do not share the same triggering condition. A positive feedback act has the triggering condition that the speaker wants the addressee to know that the speaker believes to have successfully processed the dialogue segment expressing the dialogue act to which the entailing act has a functional dependency relation (see Figure 2). For instance, an answer to a question entails that the speaker believes to have understood the question. However, can the answerer of a question be said to have the goal to inform the addressee of his understanding of the question? Since This seems very doubtful not obvious, hence we preferable not to make this assumption,

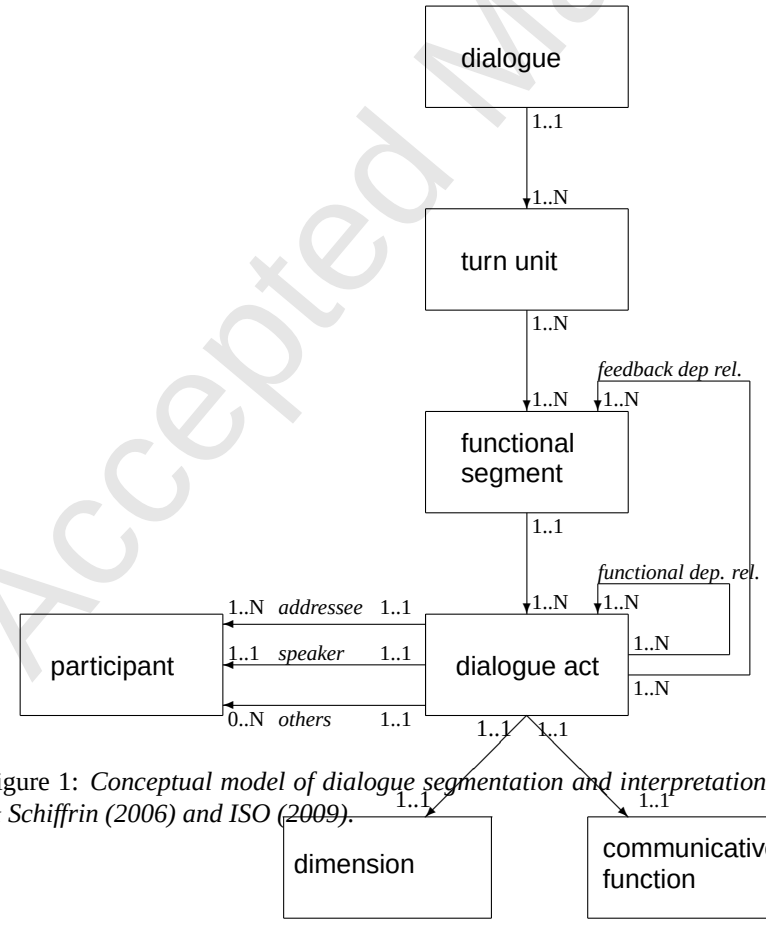



Figure 1: *Conceptual model of dialogue segmentation and interpretation. After Bunt & Schiffrin (2006) and ISO (2009).*

ACCEPTED MANUSCRIPT

Feedback acts have a different kind of 'backward-looking' function than other responsive acts, since feedback refers to the processing of stretches of communicative behaviour rather than to the dialogue acts that were realized by them. They therefore have a backward-looking relation to stretches of behaviour; this relation is called an *feedback dependency relation*. Figure 1 shows the role of this relation and that of functional dependency relations in a conceptual model of dialogue segmentation and analysis in the form of a UML diagram. This model was originally developed in the LIRICS project (see Bunt & Schiffrin, 2006) and more recently refined in ISO project 24617-2 "Semantic annotation framework, Part 2: Dialogue acts", which aims at establishing an international standard for dialogue act annotation (see ISO, 2009). The model shows that conceptually a dialogue is segmented into 'turn units' (see below) and further into functional segments. The notation '$k...N'$ at the head of an arrow in this notation means that an entity at the root of the arrow may be associated with any number between $k$ and $N$ of entities at the head, and similarly for the numerical indication at the root of an arrow. A turn unit may thus comprise any number of functional segments (zero, if it consists of a stretch of behaviour that forms part of a multi-turn functional segment). A *turn unit* is the stretch of communicative behaviour that constitutes a turn, roughly in the sense of (3), but slightly modified in order to do justice to the observation that in natural, multimodal dialogue (1) speakers hardly have periods of total inactivity, but they do have periods of speech inactivity (cf. footnote 2); (2) turns may overlap a lot – see e.g. Campbell (2008), which makes the clause saying that a turn may be *"bounded by activity of another speaker"* unrealistic. Our definition of 'turn units' is therefore as follows:

(16) **Definition**. A *turn unit* is a stretch of communicative behaviour produced by one partipant, bounded by lack of speech activity by this participant.

Note that this definition allows turn units produced by different participants to overlap.

Functional segments according to the model correspond to one or more dialogue acts. Each dialogue act has one speaker and at least one addressee, and possibly a number of other participants (such as bystanders, overhearers, or eavesdroppers, in Clark's terminology – see Clark, 1996). Each dialogue act has a communicative function and a dimension. Moreover, each dialogue act may have a functional dependency relation to other dialogue acts, and may have an feedback dependency relation to one or more turn units.

Using the concepts from the model depicted in Figure 1, we can represent the relation between a responsive dialogue act and the entailed feedback act as in Figure 2. A responsive dialogue act DA2 has a functional dependency relation to a previous dialogue act DA1, which was realized by a functional segment FS1. This segment was part of turn unit TU1. The responsive dialogue act then entails a feedback act FB1, which has an feedback dependency relation with the segment FS1 (the scope of the feedback act). So in the case of an entailed feedback act, the semantic content of that act is not related to that of the entailing act, but to the stretch of communicative behaviour that can be found by retracing the arrows in Figure 2 from the dialogue act that the entailing act responded to.

Some examples of entailment relations between dialogue acts are:

ACCEPTED MANUSCRIPT

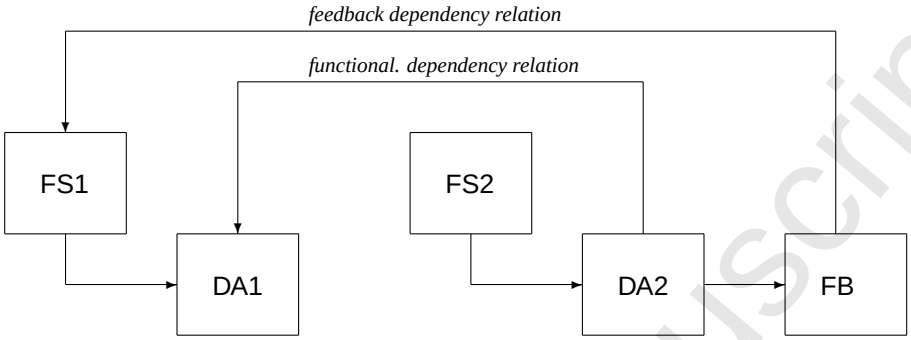



Figure 2: Entailed feedback for responsive dialogue acts

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 (Init Greeting, Apology);

4. Partner Communication Management acts entailing positive feedback acts; for instance, a Correct-misspeaking act, performed when the speaker is thought to say or have said something wrong, presuppose that this participant believes to have understood what the speaker was saying; the same goes for a Completion act.

Note that for the first kind of example we have spelled out the relation between the semantic content of the entailing act and the entailed act in (**??**) as being the identity relation; for the feedback examples of type 2 and type 3 we have described how the semantic content of the entailed act can be computed according to Figure 2. The case of example 4 can be treated in the same way, if we view a Partner Communication Management act can as responding to what the partner is currently saying.

#### 4.2.2 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

ACCEPTED MANUSCRIPT

also another communicative function due to the occurrence of a conversationally implicated act. Like all conversational implicatures, this phenomenon is context-dependent, and implicates functions are intended to be recognized. Examples are:

1. an expression of thanks implicating positive feedback at all levels of the previous utterance(s) of the addressee;
2. an expression of thanks indicating a Pre-closing act, i.e. indicating that, as far as the speaker is concerned, the dialogue can be closed.
3. positive feedback implied by shifting to a new topic, related to the previous one; more generally, by any relevant continuation of the dialogue;
4. 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 the implicating dialogue act) in a given context. Implicated functions are intended to be recognized, and correspond semantically to an additional context update operation, hence they are a true source of multifunctionality.

Two important, frequently occurring cases of implicated multifunctionality are that of so-called 'indirect speech acts' and that relating to the different levels of processing that may be addressed by feedback acts. These two cases are discussed separately in the next two sections.

#### 4.2.3 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 analyzed as expressing both a question about the addressee's abilities and, indirectly, a request to pass the salt. Evidence in support of this analysis is that it is quite common to respond to such an indirect request by saying *yes*, which seems to be an answer to the literal interpretation of the question, and subsequently carrying out the request – or by saying *no* and apologizing, or explaining why one is unable or unwilling to comply with the request.

Using DIT or another semantic approach, based on context update, 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*. Moreover, 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 inconsistencies when human dialogue participants are modeled, but the inconsistency between two beliefs of the form *S believes that A is able to do X* and *S wants to know whether A is able to do X* is too blatant to be acceptable in any context model.

ACCEPTED MANUSCRIPT

| utterance | interpretation |
|---|---|
| Do you know what time it is? | Please tell me what time it is, if you know. |
| Do you have a light? | Please give me a light, if you have one. |
| Have you seen Louise? | Please tell me where Louise is, if you have seen her. |
| Can you reach the light switch? | Please operate the light switch, if you can reach it. |
| Do you know if there are any flights to Toronto this evening? | Are there any flights to Toronto this evening, if you know? |
| Are there any flights to Toronto this evening? | Which flights to Toronto are there this evening, if any? |
| Can you pass me the salt? | Pass me the salt please. |
| I would like to have some coffee. | Please give some coffee, if you have any. |

Table 5: *Indirect questions and requests interpreted as conditional questions and requests.*

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* or *Please tell me what time it is*, respectively. On this analysis, the indirectness is considered as just a matter of politeness, not having any semantic consequences. Evidence in favour of this analysis is that there clearly are situations where it is quite common to ask an indirect question like *Can you tell me from which platform the train to Tilburg leaves?* even though one does assume that the addressee possesses this knowledge.

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 that S wants 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 and indirect requests, as Table 5 shows. So this type of 'indirect speech act' is viewed not as expressing multiple acts, but as expressing a single, conditional dialogue act.

Another kind of indirect speech act, which does not address a precondition for another speech act, is exemplified by *I would like to have some coffee.* This might be analyzed as an inform act, and indirectly as a request. The DIT analysis of such cases is as follows. Speaker S has a goal $G$ which could be achieved by successful performance of a dialogue act with communicative function $F_2$ (such as Request). The utterance is interpreted as the request to A to perform the $F_2$ act if A is able and willing to do so.

ACCEPTED MANUSCRIPT

Hence again, the utterance is viewed not as expressing two dialogue acts, but rather as a single, conditional one.

To the extent that functional segments realizing indirect dialogue acts can be analyzed as 'indirect' but conventional expressions of a single dialogue act, such as a conditional request, this is not a source of multifunctionality. Other cases, where an 'indirect' function can be determined only with the help of inference using context information, are in fact cases of 'particularized' conversational implicatures.

#### 4.2.4 Entailed and implicated feedback functions

A speaker who provides feedback about his perception, understanding, or evaluation of previous utterances, or, in the terminology introduced above, performs an auto-feedback act, 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 mostly a signal that the speaker is not sure he heard well, whereas a rephrasing of what was said is not concerned with perception but with 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 (or at least, the speaker whose feedback addresses the level of understanding does so with the assumption that there was no problem at the perceptual level).

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

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

'Evaluation' should be understood here in relation to the information-state update approach and the requirement that information states at all times remain internally consistent when update operations are applied to them. 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 check whether $p$ is consistent with his current state; if not; the update would be unacceptable. Evaluation leads to a positive result if the intended update operation is acceptable, and may be signaled by a positive feedback act referring to this level; a negative result will typically lead to a negative feedback signal. If the evaluation has a positive outcome, then the recipient can move on to the stage of execution, which is the highest level of processing of an input. For the example of the informing act with content $p$, execution would mean that the recipient inserts $p$ in his information state.

When the input is a question, then the evaluation comes down to deciding whether the input can be accepted as such, e.g. does not conflict with the belief that this particular question has already been answered. Its 'execution' is then the gathering or computation of the information needed to answer the question. If execution fails, this typically leeds to a response like *I don't know*, which is a negative feedback act at execution level.

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

ACCEPTED MANUSCRIPT

feedback at all levels $L_j$ where $i > j$; positive feedback at execution level therefore entails positive feedback at all other levels. By contrast, positive feedback at level $L_i$ *implicates* negative feedback at all levels $L_j$ where $i < j$; for instance, a signal of good perception implicates that there is a problem with understanding, for why not signal good 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.

| *feedback polarity* | *proc. level* | *implic. relation* | *feedback polarity* | *proc. level* | *level relation* |
|---|---|---|---|---|---|
| positive | $L_i$ | entails | positive | $L_j$, | $L_i > L_j$ |
| positive | $L_i$ | implicates | negative | $L_j$, | $L_i < L_j$ |
| negative | $L_i$ | entails | negative | $L_j$, | $L_i < L_j$ |
| negative | $L_i$ | implicates | positive | $L_j$, | $L_i > L_j$ |
| elicitation | $L_i$ | entails | positive allo-fb. | $L_j$, | $L_i > L_j$ |

Table 6: *Entailments and implicatures between feedback acts at different levels of processing.*

For negative feedback the entailment and implicature relations work in the opposite direction from positive feedback. For allo-feedback providing functions the same relations hold as for auto-feedback. For feedback eliciting acts, which form a subclass of the acts in the allo-feedback dimension, the situation is that an act which elicits feedback at level $L_i$ implicates (since there is no logical necessity) that the addressee had no processing problems at lower levels, hence it implicates positive allo-feedback *providing* acts at all levels $L_j < L_i$. Figure 3 summarizes these 'rules'.

The implicated feedback acts can be explained by the Cooperation Principle (Grice, 1975), in particular from the Maxim of Quantity that derives from it. If a speaker has for instance successfully understood a previous utterance, then it would not be optimally informative to say that he perceived the utterance well. Applying the Maxim of Quantity to the levels of processing says that one should provide positive feedback at the highest level of successful processing, and negative feedback at the lowest level where processing problems were encountered.

Implied feedback functions do not really constitute a separate kind of implied functions, but we distinguish them here and in the annotation strategies considered below because of there virtually ubiquitous character.

# 5 Empirical determination of multifunctionality

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

ACCEPTED MANUSCRIPT

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

(ii) the annotation strategy that is used with respect to the markup of implied functions.

## 5.1 Experiment

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

(1) dialogues over a microphone and a head set with a WOZ-simulated helpdesk, providing assistance in the use of a fax machine (from the DIAMOND corpus[7]);

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

(3) human-computer telephone dialogues about train schedules (from the OVIS corpus); [8]

(4) Dutch Map Task dialogues without visual contact between the participants.

We consider the effects of three alternative segmentation methods:

**a. turn-based:** the turn unit, as defined in (16), is taken as the unit which is annotated with communicative functions;

**b. utterance-based:** communicative functions are assigned to contiguous utterances, as defined in Section 2;

**c. functional-segment based:** communicative functions are assigned to all functional segments.

The dialogues were in fact segmented into functional segments and annotated accordingly; from this segmentation and annotation we reconstructed the annotation that would correspond to the coarser other two segmentation methods.

The following annotation strategies were compared for dealing with the various possible sources of (simultaneous) multifunctionality:

**a. strictly indicator-based:** only communicative functions are marked which are recognizable directly from features of the stretch of communicative behaviour, given the context of the preceding dialogue. In particular, only explicit feedback functions are marked, and Turn Management functions are marked only if they are explicitly indicated through lexical and/or prosodic features. (No nonverbal features were available for this corpus.)

**b. + implicated functions:** implicated functions are are marked as well, except those that occur between feedback acts addressing different levels of processing;

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

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

ACCEPTED MANUSCRIPT

**c. + turn taking:** a turn-initial segment (i.e., a functional segment occurring at the start of a turn unit) 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 treated 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:** feedback functions which are entailed by non-feedback functions are also marked, such as the positive feedback on understanding that is entailed by answering a question or accepting an offer;

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

**g. + implicated feedback levels:** signals of feedback at a certain level of processing are also marked as feedback signals at other levels, as specified in Figure 3;

**h. + entailed feedback levels:** signals of feedback at a certain level of processing are also marked as feedback at other levels, as specified in Figure 3;

**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 a and c-i were reconstructed by deleting (for strategy a) or adding the relevant implied, indirect or default functions.

## 5.2 Results

The average number of annotated communicative functions per functional segment found in the experiment are presented in Table 6. As the annotated dialogue corpus used in this study was marked up according to strategy b, it includes besides the communicative functions derived directly from function indicators (in context) also the implicated ones, except implicated functions at various feedback levels (which are taken into account in strategy g). The entailed functions that are additionally annotated when strategies c-f and h are applied, can all be derived automatically from the annotations resulting from strategy b.

The positive and negative feedback functions at certain levels of processing which are implicated by a feedback function at another level, and that are taken into account in strategy g, cannot be deduced from the strategy-b annotations, but these implicated functions can be assumed to occur by default, as they seem to always occur except in some unusual circumstances.[9]

[9]Such unusual circumstances may for example be that one is received by the king of an very traditional country with an extremely strict hierarchical political system, where the king is never to be asked to clarify or repeat what he said.

ACCEPTED MANUSCRIPT

Indirect communicative functions, which are additionally taken into account in strategy i, cannot be deduced from strategy-b annotations in a straightforward way, but require a good understanding of the dialogue context. However, we have argued above that it is questionable whether indirect speech acts should be treated as the occurrence of *both* a direct and an indirect act, and therefore that it can be argued that indirect speech acts do not really add to the multifunctionality of dialogue segments.

| *segmentation method:* / annotation strategy: | *turn unit* | *utterance* | *funct'l. segment* |
|---|---|---|---|
| a. strictly indicator-based | 2.5 | 1.7 | 1.3 |
| b. + implicated functions | 3.1 | 2.1 | 1.6 |
| c. + turn taking | 4.0 | 2.7 | 2.1 |
| d. + turn releasing | 4.8 | 3.3 | 2.6 |
| e. + entailed feedback | 5.2 | 3.6 | 2.8 |
| f. + subsumed functions | 5.6 | 3.9 | 3.0 |
| g. + implic. feedb. levels | 6.3 | 4.2 | 3.2 |
| h + entailed feedb. levels | 6.6 | 4.5 | 3.4 |
| i. + indirect functions | 6.9 | 4.7 | 3.6 |

Table 7: *Average multifunctionality of dialogue units for various annotation strategies and segmentation methods.*

## 5.3 Discussion

Three important limitations should be noted of the results in Table 6. First, the data are from speech-only dialogues, where there was no nonverbal interaction between the participants other than by the use of nonverbal sounds, such as laughs, sighs, and chuckles. Second, only dialogues with two participants are considered. These two limitations may both be expected to have a decreasing influence on the multifunctionality that is found. For the speech-only limitation this is because in full-blown multimodal communication a speaker can for instance use facial expressions to indicate a feedback elicitation act or a social obligations management act (like an apology), additional to the functions that are indicated through verbal means (see e.g. Petukhova & Bunt, 2009d). For the two-party limitation this is because in multi-party dialogue turn management is much more important than in two-party situations (see e.g. Petukhova & Bunt, 2009a), and occurs much more frequently (Petukhova & Bunt, 2009b). Especially when neither of these limitations apply we can expect for instance Turn Assign acts to frequently be performed in co-occurrence with other acts, by the speaker directing the gaze to the participant that he targets (Petukhova & Bunt, 2009c). The third and crucial limiting factor on the data is that they are based on the use of a particular tag set. Of course, the figures all depend on the richness of the sets of communicative functions and dimensions in the annotation schema; on the other hand, DIT$^{++}$ is probably the richest dialogue act tag set around, so the use of this set is not 'limiting' in the sense that using other schemas would lead to higher multifunctionality figures.

Due to these limitations and to the relatively small sample of annotated dialogue

ACCEPTED MANUSCRIPT

material on which they are based, most of the absolute figures in this table are not of great interest; more interesting are the differences that we see depending on the segmentation method that is used and on the annotation strategy.

The figures for turn-based segmentation and annotation are the least reliable, since the complexity of the turn units that one finds in dialogue may vary a lot, depending on the kind of dialogue and the roles of the participants. By contrast, functional segments as defined in this paper are highly independent of the kind of dialogue that is considered (except maybe for segments of purely nonverbal behaviour), so these figures may be considered to be reliable for the given tag set.

The first and most important thing to notice in Table 5.2 is that, when the most fine-grained segmentation is applied, thereby eliminating all forms of sequential multifunctionality and retaining only simultaneous multifunctionality, and the most conservative annotation strategy, which leaves out all kinds of implicated, entailed, and indirect functions, then the multifunctionality found in the annotations reflect the *independent multifunctionality* of functional segments. We see that the average independent multifunctionality of a functional segment is 1.3, hence on average one in every three segments has two independent communicative functions. The figures in boldface in the second row of Table 5.2 represent the *minimal degree of multifunctionality* that is found, taking implicated functions into account. (But leaving implicated and entailed feedback functions are various levels out of consideration.) These figures represent the most conservative way of determining the average multifunctionality of functional segments.

Another interesting point to observe, going down in the rightmost column in Table 6, is that the increments in multifunctionality that are incurred by using more generous annotation strategies are relatively small. On the most generous strategy we find a multifunctionality of 3.6.

The difference with the multifunctionality of *utterances* is caused by the fairly frequent discontinuity of functional segments. This is mainly caused by the occurrence of Own Communication Management acts, where the speaker edits his contribution on the fly and interrupts his utterance by stallings, retractions, restarts, and so on.

The multifunctionality of a *turn unit* is simply the sum of the simultaneous multifunctionalities of its constituent functional segments. The figures in Table 6 for unsegmented turn units show that in our corpus a turn unit on average contains one and a half contiguous utterances and nearly two functional segments. These figures may vary depending on the type of dialogue. For instance, in a meeting conversation where one participant is very dominant and produces long turns, alternated by occasional short turns from other participants, the number of utterances per turn will on average be greater. In general, the figures in the column for utterance-based segmentation have to be taken with a big grain of salt, as they depend a lot on the complexity of the turns in the dialogues that are considered.

What do these results finally say about the degree of multifunctionality that we find in spoken two-party dialogue? The ordering of annotation strategies going down in Table 6 reflects an increasingly generous approach to semantic annotation, which corresponds only in part with an increasingly generous view on multifunctionality. In particular, when performing annotation it makes sense to not encode those functions which can be inferred from already encoded ones. This applies to entailed functions

ACCEPTED MANUSCRIPT

and to those implicated functions that can be considered as default implications.

The issue of whether or not to annotate a Turn Take or Turn Accept function (depending on the context) every time a participant starts speaking, and to annotate a Turn Release function every time a speaker stops speaking, is also partly a matter of economy in the annotation process. Such functions could be considered as entailments from the starting/stopping speaking behaviour of participants, and it would therefore be redundant to annotate them. And even worse: we don't want annotate a Turn Keep function every time a speaker *continues* to speak... Every small stretch of speech could be claimed to have a turn management function, since a speaker who is not inactive is either starting, continuing, or ceasing to speak, but clearly this would be a rather fruitless approach from the point of view of annotation. We would suggest that turn management acts are more fruitfully defined as acts which a speaker performs with the purpose of signaling that he wants to obtain, to release, or to assign the turn, and that cases where a speaker starts asking a question, making an offer, etc. do not have that purpose, and neither is the case when a speaker continues to speak in order to express another dialogue act. Note also that for *automatic* annotation there would be no objection against adding all the entailed or default functions.

In view of the conceptual analysis of various types of implied functions in section 4, the functions that are encoded in the annotation strategies a-i contribute to multifunctionality as follows:

**a. strictly indicator-based annotation:** all functions that are annotated are communicative functions of the annotated segment. The multiplicity of the annotation tags shows the independent multifunctionality of functional segments.

**b. + implicated functions:** implicated functions, except those that occur between feedback acts addressing different levels of processing, are additional functions and add to the multifunctionality of segments.

**c. + turn taking:** according to the analysis above, it seems best not to consider starting to speak as always carrying a Turn Take or Turn Accept function, but different views on this are certainly possible.

**d. + turn releasing:** similar to c., although speakers may deliberately use a silence to release the turn. The view that ceasing to speak by default expresses a Turn Release act is therefore somewhat more plausible than the corresponding view on default turn taking;

**e. + entailed feedback functions:** we have argued in Section 4.2.1 that from a semantic (context-update) point of view, feedback functions which are entailed by non-feedback functions are additional functions of the segment under consideration, and add to its multifunctionality.

**f. + subsumed functions:** we have argued that the entailed functions within a dimension, due to degrees of specificity, should not be viewed as *additional* functions of a segment. They should hence not be considered as adding to multifunctionality.

**g. + implicated feedback levels:** being a case of conversational implicature, the implicated positive or negative (auto- or allo-) feedback functions at other levels of processing than a level-specific explicit feedback act may well be considered to be additional functions of the segment, and thus to add to its multifunctionality.

ACCEPTED MANUSCRIPT

**h. + entailed feedback levels:** this case is conceptually similar to that of functional subsumption; hence these functions are perhaps best considered as not adding to multifunctionality.

**i. + indirect functions:** we have suggested in Section 4.2.3 that in the case of indirect speech acts, the indirect and the direct interpretation of the segment should not both be considered as functions of the segment, and so the phenomenon of indirectness does not contribute to multifunctionality.

None of the classes of functions that are encoded according to the strategies b-i can be said unequivocally to not add to the multifunctionality of functional segments, but we can order them according to their importance for grasping the full meaning of a functional segment, according to the above analysis. This corresponds to reordering Table 6 as shown in Table 7.

| *segmentation method:* / annotation strategy: | *turn unit* | *utter-ance* | *funct'l. segment* |
|---|---|---|---|
| a. only indicator-based functions | 2.5 | 1.7 | 1.3 |
| b. + implicated functions | **3.1** | **2.1** | **1.6** |
| g. + implicated feedback levels | 3.8 | 2.4 | 1.8 |
| e. + entailed feedback | 4.1 | 2.7 | 2.0 |
| d. + default turn releasing | 4.9 | 3.3 | 2.5 |
| c. + default turn taking | 5.8 | 3.9 | 3.0 |
| i. + indirect functions | 6.1 | 4.1 | 3.2 |
| f. + subsumed functions | 6.4 | 4.4 | 3.4 |
| h + entailed feedback levels | 6.9 | 4.7 | 3.6 |

Table 8: *Average multifunctionality of functional segmentation.*

Implicated functions, both feedback functions at certain levels of processing implicated by functions at other levels of processing and other implicated functions, being logically independent of the functions that implicate them, must be considered to be additional functions of functional segments. It therefore seems justifiable to claim that on average a functional segment has at least 1.8 communicative functions. We have argued in Section 5 that entailed feedback functions are also additional, from a semantic point of view, which would bring the multfunctionality to 2.0; however, this could conceivably be avoided by defining a semantic interpretation of the 'feedback dependency relations', introduced in the model in Figure 1, to the effect that the semantics of backward-looking functions takes the understanding of the 'antecedent' utterances into account. The default interpretation of stopping speaking as constituting a Turn Release act is intuitively also quite reasonable, more so than interpreting starting speaking as constituting a Turn Take or a Turn Accept act. Anyway, in speech-only two-party dialogue the multifunctionality of a functional segment, which is arguably the most sensible unit to assign a communicative function to, is somewhere between 1.8 and 3.6.

ACCEPTED MANUSCRIPT

# 6 Conclusions and future work

One conclusion that follows from this study is that the multifunctionality of stretches of dialogue is a persistent phenomenon, no matter what segmentation method is used. Taking the most-fine-grained segmentation of dialogue into functional units, thereby avoiding all forms of segmentation-related multifunctionality, and a minimal approach to the notion of multifunctionality, we find that on average two out of every three units have more than one independent communicative function. In addition, at least half of the functional segments have an entailed or implicated function.

Why are dialogue utterances multifunctional, regardless how the term 'utterance' is understood? The phenomenon of multifunctionality can be explained by considering 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; managing time, and so on. This approach has been backed up by empirical data, which show that functional segments display both what we called independent multifunctionality, having two or more functions in different dimensions due to their function indicators, as well as implicated and entailed multifunctionality.

The amount of multifunctionality that is found in dialogue depends on three main factors. First, the choice of units in dialogue which are considered as having communicative functions matters a lot. If turns are taken as units, then there is not much that can sensibly be said, due to the fact that turns may be quite complex, and therefore display quite a lot of sequential multifunctionality. Second, regardless of the choice of functional units, we have seen that the observed amount of multifunctionality depends strongly on the view that is taken on what counts as having multiple functions, and on the role that is given to implied, default, and indirect functions. Third, the choice of analysis framework and annotation schema imposes obvious limits on the multifunctionality that can be expressed.

Finally, any adequate account of the meaning of dialogue utterances will have to take their multifunctionality into consideration. Our findings confirm that the multifunctionality of functional segments can be viewed as arising only due to their meaning *in different dimensions*: a segment can be viewed as never having more than one function in any given dimension. An update semantics can thus be defined which interprets communicative functions as recipes for updating a part of the context model by means of separate updates for each dimension, which, due to the independence of dimensions, can be performed independently of each other.

## Acknowledgements

I would like to thank the members of the Tilburg Dialogue Club who have over the years been involved in the development of DIT: Jeroen Geertzen, Hans van Dam, Ielka van der Sluis, Simon Keizer, Roser Morante, Amanda Schiffrin, and in particular Volha Petukhova.

# References

[1] Ahrenberg, L., N. Dahlback & A. Jonsson (1995) Codings schemas for Studies of



Accepted Manuscript

ACCEPTED MANUSCRIPT

Natural Language Dialogue. In: Working Notes from the AAAI Spring Symposium, Stanford.

[2] Akker, R. op den, H. Bunt, S. Keizer and B. van Schooten (2005) From question answering to spoken dialogue: towards an information search assistant for interactive multimodal information extraction. In *Proceedings of the Ninth European Conference on Speech Communication and Technology, Interspeech 2005*, Lisbon, pp. 2793-2797.

[3] 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.

[4] 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.

[5] 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.

[6] Allwood, J. (1992). *On dialogue cohesion.* Gothenburg Papers in Theoretical Linguistics 65. Gothenburg Univrsity, Department of Linguistics.

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

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

[9] Bos, J., E. Klein, O. Lemon, and T. Oka (2003). DIPPER: Descriptin and formalization of an information-state update dialogue system architecture. In *Proceedings 4$^{th}$ SIGdial Workshop on Discourse and Dialogue*, pp. 115 – 124.

[10] Bunt, H. (1989). Towards a dynamic interpretation theory of utterances in dialogue. In H. Bouma and B. Elsendoorn (eds) *Working models of human perception.* New York: Academic Press, pp. 419 – 456.

[11] Bunt, H. (1990). DIT – Dynamic Interpretation in Text and Dialogue. In L. Kálmán and L. Pólos (eds) *Papers from the Second Symposium on Logic and Language.* Budapest: Akademiai Kiadó.

[12] Bunt, H. (1994). Context and Dialogue Control. *Think Quarterly* 3 (1), 19–31.

[13] Bunt, H. (1995). DIT – Dynamic interpretation and dialogue theory. In M.M. taylor, F.Néel, and D.G. Bouwhuis (eds) *Proceedings from the Second Venaco Workshop on Multimodal Dialogue*. Amsterdam: Benjamins.

ACCEPTED MANUSCRIPT

[14] Bunt, H. (1996). Interaction management functions and context representation requirements. In *ProceedingsTwente Workshop on Dialogue Management in Natural Language Systems.* Enschede: Universiteit Twente

[15] Bunt, H. (1999). Context Representation for Dialogue Management. In P. Bouquet, L. Serafini, P. Brézillon, M. Beneceretti, and F. Castellani (eds) *Modeling and Using Context.* Berlin: Springer.

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

[17] 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`

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

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

[20] Bunt, H., R. Morante and S. Keizer (2007) An empirically based computational model of grounding in dialogue. In *Proceedings SIGDIAL 2007, 8th Workshop on Discourse and Dialogue,* Antwerp, September 2007..

[21] Bunt, H., and A. Schiffrin (2006). Methodological aspects of semantic annotation. In *Proceedings Fifth International Conference on Language Resources and Evaluation LREC 2006*, Genova.

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

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

[24] Clark, H.H. (1996) *Using Language.* Cambridge, UK: Cambridge University Press.

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

[26] Core & Allen (1997) Coding Dialogs with the DAMSL Annotation schema. *AAAI Fall Symposium on Communicative Action in Humans and Machines*, Boston, MA.

[27] Dhillon, R., Bhagat, S., Carvey, H., and Schriberg, E. (2004). Meeting recorder project: dialogue labelling guide. ICSI Technical Report TR-04-002.



Accepted Manuscript

ACCEPTED MANUSCRIPT

[28] Di Eugenio, B. Jordan, P.W., and Pylkkaenen, L. (1998). *The COCONUT project: dialogue annotation manual*. ISP Technical Report 98-1 (1998).

[29] Geertzen, J. (2009) *The automatic recognition and prediction of dialogue acts.* Ph.D. Thesis, Tilburg Universsity.

[30] Geertzen, J. & H. Bunt (2006) Measuring annotator agreement in a complex, hierarchical dialogue act schema. In *Proc. SIGDIAL 2006*, Sydney.

[31] Grice, H.P. (1975) Logic and conversation. In *Syntax and semantics, Vol 3: Speech acts.* New York: Seminar Press.

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

[33] Keizer, S. and H. Bunt (2006) Multidimensional dialogue management. In *Proc. SIGDIAL 2006, 7th Workshop on Discourse and Dialogue,* Sydney, pp. 37 - 45.

[34] Keizer, S. and H. Bunt (2007) Evaluating combinations of dialogue acts. In *Proc. SIGDIAL 2007, 8th Workshop on Discourse and Dialogue,* Antwerp, Septmeber 2007, pp. 158 – 165..

[35] Klein, M., N.O. Bernsen, O. Davies, L. Dybkjaer, J. Garrido, H. Kasch, A. mengel, V. Pirelli, M. Poesio, S. Quazza, and C. Soria (1998). Supported Coding Schemes. EU Project MATE Deliverable D1.1.

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

[37] LIRICS (2007b). *Multilingual test suites for semantically annotated data*. eContent Project LIRICS, Deliverable D4.4.

[38] LIRICS (2007a). *Documented set of semantic data categories*. eContent Project LIRICS, Deliverable D4.3.

[39] Morante, R. (2007). *Computing meaning in interaction.* Ph.D. Dissertation, Tilburg University.

[40] Morante, R. and S. Keizer (2006). DISCUS: a dialogue simulation and context update system. In *Proceedings $10^{th}$ Workshop on the Semantics and Pragmatics of Dialogue (BRANDIAL)*, Potsdam, pp. 169 – 170.

[41] Morante, R., S. Keizer and H. Bunt (2007). A dialogue act based model for context updating. In *Proceedings of the $11^{th}$ Workshop on the Semantics and Pragmatics of Dialogue (DECALOG)*, Rovereto, pp. 9 – 16.

[42] 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.

ACCEPTED MANUSCRIPT

[43] Petukhova, V. and H. Bunt (2008) Towards a multidimensional semantics of discourse markers in spoken dialogue. In *Proceedings IWCS-8, 8$^{th}$ International Conference on Computational Semantics*. Tilburg, January 2008.

[44] Petukhova, V. and H. Bunt (2009a) *Dimensions of communication.* Technical Report 2009-003, Tilburg centre for Creative Computing, Tilburg University.

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

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

[47] Petukhova, V. and H. Bunt (2009d) Grounding by nodding. In *Proceedings of GESPIN, Conference on Gestures and Speech in Interaction*, Poznán, September 2009.

[48] Poesio, M. and D. Traum (1997) Conversational actions and discourse situations. *Computational Intelligence* 13(3), pp. 309 – 347.

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

[50] 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.

[51] 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.

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

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

[54] Traum, D., and E. Hinkelman (1992). *Conversation Acts in Task-oriented Spoken Dialogue*. Computational Intelligence, 8(3), pp. 575–599.

[55] Traum, D., and S. Larsson (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.



Accepted Manuscript