Multifunctionality and multidimensional dialogue semantics

Harry Bunt
Tilburg Center for Creative Computing
Tilburg, the Netherlands
harry.bunt@uvt.nl

Abstract

This paper addresses the following questions: (1) Is it true, as is often claimed, that utterances in dialogue tend to have multiple functions? (2) If so, then what are the reasons for that? (3) How many functions does a dialogue utterance typically have, and which factors determine this? (4) What consequences does this have for the computational semantics of dialogue utterances? Answers to these questions are sought by investigating a dialogue corpus annotated with communicative functions using various segmentation and annotation strategies.

1 Introduction

Traditional approaches to the analysis of sentence meaning notoriously fail when applied to dialogue utterances. 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 (ISU) or context-change (CC) approaches Traum & Larsson, 2003) , which view utterance meanings in terms of changes in the information states (or 'contexts') of the dialogue participants. These approaches closely relate to the ideas of speech act theory, which regard the use of language as the performance of communicative actions.

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

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

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

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

This example illustrates that dialogue utterances often do not correspond to a single speech act, but to sets of speech acts. Moreover, some of these speech act types, such as feedback acts and turn-taking acts have hardly 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 dialogue contexts 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 2).

[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.*

Proceedings of the 13th Workshop on the Semantics and Pragmatics of Dialogue, June 24–26, 2009, Stockholm, Sweden.

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 as well: in the first part of B's utterance 2 the speaker signals that he is stalling for time through the use of the expression *Let me see* and slowing down; by implication 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 3 we will discuss the issue of implied functions in more detail, as well as other reasons why dialogue utterances often have multiple functions.

In the literature, claims about the multiple functionality of dialogue utterances are often motivated by isolated examples like (1), rather than by quantitative studies of corpus data; moreover, the claimed multifunctionality of utterances is highly dependent on what is meant by 'utterance', as well as by the spectrum of communicative functions that is considered. In Section 3 we will discuss the definition of 'utterance' in the light of segmenting a dialogue into meaningful units, and in Section 2 we will introduce a rich, well-motivated taxonomy of communicative functions for the analysis in the rest of the paper. In Section 4 we discuss the various ways in which one dialogue act may imply another. Section 5 is devoted to an empirical study of the multifunctionality of utterances in a dialogue corpus, and Section 6 ends the paper by summarizing the answers to the questions that were raised in the abstract.

2 Theoretical framework

2.1 Dialogue acts and utterance meanings

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

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

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

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

Information Transfer Functions
information-seeking functions
Direct Questions
propositional question, set question, alternatives question, check question, etc.
Indirect Questions
indirect propositional question, set question, alternatives question, check question, etc.
information-providing functions:
informing functions:
inform, agreement, disagreement, correction; *informs with rhetorical functions such as:*
answer functions:
propositional answer, set answer, confirmation, disconfirmation
Action Discussion Functions
Commissives
offer, promise, address request
other commissives, expressable by means of performative verbs
Directive functions:
instruction, address request, indirect request, (direct) request, suggestion
other directives, such as advice, proposal, permission, encouragement, urge,..., expressable by means of performative verbs

Table 1: Structure of the DIT^{++} taxonomy of general-purpose communicative functions.

10. Social Obligations Management: dialogue acts that take care of social conventions such as greetings, apologies, thanking, and saying goodbye.

One of the products of DIT is a multidimensional taxonomy of communicative functions, called the DIT^{++} taxonomy, designed for the purpose of dialogue act annotation and dialogue system design across a wide range of domains,[3] and which includes 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). Multidimensional taxonomies support dialogue utterances to be coded with multiple tags and have a relatively large tag set; such a tag set may benefit in several respects from having some internal structure.

First, clustering semantically related tags improves the transparency of the tag set for human users, as the clusters indicate the kind of semantic information that is addressed. Second, introducing a hierarchical or taxonomical structure which is based on semantic clustering may support the decision-making process of human annotators: an initial step in such a process can be the decision to consider a particular cluster, and subsequently more fine-grained distinctions may be tested in order to decide on a specific tag within the cluster. Third, a hierarchical organisation in the tag set may also be advantageous for automatic annotation and for achieving annotations which are compatible though not identical with those of human annotators (namely, the automatic annotation may use less specific tags than the human annotation). In general, a structured tag set can be searched more systematically (and more 'semantically') than an unstructured one, and this can clearly have advantages for dialogue annotation, interpretation, and generation.

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

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

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

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

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

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

Table 2: Examples of dimension-specific communicative functions and representative expressions for each dimension.

When we view a dimension in dialogue analysis in accordance with (2) as a particular aspect of interacting, like the 10 dimensions mentioned above, we see that dialogue acts like question and answer do not belong to any dimension. This is because one can ask a question about something in the task, or a about agreeing to close a topic, or about whose turn it is to say something, or about any other aspect of interacting, so questions can belong to *all* these dimensions. Every *occurrence* of a question function, as the function of a dialogue act that is performed, falls within one of the dimensions; which dimension is determined by the type of semantic content. Similarly for answers, statements, requests, offers, agreements, (dis-)confirmation, and so on. Clusters of such general types of dialogue acts therefore do not form a dimension, but can be used in any dimension; they are called *general-purpose functions*. This in contrast with communicative functions that are specific for a particular dimension, such as Turn Keep, Turn Release, Introduce Topic, Change Topic, Apology and Thanking. The DIT^{++} taxonomy therefore consists of two parts: (1) a taxonomy of *general-purpose functions*; (2) a taxonomy of *dimension-specific functions*. Table 1 shows the structure of the taxonomy of general-purpose functions; Table 2 lists examples of dimension-specific communicative functions in each of the DIT^{++} dimensions.

In order to define a context-change semantics for all the types of dialogue acts in the DIT^{++} taxonomy, the context models on which the semantics is based should contain all the types of information addressed by these dialogue acts.

Table 3 lists these types, and illustrates their use by dialogue utterances whose update semantics involves these types of information.

3 Multifunctionality and segmentation

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

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

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 *"unit in spoken dialogue which corresponds to a stretch of speech from one speaker, bounded by lack of activity or another communicator's activity."* Utterances in this sense, which are more commonly called *turns* are often quite complex, and it is no wonder that they are often sequentially multifunctional. It is therefore more common to consider smaller functional units within turns, and refer to these units as 'utterances', as we shall also do in the rest of this paper.

Utterances in the latter sense are defined as contiguous stretches of linguistic behaviour which form grammatical units that have a communicative function. 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 (as opposed to turn 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.

Segmenting a dialogue into utterances has the effect of eliminating sequential multifunctionality. There are however other, segmentation-related forms of multifunctionality that remain, 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.

(4) 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. Example (4) 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.

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

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

Here we see the discontinuous statement *I think twenty five euros for a remote [...] is too much money to buy an extra remote or a replacement one* interleaved with the discontinuous question *is that locally something like fifteen pounds [...] or is it even more in pounds?* These examples show that the segmentation of dialogue into utterances in the usual sense does not lead to distinguishing the stretches of behaviour that form functional units. Instead, such units should be allowed to be discontinuous, to overlap, and to be interleaved. To avoid terminological confusion, we use the term *functional segment* for this purpose (see further Geertzen et al., 2007).[4]

4 Types of multifunctionality

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

independent: a functional segment has more than one communicative function, due to having features expressing each of these functions;

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

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

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

We discuss each of these types of multifunctionality in turn.

[4] A functional segment may also spread over multiple turns, as the following example shows:
A; Could you tell me what departure times there are for flights to Frankfurt on Saturday?
B: Certainly. There's a Lufthansa flight leaving at 08:15,
A: yes,
B: and a KLM flight at 08:50,
A: yes,
B: then there's a flight by Philippine airlines,...
In this example the A's question to consists of a list of items which B communicates one by one in separate turns in order not to overload A.

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

Table 3: Semantic information categories as related to dialogue act types, and example utterances.

4.1 Independent multifunctionality

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

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

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

4.2 Implied communicative functions

4.2.1 Entailed functions

It was noted in Section 1 that one of the reasons why utterances may have multiple functions, is that one function may imply another. The two implication relations that we see in example (1) above are of a different nature. The turn-taking act that is implied by the first part of utterance 2 follows from the fact that there is a stalling act in turn-initial position; the feedback act implied by the answer in the second part of 2 follows from the fact that giving an answer presupposes understanding the corresponding question. The latter case corresponds to a logical entailment relation between answers and positive feedback acts, whereas the former is context-dependent, and more like a conversational implicature.

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

Some examples of entailment relations between dialogue acts are:

1. Justification, Exemplification, Warning all entailing Inform; Agreement, Disagreement, Correction entailing Inform; Confirmation and Disconfirmation both entailing Propositional Answer; Check Question entailing Propositional Question;
2. Answer, Accept Offer, Reject Offer, Accept Suggestion, Reject Suggestion entailing positive feedback;
3. Responsive dialogue acts for social obligations management, such as Return Greeting and Accept Apology entailing positive feedback on the corresponding intiating acts (such as Init Greeting and Apology);
4. Evaluative feedback entailing positive feedback on perception and understanding; Negative feedback on perception entailing negative feedback on understanding (see below, Section 4.4).

Entailment relations typically occur between dialogue acts within the same dimension, and which have the same semantic content but communicative functions that differ in their level of specificity. More specific dialogue acts entail less specific ones with the same semantic content. Dialogue acts in different dimensions are concerned with different aspects of the interaction; therefore with different types of information, and hence there is usually no relation of entailment or other semantic relation between them.

Entailed functions within the same dimension correspond to the context update operation representing the entailed interpretation being subsumed by the update operation of the entailing one. They are thus semantically vacuous, and it therefore does not seem to make much sense to consider such cases as multiple functions that can be assigned to a functional segment.

Entailments may also occur also between an act in a non-feedback dimension and a feedback act. An answer, for example, is semantically related to a question, which has been expressed in a preceding utterance or sequence of utterances contributed by the dialogue partner. Relations such as the one between an occurrence of an answer and the corresponding question, are called *functional dependency relations*[5], and are part of the annotations in the corpora that we will consider in Section 5. This type of relation is relevant for answers, responses to directive dialogue acts (such as Accept Request and Reject Offer), and more generally to those dialogue acts that have a 'backward-looking function' (Allwood, 2000; Allen & Core, 1997), for which the functional dependency relation indicates the dialogue act that is responded to. 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.

Entailed feedback acts corresponds to context-changing effects in the component of the context model that contains the speaker's assumptions about his own and his partner's processing of previous utterances. These context-changing effects are additional to those that express the semantics of the entailing responsive act, and should therefore be considered as adding an extra communicative function to the corresponding utterance.

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

4.3 Implicated functions

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

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

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

4.4 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 typically 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:

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

'Evaluation' should be understood here in relation to the information-state update approach followed in DIT, and the requirement that information states at all times be internally consistent, also 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 viewed as 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 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.

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

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.

4.5 Indirect speech acts

The phenomenon known as 'indirect speech acts' is another potential source of multifunctionality. An utterance such as *Can you pass me the salt?* has been analysed as expressing both a question about the addressee's abilities and, indirectly, a request to pass the salt. Using DIT or another semantic, ISU-based approach, such an analysis does not make much sense, however, since a request to do X is normally understood to carry the assumption (on the part of the speaker, S) that the addressee (A) is able to do X; hence the interpretation of the utterance as a request would lead to an update of the context to the effect that A believes that S believes that A is able to pass the salt, while 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, resulting in an inconsistent information state.

The DIT analysis of such cases is as follows. S has a goal G that could be achieved by successful performance of a dialogue act with function F_1; however, F_1 has a precondition p_1 of which S does not know whether it is satisfied, and which S believes A knows whether it is satisfied (for instance, a property of A). S therefore asks A whether p_1. A understands this situation (in fact, S and A mutually believe this situation to obtain), and 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*. Similarly, an utterance like *Do you know what time it is?* is understood as *Please tell*

me what time it is, if you know, and *Are there any flights to Toronto this evening?* as *Which flights to Toronto are there this evening, if any?* 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 is exemplified by *I would like to have some coffee.* This might be analysed as an inform act, and indirectly 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. Hence again, the utterance is viewed not as expressing two dialogue acts, but rather as a single, conditional one.

Whether all types of indirect speech act can be analysed in a similar way, as corresponding to a single conditional dialogue act rather than to multiple acts, is an issue for further research. If the answer is positive, then indirect speech acts are in fact not a source of multiple functionality. If the answer is negative, or if the DIT analysis is not adopted, then it is.

5 Empirical determination of multifunctionality

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

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

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

5.1 Experiment

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

(1) dialogues over a microphone and head set with a WOZ-simulated helpdesk, providing assistance in the use of a fax machine (from the DIAMOND corpus[6]);

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

(3) human-computer telephone dialogues about train schedules (from the OVIS corpus); [7]

(4) Dutch Map Task dialogues.

We compared three alternative segmentation methods:

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

b. utterance-based: every turn is chopped up into contiguous, non-overlapping grammatical units which have one or more communicative function;

c. functional-segment based: functional segments are distinguished for each (possibly discontinuous) stretch of behaviour which has one or more communicative function, where functional segments may be discontinuous, overlapping, and interleaved, and may spread over more than one turn.

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

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

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

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

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

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

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

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

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

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

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

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

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

5.2 Results

The results are summarized in Table 2. The absolute figures in this table are not of great interest, given the small sample of annotated dialogue material on which they are based; relevant are especially the differences that we see depending on the segmentation method that is used and on what is considered to count as multifunctionality.

Table 4: Cumulative multifunctionality for various annotation strategies and segmentation methods.

segmentation method: / annotation strategy:	*turn*	*utterance*	*funct'l. segment*
a. strictly feature-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. + inherited 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	6.7	4.6	3.5

5.3 Discussion

As noted above, the annotated dialogue corpus used in the present study was marked up according to strategy b, i.e. it includes besides the communicative functions derived from utterance features also the implicated ones, except implicated functions at various feedback levels (which are taken into account in strategy g). The entailed and default 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 that 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 very unusual dialogue situations.[8]

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 (or a large corpus of examples in context, from which the indirect understanding might be learnable). However, we have argued above that in an ISU-based semantic framework it is highly 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 add to the multifunctionality that is found in dialogue.

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

All in all, the figures in the second row in Table (5.2) represent the *minimal degree of multifunctionality* that is found.

When the most fine-grained segmentation is applied, using functional segments, then all sequential multifunctionality is eliminated and only purely simultaneous multifunctionality remains. Using annotation strategy a, where all kinds of implicated, entailed, indirect, and default functions are left out of consideration, the annotations reflect purely the *independent multifunctionality* of functional segments. Table (5.2) shows that our data indicate that on average one in every three segments has two independent communicative functions. The minimal multifunctionality of functional segments, as just argued, is found when annotation strategy b is followed, and turns out to be 1.6 in our data. This means that on average two in every three segments have two independent communicative functions.

When utterance-based segmentation is used, we find that on average each utterance has two communicative functions. The difference with the multifunctionality of functional segments is caused by the fact that functional segments are often discontinuous. The main cause of this is the occurrence of Own Communication Management acts, where the speaker edits his contribution on the fly, interrupting his utterance by stallings, retractions, restarts, and so on.

The multifunctionality of a turn is simply the sum of the simultaneous multifunctionalities of its constituent utterances. It follows, from the figures in Table (5.2) for unsegmented turns, that in our corpus a turn 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 per 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.

6 Conclusions and future work

Returning to the three questions formulated at the start of this paper, we have in fact arrived at the following answers.

In response to the question whether dialogue utterances tend to have multiple functions, the answer is yes, definitely! Utterances in the usual sense, of contiguous stretches of linguistic behaviour with a grammatical status, have on average at least two functions. And if we take the most-fine-grained segmentation of dialogue into functional units and a minimal approach to the notion of multifunctionality, we still find that on average two out of every three units have more than one communicative funciton. These quantitative findings answer the first part of question 3: how many functions does an utterance typically have?

Question 2, why dialogue utterrances are multifunctional, has been answered in a theoretical sense 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 correct 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 functions in different dimensions, as well as implicated multifunctionality where the implicated function belongs to the feedback dimension(s). Entailment relations between dialogue act and default and indirect functions add further to the mulltifunctionality that can be observed.

Question 3 asks which factors influence the amount of multifunctionality that is found. The answer to this question is: 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 sequential multifunctionality. 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.

Finally, what are the consequences of the findings, reported and discussed in this paper, for the semantic interpretation of dialogue utterances? 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 never has more than one function in any given dimension. (See the arguments above about entailed functions within a dimension being semantically vacuous.) This supports the view that an update semantics which interprets communicative functions as recipes for updating a part of the information state can be developed which uses separate updates for each dimension, which, due to the independence of dimensions, can be performed by autonomous software agents, one for each dimension.[9]

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, in particular Volha Petukhova, Jeroen Geertzen, Hans van Dam, Ielka van der Sluis, Simon Keizer, Roser Morante, and Amanda Schiffrin.

References

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

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

Allen, J. & M. Core (1997) DAMSL: Dialogue Act Markup in Several Layers (Draft 2.1). Technical Report, Multiparty Discourse Group, Discourse Resource Initiative, September/October 1997.

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

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

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

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

[9]This view also underlies the PARADIME dialogue manager (Keizer & Bunt, 2006; 2007).

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

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

Bunt, H. (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.

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

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

Geertzen, J., V. Petukhova, and H. Bunt. 2007. A Multidimensional Approach to Utterance Segmentation and Dialogue Act Classification. In: *Proc. SIGDIAL 2007*, Antwerp, pp. 140–149.

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

Keizer, S. and H. Bunt (2006) Multidimensional dialogue management. In Proc. *SIGDIAL 2006*, Sydney.

Keizer, S. and H. Bunt (2007) Evaluating combinations of dialogue acts. In *Proc. SIGDIAL 2007*, Antwerp.

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

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

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

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

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

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