Multifunctionality and Multidimensional Dialogue Act Annotation

Harry Bunt,
Department of Communication and Information Sciences
Faculty of Arts, Tilburg University, Netherlands
harry.bunt@uvt.nl

Abstract

Communicative behaviour in dialogue has often been observed to be multifunctional; in particular, Allwood (1992) has claimed that an utterance in dialogue tends to be both sequentially and simultaneously multifunctional. In this paper we examine this claim using empirical data obtained in dialogue annotation experiments.

Allwood considers as 'utterances' the stretches of communicative behaviour contributed by one participant, which is what many other researchers call a 'turn'. We consider finer-grained notions of utterance, corresponding to alternative ways of dialogue segmentation: *linear segmentation*, where turns are divided into functionally relevant subsequences; and *multidimensional segmentation*. The latter belongs to a multidimensional approach to dialogue annotation, where communicative functions are viewed as operating in a certain 'dimensions' of participating in a dialogue, such as task/activity performance; turn management; feedback; and time management. Upon multidimensional segmentation, dialogue behaviour is segmented in multiple ways, distinguishing functional segments *per dimension* (and allowing segments to be discontinuous and to overlap). We also consider alternative annotation strategies, which differ in the way they deal with potential sources of simultaneous multifunctionality.

Our conclusion is that, whatever segmentation method is used and whatever annotation strategy, multifunctionality never goes away. Upon a restrictive approach to annotation, for instance not marking up entailed, inherited or indirect functions, entire turns have on average in the order of five communicative functions; about half of that is due to sequential multifunctionality. When using multidimensional segmentation, all sequential multifunctionality is eliminated, and functional segments usually have two or three communicative functions. Upon linear segmentation, some sequential multifunctionality remains, resutling in a slightly higher overall multifunctionality than in the case of multidimensional segmentation.

Keywords: multifunctionality, dialogue annotation, multidimensional annotation, dialogue segmentation

1 Introduction

Utterances in dilogue are often multifunctional, i.e. they have more than one communicative function. In his 1992 paper *On dialogue cohesion*, Jens Allwood writes:

> *An utterance can be mono- or multifunctional. If it is multifunctional, its multifunctionality can be simultaneous or sequential. For instance, A's utterance in the following example contains sequentially the functions feedback giving, request, request, statement and response elicitation.*
>
> *(1) A: Yes! Come tomorrow. Go to the church! Bill we be there, OK?*
> *B: The church, OK.*
>
> *Furthermore, the statement 'Bill will be there' could simultaneously be a promise and thus illustrate simultaneous multifunctionality.*

At the time when this was written, the observation that communicative behaviour may be multifunctional was mainly of analytical interest. In more recent years, it has played a major role in the design of dialogue annotation systems.

Due to, on the one hand, the general movement in linguistics towards corpus-based research, and on the other hand the development of spoken dialogue systems, recent years have witnessed a growing interest in the construction of corpora of dialogues annotated with *dialogue act* information. Dialogue acts are empirically-based and computationally well-defined combinations of a communicative function and a propositional/referential semantic content (where the communicative function defines the way the addressee's information state is updated with the semantic content when he understands the speaker's communicative behaviour ([7]), and have proved to be very useful for designing dialogue systems (see e.g. [20]). Several efforts have been undertaken to develop dialogue act annotation schemes; existing schemes differing not only in their precise sets of tags, but more importantly with respect to (1) underlying approach to dialogue modelling; (2) definition of the related concepts; and (3) level of granularity of the defined tag set. Generally, annotation schemes can be divided into one- and multidimensional ones.

One-dimensional schemes allow coding dialogue utterances with only one tag, and their tag sets are as a rule kept very simple. Because of their simplicity, they are thought to be reliable and to take less effort to apply consistently by annotators. Some researchers, e.g. [22], [23], note, however, that one-dimensional annotation schemes also have serious disadvantages. Allen and Core in [2], [16] note that annotating with a single set of mutually exclusive categories does not allow to account for the fact that utterances may perform multiple actions simultaneously, and they criticise traditional speech act theory in this respect.

Mutidimensional approaches to dialogue act annotation allow coding utterances with multiple tags. They incorporate a multifunctional view on dialogue behaviour, and have been recognised by many researchers (e.g. [5], [7]) as empirically better motivated and allowing a more accurate modelling of theoretical distinctions. While the multifunctionality of dialogue utterances has been widely recognised, computationally oriented approaches to dialogue generally see multifunctionality as a problem, both for the development of annotation schemes and for the design of dialogue systems [28]. Information that may be obtained through a multifunctional analysis is therefore often sacrificed for simplicity in computational modelling. As a consequence, the actual multifunctionality of dialogue utterances and related phenomena are still understudied, and have so far escaped extensive description and formalisation.

Dialogue act annotation schemes, i.e. collections of dialogue act tags in themselves may be neutral with respect to whether annotators should or should not assign multiple tags to an utterance, although the simpler schemes, such as the HCRC Maptask scheme ([12]) and the LINLIN scheme ([1]) are intended for one-dimensional annotation, since the various tags are meant to be mutually exclusive.

The DAMSL scheme (Dialogue Act Markup using Several Layers, Allen & Core, 1997) was designed for multidimensional annotation. In a study of the use of multiple DAMSL tags in annotating meeting recordings, Clark and Popescu-Belis ([14]) found that only a very small percentage of the possible combinations of tags were in fact used by annotators (220 out of approx. 4 million). This is due to the fact that many tags are supposed to be mutually exclusive; moreover, sometimes the assignment of one tag forces the assignment of another one. In DAMSL some of the dependencies between dialogue act tags are captured in the annotation scheme, and some are captured in the accompanying annotation software, while others are not captured at all. We have found that it is possible to design truly multidimensional annotation schemes that support not only the assignment of multiple dialogue act tags to dialogue utterances, but do so in a principled way, based on a well-defined notion of dimension in dialogue. This work is inspired by efforts to develop a widely shared set of tools and concepts for semantic annotation in the EU project LIRICS (see [10]) and in ISO TC 37/SC 4/TDG 3 (see http://let.uvt.nl/research/ti/tdg3).

In the next section we will return to what it means for some communicative behaviour to have multiple communicative functions; first we consider what are the entities that may be said to have a communicative function. Allwood calls these entities 'utterances' and defines them as follows (Allwood 2000: 67):

> *The basic individual communicative unit in interaction I will call a contribution. A contribution can be defined as a sequence of communicative behaviour bounded by lack of activity or another communicator's activity. If the speaker's activity should cease during a contribution (e.g. by pausing while speaking), the pause must not be filled by another communicator's contribution, nor must it be so long that it is more reasonable to regard renewed activation as a*

new contribution. The unit in spoken dialogue corresponding to a contribution is an utterance.

Note that (a) this notion of contribution is not to be confused with the same term as used in Clark and Schaefer's Contribution Model of dialogue [15]; (b) utterances according to this definition are units of spoken language only (which we believe to be an unfortunate restriction); and (c) an utterance in this sense is everything that a speaker contributes in what is often called a *turn*.

Are utterances in Allwood's sense indeed the appropriate units to be assigned communicative functions? The reason why utterances may be sequentially multifunctional, as Allwood observes, is simply that utterances in this sense can be quite complex, being made up of smaller parts that have communicative functions. Allwood calls these smaller parts 'grammatical units', without giving a definition. A's contribution in dialogue (1), for example, would contain five grammatical units, three of which are sentences and two feedback morphemes. This suggests in fact that the appropriate units in dialogue that carry communicative functions are not utterances, but smaller entities, such as the linguistic objects that Allwood calls grammatical units. In many people's vocabulary, these units are called 'utterances' – for instance, Levinson (1983), writes:

An utterance is the issuance of a sentence, a sentence-analogue, or sentence-fragment, in an actual context.

This seems a correct, albeit vague characterization of the grammatical units in (1). We will follow this terminology, using 'utterance' where Allwood uses 'grammatical unit', and 'turn' where Allwood uses 'utterance'. Using this terminology we can say at this point that:

1. Utterances can have several communicative functions simultaneously, as exemplified by promises also being statements.

2. Turns may consists of several utterances and thus be sequentially multifunctional, inheriting moreover the simultaneous multifunctionality of its component utterances.

We will return to the definition of functional units in dialogue in more detail in section 4.

2 Multifunctionality and Annotation

The sequential multifunctionality of turns is obvious, and simply a consequence of considering the functionality of stretches of lingusitic behaviour which may be quite complex. The more interesting question, which is central to this paper, is whether or to what extent the multifunctionality of linguistic (or, more generally, communicative behaviour) disappears when we assign communicative functions to simpler, smaller units. When taking sufficiently smaller units, it should be possible to get rid of sequential multifunctionality; the more fundamental question is whether or how much simultaneous multifunctionality can also be eliminated in this way. So simultaneous multifunctionality is the more interesting phenomenon to investigate.

2.1 Simultaneous multifunctionality

Allwood (1992) mentions the example of a promise being also a statement to illustrate simultaneous multifunctionality. This form of multifunctionality is a consequence of the logical relation between the concepts of promise and statement: with a promise, the speaker informs the addressee of his commitment to perform a certain action (or bring about a certain state of affairs), and as such the promise has the function of a statement, but compared to a statement the speaker additionally commits himself to performing that action. It is debatable whether this is truly a form of multifunctionality, however. An action is multifunctional when it serves multiple purposes; does a promise, besides promising something, serve the additional purpose to inform the addressee of something? I don't think so: informing the addressee of the speaker's commitment is *part of* a promise. The action of committing oneself to perform a certain action without communicating that is not a promise, as it's not a communicative act (but something like making a resolution). We would therefore say that an utterance such as *Bill will be there* is either intended by the speaker as just a statement, i.e. as a description of a state of affairs that he wants to bring to the addressee's attention, or as a promise, and in the latter case it would *not* be a statement as well. In no case would the utterance have both the functions of a statement and a promise. However, we believe that there are deeper reasons why utterances may be be multifunctional.

In natural dialogue there is often a certain activity or task for which one or both of the participants want the dialogue to be instrumental. Studies of human dialogue behaviour indicate that natural dialogue involves several activities beyond those strictly related to performing the motivating task or activity; dialogue participants also give communicative feedback (i.e., they share information about the processing of each other's messages); they also elicit feedback; they also monitor a variety of aspects of the interaction such as contact and attention, taking turns, and the use of time; and they also deal with social obligations such as apologizing and thanking. They often exploit the availability of multiple communication channels to address several of these aspects simultaneously, but also when there's only one channel available, as in telephone conversations, they still have this multiplicity of aspects to take into account. This is why dialogue utterances are often multifunctional, serving multiple purposes at once.

For example, an utterance may answer a question, provide positive feedback about the understanding of the question, and pass the turn to the dialogue partner, as in the second utterance in the following dialogue fragment. The repetition of material from the question can be seen as providing feedback on understanding, while the sentence-final intonation in combination with a pause indicates the turn giving function:

(2) 1. U: Can you tell me what time is the first train to the airport on Sunday morning?
2. S: On Sunday morning the first train to the airport is at 5.32.
3. U: Thank you.

The multifunctionality of S's answer (2) also having a feedback function is of a very interesting kind. In an information-state update approach to dialogue analysis, like DIT ([7]) or the TRINDI/GODIS approach ([29]), characteristic of a dialogue act with the communicative function 'WH-Answer' is that the answerer believes that the addressee wanted to know the information that he supplies. As this belief has been created by his understanding of the preceding utterance as a question, by answering the question he indicates that he believes to have understood the question. An answer therefore implies positive feedback on understanding the question. We will therefore say that an answer utterance has a positive feedback function as an *entailed function.* There are many more instances of entailed functions; for example, accepting or rejecting an offer or a request implies understanding of the offer/request.

Note that utterance 3 in (2) is also multifunctional, as it expresses gratitude but also provides positive feedback about the speaker's processing of the answer. (Moreover, we have found that thanking also tends to function as an indication of dialogue closure.)This is a form of multifunctionality which differs from that of entailed functions or from the alleged multifunctionality of the promise/statement variety, which is due to inheritance relations in a hierarchical annotation schema. Thanking and positive feedback do not have a *logical* relation, but a conventional, 'pragmatic' relation. We therefore call these functions *implicated.*

Indirect speech acts are sometimes also considered as giving rise to multifunctionality. An utterance such as *"I would like to know what time it is"* may be argued to have both the function of a statement and that of a question. We believe that, similar to the alleged multifunctionality of the promise/statement variety, the use of indirect speech acts should not be regarded as a form of multifunctionality. The speaker who utters *"I would like to know what time it is"* in order to get the addressee to tell the time does not *in addition to that* also want to tell the addreessee that, he (the speaker) would like to know what time it is. One of the aspects of asking the question what time it is, is that the speaker is making the addressee aware that the speaker would like to know what time it is. In an information-state update approach to utterance meaning, like DIT, this becomes especially clear since the update of the addressee's information state that occurs when he understands *"I would like to know what time it is"* as a question about the time, includes adding to that information state the information that the speaker would like to know what time it is. So the question function of the indirect interpretation entails the statement function of the direct interpretation.

In the DIT annotation scheme indirect questions are treated as a separate type of dialogue act, since they have slightly different appropriateness conditions than 'direct' questions: while the direct question *"What time is it?"* carries the assumption that the addressee kows what time it is, indirect formulations such as *"I would like to know what time it is"* and *"Do you know what time it is?"* do not have this assumption. Direct and indirect questions are therefore hierarchically related in the DIT taxonomy of communciative functions, just as promise and statement.

3 Dialogue Dimensions

A well-designed multidimensional dialogue act annotation scheme should support annotators in assigning multiple function tags to utterances, while avoiding the consideration of theoretically impossible tag combinations. Such support should come from conceptual clarity of the ingredients of an annotation scheme, in particular of (i) a conceptually clear notion of dimension; and (ii) clear definitions of communicative functions within dimensions. The term 'dimension' in this context is justified only if it allows to locate objects in a multidimensional space in the usual way: the object has one value for each dimension, and never more than one value. Since utterances may have a value only in certain dimensions and not in others, this means that theoretically in all dimensions there is a value 'none', corresponding more or less to 'zero' in a numerical multidimensional space. Moreover, dimensions are assumed to be orthogonal: the assignment of a value in one dimension is independent of that assigned in other dimensions.[1]

DAMSL being the most widely used annotation scheme for multidimensional dialogue act annotation, let us examine to what extent DAMSL approaches these ideals.

3.1 Multidimensionality in DAMSL

DAMSL makes a top-level distinction of four types of information, called "layers": Forward-looking functions; Backward-looking functions; Information Level, and Information Status. The layers of Forward- and Backward-looking functions, also called 'dimensions', contain DAMSL's communicative functions. The assumption behind this dichotomy is that all utterances can have both backward- and forward-looking functions (BLFs and FLFs). The FLFs are subdivided into the following 8 'dimensions' ([2]):

1. Statement: Asserts and other acts where the speaker makes a claim about the world (modified in Core et al., 1998 to also allow statements to be claims about the communication).
2. Info-request: Speaker requests Hearer (by just asking or in another, indirect way) to provide information.

[1]This may not seem entirely true, since e.g. it seems impossible to answer a question and at the same time signal not to have understood the question (negative feedback). However, this may be regarded as a fact about communication (not all points in the multidimensional space are occupied by instances of communicative behaviour), rather than as a dependency between dimensions.

3. Influencing-addressee-future-action: Speaker is suggesting potential action to Hearer, beyond answering a request for information.

4. Committing-speaker-future-action: Speaker is potentially committing himself to perform a future action.

5. Conventional: Opening or Closing, i.e. Speaker summons Hearer and/or starts the interaction, or Speaker closes the dialogue or is dismissing Hearer.

6. Explicit-performative: Speaker is performing an action by virtue of making the utterance.

7. Exclamation (no definition supplied).

8. Other-forward-looking-function: No definition given; supposedly any FLF that does not fit into the categories 1-7.

The backward-looking functions are subdivided into the following 4 dimensions:

1. Agreement: Speaker is addressing a previous proposal, request, or claim, with the possibility of accepting or rejecting all or part of the proposal, request or claim; of withholding his attitude towards the proposal, request, or claim; or stating his attitude while being non-committal to the proposal, request, or claim.

2. Understanding: Utterances concerning the understanding between Speaker and Hearer, ranging from merely hearing the words to fully identifying intention.

3. Answer: Standard reaction of Speaker to an Info-request action by Hearer.

4. Information-relation: Tag which should capture how the content of this utterance relates to the content of its antecedent (e.g. in terms of the relations of Rhetorical Structure Theory).

There is no theoretical underpinning of this choice of dimensions, and in fact it uses the rather naive notion of dimension as a set of mutually exclusive values. This is not a satisfactory notion of dimension, however, as can be seen in the following example.

(3) 1. S: And what possibilities do you have on Tursday?
2. U: Did you say Thursday?

Utterance 2 expresses a problem in U's understanding of the previous utterance, and as such has a function in the Understanding dimension. This dimension in DAMSL has two possible values: Signal-understanding and Signal-non-understanding. The latter tag would clearly be too crude, since S expresses an understanding problem concerning only a small part of the previous utterance. Something like 'Partial-signal-non-understanding' would be closer, if it existed, but would still not be accurate. An accurate description would be: S wants to make sure that he correctly understood the designated element in the previous utterance. But *S wants to make sure that X* is the essence of the communicative function known as Check, a function in the Info-request dimension (see [17]). So an attempt to accurately characterize the utterance in the Understanding dimension leads to characterizing it in the Info-request dimension! Similarly, an accurate characterization of *I did not quite hear if you said 'Thursday'* in the Understanding dimension would result in a characterization in the Statement dimension!

The point of examples such as these is that questions, assertions, checks and answers can be not just about the task or activity at hand, but can also be about understanding; in fact, questions, assertions, answers, etc. can be about *any* aspect of the communication. For example, *Are you there?* may be a question about the presence or attention of the dialogue partner; *I think we're done* probably asserts that the dialogue can soon be ended, and *That's what I said* may be an answer in the feedback dimension. Questions, assertions, answers, etc. are therefore functions that do not constitute separate dimensions, as DAMSL has it, nor do they belong to any other particular dimension, and so the DAMSL categories Info-request, Statement, and Answer do not constitute proper dimensions. A similar analysis applies to requests, offers, instructions, suggestions, and other directive and commissive acts. Dialogue acts with these functions can also be about any aspect of the dialogue, as is illustrated by such examples as *Please repeat that* (Request) and *Shall I repeat that?* (Offer) and *I'll be right back* (Promise). For this reason the DAMSL categories 'Committing-speaker-future-action' and 'Influencing-addressee-future-action', likewise, do not qualify as proper dimensions.

3.2 Dimensions in DIT

In order to design a dialogue act annotation schema that is truly multidimensional, we start not just from possible combinations of communicative functions but from the conceptual view that a participant in a dialogue has a number of things to manage, besides trying to perform the task or activity that motivates the dialogue, as mentioned in section 2.1. According to DIT, these conceptually distinct aspects of participating in a dialogue qualifies as a proper dimension of dialogue behaviour if the following two conditions are satisfied:

(4) 1. This aspect of participating in a dialogue can be addressed by means of dialogue acts that have a communicative function especially for this purpose;
2. This aspect of participating in a dialogue can be addressed independently of other aspects, i.e., dialogue utterances can have a communicative function for this aspect, independent of its functions in other dimensions.

The first of these criteria means that we are considering an aspect of communication that cannot just be distinguished according to some abstract conceptual analysis, but that corresponds to empirically observable dialogue behaviour. The second condition requires dimensions to be orthogonal.

Using the first criterion in (4), we have examined a wide variety of dialogues (information-seeking; assistance giving; problem-solving) in a range of different situations (face-to-face; telephone; through keyboard + screen; WOZ-simulated human-computer dialogues) and have identified the following seven aspects of communication that may be addressed by dialogue acts, expressed fully or partly linguistically in Dutch or English:

1. task/activity performance
2. contact and attention monitoring
3. feedback on understanding and other aspects of the processing of dialogue utterances
4. turn management
5. time management
6. dialogue and topic structure management
7. editing of ongoing speech or text
8. social obligations management.

We subsequently applied the second criterion of (4) in order to determine if these aspects would qualify as proper dimensions.

Consider for example the category of time management. Utterances that address time management include those where the speaker wants to gain a little time in order to determine how to continue the dialogue; this function is called Stalling. Speakers indicate this by slowing down in their speech and/or by using fillers, such as *ehm, well, you know...* The devices used to indicate the Stalling function can be applied to virtually any utterance, which can have have any other function in any other dimension. Time management therefore satisfies the second criterion, and qualifies as a proper dimension.

A similar analysis can be applied to the other candidate dimensions mentioned above. Of these, the feedback category should be divided into two, depending on whether a speaker gives feedback on his own processing, or whether he gives or elicits feedback on the addressee's processing; we call these dimensions 'Auto-feedback' and 'Allo-feedback', respectively (see [6]). Similarly, speech editing functions should be divided those concerned with editing the speaker's own contribution ('Own Communication Management', to use Allwood's terminology ([3]), and those where the speaker edits the partner's speech (Partner Communication Management), which occurs for instance when a speaker assists the dialogue partner in producing a contribution (such as completing an utterance which the dialogue partner is struggling to complete).

Altogether, this leads to the distinction of the following 10 dimensions:

Task/Activity: dialogue acts whose performance contributes to performing the task or activity underlying the dialogue;

Auto-Feedback: dialogue acts that provide information about the speaker's processing (perception, interpretation, evaluation, or application) of the previous utterance or some particular previous utterance(s);

Allo-Feedback: dialogue acts used by the speaker to express opinions about the addressee's processing (perception, interpretation, evaluation, or application) of the previous utterance or some particular previous utterance(s), or that solicit information about that processing;

Turn Management: dialogue acts concerned with grabbing, keeping, giving, or accepting the speaker role;

Time Management: dialogue acts signalling that the speaker needs a little bit of time to formulate his contribution to the dialogue, or that his preparation for producing a contribution requires so much time that a pause is necessary;

Contact Management: establishing whether the dialogue partner is present and paying attention; also indicating the speaker's presence and attention;

Own Communication Management: dialogue acts to indicate that the speaker is editing the contribution to the dialogue that he is currently producing;

Partner Communication Management: the agent who performs these dialogue acts has the addressee rather than the speaker role, and assists the dialogue partner in his formulation of a contribution to the dialogue;

Discourse Structuring: dialogue acts for opening or closing a (sub-)dialogue or announcing that the speaker is going to perform certain dialogue acts, possibly conditional to the addressee's consent that he do so; start or close a topic, or negotiate a topic shift;

Social Obligations Management: dialogue acts that take care of social conventions such as welcome greetings, apologies in case of mistakes or inability to help the dialogue partner, and farewell greetings.

Table 1 gives some examples of communicative functions within each of these dimensions, with typical utterance forms in English. Note that in natural dialogue many of these functions are often indicated not or not only linguistically, but (also) through nonverbal means such as facial expressions, head movements, direction of gaze, and hand gestures. Nonverbal expressions corresponding to many dialogue acts in the various dimensions mentioned here have been identified by Petukhova ([26]) in a multidimensional analysis of recorded conversations in meetings, as part of the EU project AMI.[2]

[2]See http://www.amiproject.org/

Dimension	Dim.-specific functions	Typical expressions
Task/Activity	Open-, CloseMeeting; Appoint; Hire	domain-specific fixed expressions
Auto-Feedback	PerceptionNegative	*Huh?*
	EvaluationPositive	*True.*
	OverallPositive	*OK.*
Allo-Feedback	InterpretationNegative	*THIS Thursday.*
	EvaluationElicitation	*OK?*
Turn Management	TurnKeep	final intonational rise + pause
	TurnGrabb	hold sign with hand
	TurnGive	*Ye-es.*
Time Management	Stalling	slowing down speech; fillers
	Pausing	*Just a minute*
Contact Management	ContactChecking	Hello?
Own Commun. Man.	SelfCorrection	*I mean...*
Partner Comm. Man.	PartnerCompletion	completion of utterance
Discourse Structuring	AnnounceDialogueAct	*Question.*
Social Obligations Management	Apology	*I'm sorry*
	Greeting	*Hello!* *Good morning.*
	Thanking	*Thanks.*

Table 1: Examples of dimension-specific communicative functions and typical expressions per dimension.

Note that the total set of communicative functions that is available for constructing dialogue acts in this multidimensional system, consists of the dimension-specific functions of which some examples are mentioned above and listed in Table 1, plus all the general-purpose functions, which include:

1. **Information-seeking functions:** all sorts of questions (WH-questions, yes/no-questions, multiple-choise questions, checks, etc.)
2. **Information-providing functions:** informs, agreements, disagreements, corrections, etc., but also informs with additional rhetorical functions such as explanation, justification, exemplification, etc. answers, i.e. information-providing dialogue acts in response to an information need signaled by the dialogue partner, including also confirmations, disconfirmations.
3. **Commissive functions,** where the speaker commits himself to performing some action(s) or to bring about a certain state of affairs, possibly conditional on the addressee's consent that he do so, such as offers, promises, and acceptance of requests or other directive functions;
4. **Directive functions,** where the speaker wants the addressee to consider some action(s) to perform, potentially putting pressure on the addressee to do so, such as instructions, requests, and suggestions.

4 Dialogue Segmentation and Annotation

In section 1 we tentatively settled on ‘utterances’ as the units of dialogue that we assign communicative functions to. The notion of ‘utterance’ (in Levinson’s sense) has too vague a definition to be really useful, however. A given turn in a dialogue can usually be segmented into ‘utterances’ in very many ways, most of which would be irrelevant from the point of view of dialogue act annotation. So how do we identify the relevant units?

The segmentation of dialogues into functionally meaningful units has been approached as chopping up the dialogue into relevant parts, This approach runs into problems, however. For example, Core & Allen mention the following problematic examples ([16]):

(5) a. We’ll go Tuesday I mean Friday

b. u: take the product to
s: yes?
u: to CorningI would like to

Concerning the first example, they remark that “Labeling *I mean* as a separate utterance from *We’ll go on Tuesday* would mean cutting off *Friday* from *We’ll go Tuesday*’. Indeed, from a functional point of view one would like to identify *We’ll go on Friday* as one functional segment and *Tuesday I mean Friday* as another. Allen & Core’s ‘solution’ is to first define ‘utterance’ from a functional point of view as follows:

> *An utterance is a set of words by one speaker that is homogeneous with respect to Information Level and Forward and Backward Communicative Functions.*

and second to take the position that “DAMSL is not designed to annotate speech repairs”, hence communicative functions relating to speech editing are not part of the Forward or Backward-looking functions, and therefore utterance (5) can be claimed to be “homogeneous with respect to Forward and Backward-looking functions”. This is of course not a real solution, as it precludes the annotation of speech repairs with functional tags.

Concerning the second example, Allen & Core say that “Short interruptions by another speaker do not break up an utterance... *take the product to to Corning* is treated as one utterance.” This is possible by not treating the intervention *yes?* as having a communicative function. This is clearly wrong, since DAMSL does have feedback functions, and the *yes?* clearly does have such a function (plus a turn-giving function, which is not covered in DAMSL).

Other examples would be even more problematic for this approach, such as the following:

(6) I think twenty five euros for a remote... how much is that locally in pounds?... is too much money to buy an extra remote or a replacement one.

(From the AMI corpus). Using DAMSL's terminology, we have here an Assert act interrupted by an Info-request. Since these are both Foward-looking functions, this stretch of dialogue is not 'homogeneous' with respect to these functions, and the utterance would have to be segmented into three parts, which is clearly wrong.

One obvious solution to the problems posed by these examples is to abandon the idea of chopping up a dialogue into contiguous segments, and to allow functional segments to be discontinuous. This still leaves us with the problem of overlapping functional segments, an example of which occurs in the following dialogue fragment:

(7) 1. U: When is the first train to the airport on Sunday?
2. S: On Sunday the first train to the airport is at 5.32.
3. U: Thank you.

Here S's utterance as a whole has the function of an answer to the preceding question; the first part of it *On Sunday the first train to the airport* can be viewed as having the function to provide feedback information about what S understood the information request to be. Both the feedback function and the answer are Backward-looking functions, hence the utterance as a whole would, according to Allen & Core's definition, have to be split up somehow, but how? The only segmentation that makes sense here is one in overlapping segments, so the conclusion seems unescapable that functional segments should be allowed to be overlapping as well as discontinuous.

Modern annotation tools such as ANVIL allow dialogues to be annotated using multiple tiers, where different tiers can have different segments, and where a segment in one tier may overlap with a segment in another. This may seem to be a solution to segmentation problems, but in fact it only gives annotators complete freedom to choose functional segments as they like, without any guidance or restrictions. Where ideally an annotation schema should support annotators in considering only sensible tag combinations, by the same token an annotation theory should also support annotators in considering only sensible segmentations.

We believe that such support can be based on the notion of dimension as introduced in section 3. Since a functional segment should have at most one communicative function per dimension, a multi-tier annotation should have at most one tier for each dimension, and one functional tag in each tier. In other words: tiers should correspond to dimensions, not to some segmentation that an annotator considers convenient. Such an organization supports annotators in considering potential multiple segmentations, and may be expected to support a more consistent use of a multidimensional annotation schema.

Notice that upon multidimensional segmentation the phenomenon of simultaneous multifunctionality appears in the annotation as a certain segment occurring with a functional tag in more than one dimension.

5 Degrees of multifunctionality in multidimensional annotations

The claim that utterances in dialogue are multifunctional can be empirically investigated given a well-defined annotation scheme and an approach to annotation that allows for assigning multiple communicative function. We investigated the amount of multifunctionality that is observed when annotating with the DIT^{++} scheme,[3] depending on two factors: (i) the segmentation method and (ii) the annotation strategy with respect to various (alleged) forms of simultaneous multifunctionality. Six 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. (See [19]) for a report on the first part of this assessment.) The dialogues were of several kinds:

(1) dialogues over a microphone plus head set with a WOZ-simulated helpdesk, providing assistance in the use of a fax machine (from the DIAMOND corpus[4]);

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

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

(4) Dutch Map Task dialogues.

Two expert annotators also marked up these dialogues, and constructed together, on the basis of all the annotations, a gold standard annotation for a part of this material. The communicative behaviour in these dialogues varies a lot, depending on how formal and constrained the dialogue situation appeared to the participants. The Map Task dialogues approach natural conversation more than the other types of dialogue; the observed degrees of multifunctionality reported below concern these dialogues.

We consider three alternative segmentation methods:

a. turn: the turn (Allwood's 'utterance') is taken as the unit which is annotated with communicative functions;

b. linear: every turn is chopped up into contiguous, non-overlapping segments. This was achieved using the DitAT annotation tool (Geertzen, 2007);

c. multidimensional: turns are divided into functional segments for each dimension (that is addressed).

We considered the following strategies for dealing with possible (alleged) sources of simultaneous multifunctionality:

[3] See http://dit.uvt.nl.
[4] See http://ls0143.uvt.nl/diamond/.

a. minimal: only communicative functions are marked that are recognizable from utterance features (lexical, syntactic, prosodic), 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;

b. + implicated functions: pragmatically implicated functions, such as an expression of thanks also expressing positive feedback, are marked as well;

c. + Turn Take: a turn-initial segment is marked as having a TurnTake function if it does not already have a TurnGrab function (i.e., it forms an interruption) or a TurnAccept function (i.e., the speaker accepts the turn that was explicitly given to him by the other interlocutor). In other words, starting to speak is by default annotated as an indication of the TurnTake function;

d. + Turn Release: a turn-final utterance is marked as having a TurnRelease function if it does not already have a TurnKeep or a TurnGive function. In other words, ceasing to speak is by default annotated as an indication of the TurnRelease function;

e. + entailed functions: logically entailed 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: functions which are inherited because of logical relations in the annotation scheme are also marked, such as a Check also being a YN-Question, a Confirm also being a YN-Answer, and an Elaborate being an Inform. A special case of inherited functions, which is considered separately as strategy **g**, concerns feedback at different levels of processing;

g. + feedback levels: signals of positive feedback at some level of processing are also marked as positive feedback at lower levels (e.g. positive feedback about acceptance is by inheritance also positive feedback about perception and understanding), and negative feedback at a certain level is also marked as negative feedback at higher levels (e.g. negative feedback about perception is also negative feedback about understanding and acceptance);

h. + 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 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 2: Cumulative degrees of multifunctionality for different annotation strategies and segmentation methods.

segmentation method: / annotation strategy:	*turn*	*linear*	*multidim.*
a. minimal	2.5	1.7	1.3
b. + implicated functions	3.1	2.1	1.6
c. + Turn Take	4.0	2.7	2.1
d. + Turn Release	**4.8**	**3.3**	**2.6**
e. + entailed functions	**5.2**	**3.6**	**2.8**
f. + inherited functions	5.6	3.9	3.0
g. + feedback levels	6.6	4.6	3.5
h. + indirect functions	6.7	4.6	3.5

As noted above, when applying multidimensional segmentation, the simultaneous multifunctionality of a functional segment (defined as a possibly discontinuous stretch of dialogue that has a communicative function in at least one of the dimensions) shows up in the annotation by the fact that the same stretch of dialogue is marked up in more than one dimension. The multifunctionality of a turn is simply the sum of the simultaneous multifunctionalities of its constituent functional segments, so the figure for unsegmented turns means that in the dialogues under consideration a turn has on average two functional segments when applying multidimensional segmentation. One thing that the table does not show, but which is good to know, is that the most important source of the multifunctionality of the units in linear segmentation is the occurrence of Owcn Communication Management acts, where the speaker edits his own contribution on the fly). Multidimensional segmentation leads to distinguishing some 20% more functional units than linear segmentation.

Concerning the influence of various annotation strategies, we believe that strategies f-h should not be applied for determining the multifunctionality of units of behaviour. We have argued in section 2.1 that 'inherited functions', such as a promise being a statement, should not be considered as additional functions. This applies to all cases where communicative functions are hierarchically related in a taxonomy of functions. In the DIT^{++} taxonomy this happens for the following cases which occur in the dialogues on which Table 2 is based:

- Inform is inherited by Agreement, Disagreement, Correction, Elaborate, Explain
- YN-Question is inherited by Check and Posi-Check
- YN-Answer is inherited by Confirm, Disconfirm

Moreover, all questions (including WH-Questions and Alternatives-Questions) inherit from indirect questions, which we have considered separately (h); all answer types also inherit from 'Uncertain-Answer', which we have left out of consideration in this paper; and feedback acts at a certain level of processing mostly inherit feedback functions at other levels of processing, which we have considered separately (g). The arguments against counting inherited functions apply to all these cases.

With respect to implicated functions, we believe this is a true source of multifunctionality. From the annotation work from which the figures in Table 2 are derived, we have the impression that implicated functions occur more frequently in some other types of dialogue than the our Map Task dialogues, which contain a lot of straightforward instructions and explicit feedback.

We see that the amount of multifunctionality that we find is strongly influenced by the strategy that is followed with respect to Turn Management. It is for annotators often hard to decide in a consistent manner whether a turn-final segment has a turn-giving or a turn release function, or no Turn Management function at all. The '+ Turn Release' strategy is based on the consideration that, when a turn goes smoothly from one participant to the other, apparently the new speaker viewed the situation as allowing him to continue, in other words as the previous speaker releasing the turn; similarly for the '+ Turn Take' strategy. We consider this as a perfectly reasonable approach, so we take the added functionality that follows from such strategies quite seriously.

With respect to counting entailed communicative functions the situation is not so clear, As in the case of inherited functions, we instruct annotators not to mark these; that would be redundant. However, an entailed function can be quite important. Different approaches may be possible here: for example, the feedback function of the acceptance of a suggestion may be viewed on the one hand as serving a different purpose than the acceptance as such, but may on the other hand also be viewed as being an intrinsic part of an acceptance. Depending on the view that one takes here, either the row marked '+ Turn Release' or the row '+entailed functions' contains the most representative counts of multifunctionality. We have therefore marked these rows in boldface.

We may conclude that, whatever segmentation method is used and whatever annotation strategy, multifunctionality never goes away. When a coarse segmentation method is used, that considers entire turns as the units of communicative behaviour, then we find that the functional units have on average in the order of 5 communicative functions, about half of which is due to sequential multifunctionality and half to the multifunctionality of the smallest possible functional units that may be distinguished within a turn. These smallest possible units, which correspond to the functional segments of a multidimensional segmentation, have usually two or three communicative functions; functions for turn management are responsible for about half of this.

Acknowledgments

I would like to thank Jeroen Geertzen, Olga Petukhova and the other members of the dialogue group in the Department of Communication and Information

Sciences at Tilburg Universities for the collaborative dialogue annotation work that is the basis of the study reported here, and for the many discussions that have helped to shape the DIT^{++} annotation scheme.

References

[1] Ahrenberg, L., N. Dahlbäck & A. Jönsson (1995) *Coding Schemes for Studies of Natural Language Dialogue.* In: Working Notes from the AAAI Spring Symposium, Stanford.

[2] Allen, J., Core, M. (1997). *Draft of DAMSL: Dialog Act Markup in Several Layers.* http://www.cs.rochester.edu/research/cisd/resources/damsl/.

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

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

[5] Allwood, J. (2000). *An activity-based approach to pragmatics.* In: Abduction, Belief and Context in Dialogue, H. Bunt and W. Black (eds.), Amsterdam: John Benjamin, 47-81.

[6] Bunt, H. (1999). *Dynamic interpretation and dialogue theory.* M. Taylor, F. Neel, and D. Bouwhuis (eds.), *The structure of multimodal dialogue II*, Amsterdam: Benjamins, 139-166.

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

[8] 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

[9] Bunt, H. (2006). *Dimensions in Dialogue Act Annotation.* Proceedings LREC 2006, Genova.

[10] Bunt, H. & A. Schiffrin (2006) *Methodological aspects of semantic annotation.* In Proceedings LREC 2006, Genova, May 2006.

[11] Butterworth, B. (1980). *Evidence from pauses in speech.* Language Production: Speech and Talk, vol.1, B.Butterworth(eds.), London: Academic Press, 155-177.

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

[13] Cassell, J., Torres, O., Prevost, S.(1999). *Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation.* Machine Conversations, Wilks, Y.(eds.), The Hague: Kluwer, 143-154.

[14] Clark, D. & A. Popescu-Belis, A. (2004) *Multi-level Dialogue Act Tags.* In Proccedings of SIGDial 2004, 5th SIGDial Workshop on Discourse and Dialogue, Cambridge, MA, 163–170.

[15] Clark, H., E. Schaefer (1989). Contributing to discourse. *Cognitive Science* 13, 259–294.

[16] Core, M. and Allen, J. (1997). *Coding Dialogs with the DAMSL Annotation Scheme*. AAAI Fall Symposium on Communicative Action in Humans and Machines, Boston, MA, November 1997.

[17] Core, M., M. Ishizaki, J. Moore, C. Nakatani, N. Reithinger, D. Traum & S. Tutiya (1998) *The Report of The Third Workshop of the Discourse Resource Initiative,* Chiba University, Chiba, Japan. http://cogsci.l.chiba-u.ac.jp/∼tutiya/Publications/DRI98Report.pdf

[18] Geertzen, J.(2007) DitAT: A flexibile tool to support web-based dialogue annotation. In *Proceedings of the Seventh International Workshop on Computational Semantics (IWCS-7)*, Tilburg, January 2007, pp. 320–323.

[19] Geertzen, J., Bunt, H. (2006). *Measuring annotator agreement in a complex hierarchical dialogue act scheme*. Proceedings of SIGDIAL 2006, Sydney.

[20] Jurafsky, D., Martin, J. (2006) *Speech and Language Processing,* 2nd Edition, Chapter 23. Prentice-Hall, New Jersey.

[21] Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: University Press.

[22] Klein, M. (1999). *Standardization efforts on the level of dialogue act in the MATE project*. http://acl.ldc.upenn.edu/W/W99/W99-0305.pdf.

[23] Larsson, S. (1998). *Coding Schemas for Dialogue Moves*. Technical report from the S-DIME project.

[24] Lendvai, P., Bosch, van den A., Krahmer, E. (2003). *Machine Learning for Shallow Interpretation of User Utterances in Spoken Dialogue Systems*. Proceedings of EACL-03 Workshop on Dialogue Systems: interaction, adaptation and styles of management, Budapest.

[25] Levinson, S. (1983) *Pragmatics*. Cambridge, UK: Cambridge University Press.

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

[27] Popescu-Belis, A. (2005). *Dialogue Acts: One or More Dimensions?*. ISSCO working paper.

[28] Traum, D. (2000). *20 questions on dialogue act taxonomies.* Journal of Semantics, 17(1), 7-30.

[29] 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.