Introducing Communicative Function Qualifiers

Volha Petukhova and Harry Bunt
Tilburg Center for Creative Computing
Dept. of Communication & Information Sciences,
Tilburg University, The Netherlands,
{v.petukhova,h.bunt}@uvt.nl

Abstract

This paper addresses the annotation and representation of multimodal dialogue behaviour expressing intentions with various possible qualifications, relating to uncertainty, conditionality, partiality or a speaker’s emotional state and attitude towards the information or action is addressed, or towards an addressee. We present a conceptual and empirically-based analysis of this behaviour, propose the introduction of communicative function qualifiers, and describe their use in a dialogue act annotation scheme.

1 Introduction

Participants in a dialogue do not just exchange information by simple statements, direct questions and clear-cut answers. They may be less straightforward in expressing their communicative intentions, formulating a question indirectly, accepting a request conditionally, or expressing partial agreement. They often indicate their attitude toward their communicative partners, toward what they are saying, or toward things that they intend to do. They emphasize, express doubts, criticize, show interest, and so on. All this can be signalled in various ways, e.g. by using verbal indicators like modals, by intonation and by utilizing body language and facial expressions. Approaches to the analysis, annotation, or computational modelling of dialogue behaviour struggle with these phenomena. This is especially true for attempts to annotate spoken and multimodal dialogue with information about the communicative actions (‘dialogue acts’) that the participants perform.

In the context of the ISO project 24617-2 “Semantic annotation framework, Part 2: Dialogue acts”, which aims to design a standard for annotating dialogues with dialogue act information, an approach has been developed for dealing with these phenomena which is explained and motivated in this paper.

This paper is organised as follows. We first define qualifiers by describing their semantics (Section 3). In Section 4 four types of qualifiers are proposed. We discuss theoretical and empirical considerations for their introduction and illustrate their possible meanings with examples from dialogue corpora. Conclusions are drawn in Section 5.

2 Related work

Previous efforts toward the development of standards for dialogue act annotation include the MATE project, which analysed the majority of existing dialogue act annotation schemes in order to identify commonalities and differences in approaches to the annotation of dialogue, especially for designing tools for mapping, extraction, visualization and evaluation of annotated dialogue data (Klein, 1999).

In the LIRICS project the methodological factors which should be taken into consideration when isolating appropriate semantic annotation concepts were studied (Bunt and Schifrin, 2006), and a preliminary set of data categories for dialogue act annotation was proposed.

---

1Linguistic InfRrastructure for Interoperable ResourCes and Systems (http://lirics.loria.fr)
The AAMAS workshop ‘Towards a Standard Markup Language for Embodied Dialogue Acts’ addressed the relevance of existing dialogue act taxonomies for the analysis of multimodal dialogue data, including nonverbal behaviour.

To identify commonalities and differences in alternative approaches to the annotation and representation of dialogue information is a crucial step in defining an annotation standard. We analysed 18 well-known dialogue act annotation schemes and came to the conclusion that virtually every dialogue act taxonomy fails to capture the nuances in the performance of communicative actions relating to modality (e.g. uncertainty), partiality, conditional-ity and emotional and attitudinal state. For example:

1. A: Would you like to have some coffee and cake?
2. B: Only if you have it ready.
3. B: Some coffee could be nice, but what time is it now?

Response 2 in (1) can be characterized as conditional acceptance of offer and response 3 as partial and modal acceptance of offer.

Some dialogue act taxonomies pay attention to these phenomena. For instance, DAMSL and DAMSL-based schemes like SWBD-DAMSL, MRDA and Coconut distinguish such functions as Reject-Part or Accept-Part. To address uncertainty DIT++ has an Uncertain form of Answer, Agreement, Disagreement, Confirm and Disconfirm, and conditionality is captured by introducing indirect speech acts for Request and Questions. This is not really a way to go, however, since for example an answer can be uncertain and partial at the same time, and expressed with some type of attitude as well, so this would lead to an explosive growth of the tagset, undermining its transparency. Instead, we propose to add a set of qualifiers that can be attached to communicative function in order to describe the speaker’s behaviour more accurately.

We argue, however, that the introduction of specific qualifiers should be (1) theoretically justified, e.g. studied in literature and defined in some existing approaches to dialogue act annotation, and (2) empirically grounded, e.g. observed in real dialogue data and successfully recognized by annotators. For the empirical evidence relating to communicative function qualifiers we analysed the AMI meeting corpus (3,897 utterances); and the MapTask English instructing dialogues (386 utterances).

3 The Semantics of Qualifiers

A qualifier is an additional element in the description of dialogue actions. Semantically, qualifiers describe and provide more accurate definitional meaning for another element. In dialogue we deal with utterances which have a certain semantic content (propositional, referential) that corresponds to what the utterance is about, and which have a communicative function that specifies the way semantic content is to be used by the dialogue partner to update his information state when he understands the utterance. Communicative function qualifiers do not change but rather specify the way the act’s semantic content changes the addressee’s information state, e.g. by expressing the strength or weakness of certain assumptions and beliefs, or the physical and emotional abilities and state of a dialogue participant. In other words, qualifiers provide a more detailed description of the speaker’s intention.

Qualifiers can limit the scope of a communicative function by expressing partiality. A participant may accept or reject part of an offer or a suggestion, provide a partial answer to a question, or partly agree with a claim. For example:

2 A: The new student is brilliant and imaginative.
B: He’s imaginative.

In (2) B agrees with part of A’s contribution.

Qualifiers express a modality or conditionality. They may, for example, indicate the strength of a speaker’s beliefs about the information being provided or about the partner’s abilities to perform a requested action.

Most existing dialogue act taxonomies consider only two possible responses to an offer, a suggestion, or a request: accepting it and rejecting it. However, people often respond in a less straightforward way, e.g. accepting the offer conditionally or with a certain modality. Consider the following example:

Augmented Multi-party Interaction (http://www.amiproject.org/)

Detailed information about the MapTask project can be found at http://www.hcrc.ed.ac.uk/maptask/
A: Would you like to have some coffee?
1. B: I’m not sure I want any.
2. B: Maybe later?
4. B: Yes, please, if you don’t mind to bring it for me.
5. B: Coffee? At midnight?

Response 1 can be seen as modal acceptance/rejection of the offer in (3), expressing uncertainty; in response 2 modal acceptance is expressed by communicating probability; response 3 can be characterised as modal acceptance where certainty is expressed; response 4 is a conditional acceptance; and in response 5 the speaker signals surprise.

Response 5 in (3) shows that many dialogue acts can be performed with additional expression of the speaker’s emotional state with respect to the semantic content of the act or attitude towards the addressee, or towards the content of a proposition and towards the intended possible actions. Dialogue contributions may be emotionally or attitudinally loaded, performed in particular mode.

To summarize, at least four categories of qualifiers, modality, conditionality, partiality and mode, deserve to be analysed in more details. This is the topic of the next section.

4 Types of communicative function qualifier

4.1 Modality

Generally, modality is seen as a category of linguistic meaning which is concerned with expressions of certainty. Mindt (1998) distinguishes 17 modalities: (i) possibility/high probability, (ii) certainty/prediction, (iii) ability, (iv) hypothetical event/result, (v) habit, (vi) inference/deduction, (vii) obligation, (viii) advisability/desirability, (ix) volition/intention, (x) intention, (xi) politeness/downtoning, (xii) consent, (xiii) state in the past, (xiv) permission, (xv) courage, (xvi) regulation/prescription, and (xvii) disrespect/insolence. Leech (1971) proposed to differentiate between 11 modal meanings: (i) possibility (theoretical, factual), (ii) ability, (iii) permission, (iv) exclaimatory wish, (v) obligation/requirement, (vi) rules and regulations, (vii) local necessity, (viii) prediction/predictability, (ix) willingness (weak volition), (x) intention (intermediate volition), and (xi) insistence (strong volition). The most widely used division of the modal domain distinguishes between (i) alethic, (ii) deontic, (iii) dynamic and (iv) epistemic.

Alethic modality is concerned with degrees of certainty of the truth of a proposition; this is a category of modal logic for which it is not easy to find examples in natural language. Deontic modality is concerned with what is possible, necessary, permissible or obligatory according to law or social and moral obligations, and refers to actions and events. Dynamic modality refers to physical necessity or possibility and is concerned with expressions of ability, power, futurity, prediction and habit. This modality is applicable to propositions as well as to actions. Deontic and dynamic modals are closely related to communicative action. Some dialogue acts are inherently modal. For instance, directives often express ‘deontic possibility’ or ‘deontic necessity’ as in the following example:

(4) We should investigate whether it needs a battery at all.

With a directive the speaker puts some pressure on the addressee to perform certain actions. Accepting an Offer puts pressure on the speaker to perform the offered action if the offer is accepted (as in examples (1) and (3). Indirect directives could be interpreted as dynamic modals. For example:

(5) Can you pass me my notepad?

In 5 the speaker wants the addressee to perform the requested action, conditional on his ability to do so. Commisive acts put pressure on the speaker to perform a certain action, possibly dependent on certain conditions, and possibly dependent on the addressee’s consent, as in the following examples:

(6) A: Can I help you?
(7) A: I’ll talk about the new project I’ve just received

In (6) the speaker expresses his/her ability to perform a certain action. In 7 the speaker commits himself to perform an action. Both cases can be viewed as dynamic modality.

Epistemic modality is concerned with what is possible given what is known and what evidence is available. Epistemic modals form an interesting category to be studied ‘because their semantics is bound up both with our information about
The world and with how that information changes as we share what we know (von Fintel and Gillies, 2007). The semantics of epistemically modalized utterances, which is context-dependent, is still under debate. Von Fintel and Gillies suggest that utterances with epistemic modals are used to perform more than one speech act. For example:

(8) There might have been a mistake in calculation

They argue that an utterance like (8) is (1) an assertion and (2) an explanation. This analysis is not correct because by the assertion the speaker wants to make something known to the addressee, and explanation always subsumes an assertion, in other words, making an assertion plus an explanation is semantically the same as an explanation.

Potts (2003) and Swanson (forthcoming) propose to treat epistemic modals as 'speech act modifiers’. Swanson suggests that an unmodalized sentence has to be interpreted as an assertion and a modalized sentence as ‘assertion with with tempered force’ which could have the appropriate kind of context change potential.

This approach is potentially promising. Epistemically modalized utterances may be considered as having a qualified communicative function.

Epistemic modal qualifiers are concerned with expressions of the speaker’s degree of certainty regarding the validity of a proposition. For example:

(9) 1. I think that for the next meeting we have market data

2. I guess generic remote is what we’re aiming for

In the utterances in (9) the speaker weakly believes that the propositions are true.

Uncertainty is often communicated through expressions of ‘probability’. For example:

(10) 1. It will probably be sold separately

2. That might be a fairly good target group for us

In the utterances in (10) the speaker does not eliminate other possibilities, but assigns a higher value to one possibility.

Our corpus analysis shows that dialogue participants often express assessments of the validity of their propositions. About 47% of all utterances are modalized (34.5% uncertain, 12.6% certain). A degree of certainty can be expressed verbally as well as nonverbally. Table 1 gives an overview of observed expressions.

![Table 1: Expressions of modality](image)

<table>
<thead>
<tr>
<th>Modality</th>
<th>Verbal expressions</th>
<th>Vocal prosody</th>
<th>Gaze direction</th>
<th>Head movement</th>
<th>Facial expression</th>
<th>Gesture</th>
<th>Posture orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertainty</td>
<td>may (not)</td>
<td>high standard deviation in pitch;</td>
<td>aversion</td>
<td>waggle</td>
<td>lip-compression; adaptors, e.g. self-touching;</td>
<td>posture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>might (not)</td>
<td></td>
<td>high standard deviation in pitch;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>could (not)</td>
<td>jitter;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>should (not)</td>
<td>high standard deviation in pitch;</td>
<td>aversion</td>
<td>waggle</td>
<td>lip-compression; adaptors, e.g. self-touching;</td>
<td>posture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>probably(not)</td>
<td>high standard deviation in pitch;</td>
<td>aversion</td>
<td>waggle</td>
<td>lip-compression; adaptors, e.g. self-touching;</td>
<td>posture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>un(likely)</td>
<td>high standard deviation in pitch;</td>
<td>aversion</td>
<td>waggle</td>
<td>lip-compression; adaptors, e.g. self-touching;</td>
<td>posture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>maybe(not)</td>
<td>high standard deviation in pitch;</td>
<td>aversion</td>
<td>waggle</td>
<td>lip-compression; adaptors, e.g. self-touching;</td>
<td>posture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>'not sure'</td>
<td>high standard deviation in pitch;</td>
<td>aversion</td>
<td>waggle</td>
<td>lip-compression; adaptors, e.g. self-touching;</td>
<td>posture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>'you know?'</td>
<td>high standard deviation in pitch;</td>
<td>aversion</td>
<td>waggle</td>
<td>lip-compression; adaptors, e.g. self-touching;</td>
<td>posture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>'I guess', etc.</td>
<td>high standard deviation in pitch;</td>
<td>aversion</td>
<td>waggle</td>
<td>lip-compression; adaptors, e.g. self-touching;</td>
<td>posture</td>
<td></td>
</tr>
<tr>
<td>Certainty</td>
<td>shall (not)</td>
<td>low standard deviation in pitch;</td>
<td>direct</td>
<td>head nod (for emphasis)</td>
<td>thin lips; pushing up the chin;</td>
<td>beat gestures</td>
<td>leaning forward</td>
</tr>
<tr>
<td></td>
<td>will(not)</td>
<td>no pauses</td>
<td>direct</td>
<td>head nod (for emphasis)</td>
<td>thin lips; pushing up the chin;</td>
<td>beat gestures</td>
<td>leaning forward</td>
</tr>
<tr>
<td></td>
<td>can(not)</td>
<td>no pauses</td>
<td>direct</td>
<td>head nod (for emphasis)</td>
<td>thin lips; pushing up the chin;</td>
<td>beat gestures</td>
<td>leaning forward</td>
</tr>
<tr>
<td></td>
<td>would(not)</td>
<td>no pauses</td>
<td>direct</td>
<td>head nod (for emphasis)</td>
<td>thin lips; pushing up the chin;</td>
<td>beat gestures</td>
<td>leaning forward</td>
</tr>
<tr>
<td></td>
<td>must(not)</td>
<td>no pauses</td>
<td>direct</td>
<td>head nod (for emphasis)</td>
<td>thin lips; pushing up the chin;</td>
<td>beat gestures</td>
<td>leaning forward</td>
</tr>
<tr>
<td></td>
<td>certainly(not)</td>
<td>no pauses</td>
<td>direct</td>
<td>head nod (for emphasis)</td>
<td>thin lips; pushing up the chin;</td>
<td>beat gestures</td>
<td>leaning forward</td>
</tr>
<tr>
<td></td>
<td>definitely(not)</td>
<td>no pauses</td>
<td>direct</td>
<td>head nod (for emphasis)</td>
<td>thin lips; pushing up the chin;</td>
<td>beat gestures</td>
<td>leaning forward</td>
</tr>
</tbody>
</table>

4.2 Conditionality

Conditional qualifiers refer to the possibility (with respect to ability and power), necessity or volition
of performing actions, and can therefore only be attached to action-discussion functions. Consider the following examples:

(11) 1. If you're ready, maybe you make your presentation
     2. I can do this for you if you like
     3. I’ll send you an e-mail if you give me your address
     4. If we want a few more buttons maybe we could have them in a little charging station like a mobile

Utterance 1 in (11) is a conditional request; utterance 2 a conditional offer; utterance 3 a conditional promise, and utterance 4 a conditional suggestion.

Some communicative functions are inherently conditional. For instance, a request to do X can be seen as a conditional instruct to do X (the condition being that the addressee agrees to do X), and an offer to do X can be viewed as a conditional promise to do X (the condition that the addressee accepts the offer). Indirect requests are conditional on the addressee’s consent or ability to perform the requested action. For example, in (12) the speaker asks the addressee to explain something on the condition that he is able to do so:

(12) Can you explain this?

Responses to action-discussion functions can also be conditionally qualified:

(13) A: Maybe we could have something like a touch screen
     1. B: I don’t think so, unless it doesn’t take lots of space
     2. B: If we can do that, great
(14) A: Can we just go over that again
     1. B: Just very quickly. I have to hurry you on here
     2. B: We have no time, unless you make it very quickly
(15) A: I can make buttons larger
     1. B: If it’s possible, why not
     2. B: No, only if we want basic things to be visible

Response 1 in (11) expresses a conditional declining of the request. Response 2 expresses a conditional declining of the offer.

Our corpus analysis shows that about 2.6% of all utterances are conditional. The conditionality is mostly articulated by conditional clauses with ‘if’ and ‘unless’, or phrases consisting of ‘if’ followed by an adjective, e.g. ‘if necessary’, ‘if possible’.

Conditionality can be encoded using simple binary values for action-discussion qualifiers: conditional and unconditional.

4.3 Partiality

Partial qualifiers limit the scope of a communicative function to a part of the semantic content of the utterance to which the current utterance is related. Propositions that are considered to be true mostly form an exhaustive response, e.g. answer, agreement, acceptance. Asher and Lascarides call such responses ‘strong exhaustive answers’ (Asher and Lascarides, 2003).

Often, however, the speaker provides partial responses, as in the following examples:

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

The responses 1, 2 and 3 in (16) all constitute partial answers. Response 3 is also a modal answer, since its uncertainty is articulated.

Asher and Lascarides (2003) observed that responses which rule out some possible answers can also be considered as partial. For example:

(17) A: Do you know who’ll be coming tonight?
     B: Well, not Mary.

With respect to partial agreements and acceptances, the question arises whether the speaker implicitly rejects the part that is not accepted/agreed. Walker (1994) claims that by partially agreeing with the previous partner’s statement the speaker implicitly rejects the other part, e.g. in (2) B does believe that the student is imaginative but does not believe that he is brilliant. Apart from the fact that this reasoning might be wrong, e.g. because B only heard the first part of A’s claim, or meant his utterance to be ironic, we think that the part which is not addressed is an implicature and is not part of the semantic content of what B said.
Corpus analysis shows that partiality is expressed in about 6.0% of all utterances. Most of the time (about 60%) partial qualifiers are assigned to parts of long answers or conclusions, which together form a complete dialogue act, as in the following example:

(18) A1: we have four minutes left to define our functions
    B1: okay
    A2: so we want something to keep it from getting lost
    D1: yes
    A3: we want large buttons for essential things
    B2: definitely
    A4: and we want a possibility to get the hidden functions
    D2: yep

This often occurs when the speaker provides complex information, divided up into parts in order not to overload the addressee. For example:

(19) U1: Could you tell me what time there are flights to Kuala Lumpur on Monday?
    S1: There are two early KLM flights, at 7.30 and at 8:25...
    U2: Yes,...
    S2: ... and a midday flight by Garoeda at 12.10...
    U3: Yes,...
    S3: and there’s a late afternoon flight by Malaysian Airways at 17.55.

Partiality can be treated as a binary category where the values ‘partial’ or ‘complete’ (the latter as default) can be attached to a communicative function.

4.4 Mode

Mode is a broad category of qualifiers concerned with the speaker’s attitude and emotional state.

A dialogue participant may express his attitude towards the addressee(-s), or towards the content of what he is saying. Attitudes can be divided into positive and negative. Positive attitudes towards the addressee can be articulated by being polite, friendly or cooperative. Positive attitudinal expressions include compliments and expressions of appreciation of the addressee’s actions, sympathy with the addressee as well as downplaying his mistakes. Negative attitudes can be expressed by the speaker being offensive, incooperative or impolite.

Speaker attitudes can also be derived from modality and conditionality. For instance, by formulating a claim with some degree of uncertainty the speaker often wants to appear less assertive, or to ‘save the addressee’s face’. Conditional acts are often perceived as more polite than unconditional ones, e.g. indirectly formulated requests.

Attitudes towards the content of an utterance can be expressed by emphasizing its importance, and by positive or negative evaluation of partner’s previous related contributions. To stress the importance the speaker can use expressions like ‘above all’, ‘actu-
ally’, ‘believe me’, ‘by all means’, ‘indeed’, ‘really’, ‘surely’, etc. Speakers may use their bodies to indicate that what they are saying deserves special attention, e.g. hand beat gestures are known to accompany new important information, and eyebrow movements may indicate where the focus of the addressee’s attention should be positioned.

The evaluation of partner’s utterances may be both attitudinally and emotionally loaded. The attitudinal aspect is more related to mental or cognitive processing, while the emotional aspect refers to the feelings the message evokes.

Emotions can be also evoked by the addressee’s behaviour. There is a huge interest in studying emotions in interaction. No definite taxonomy of emotion exists. One of the most well known is in Ekman’s is pioneering work in the study of emotions (Ekman, 1972). He distinguishes 6 basic emotions: anger, disgust, fear, happiness, sadness and surprise. In his later work, Ekman (1999) expanded his taxonomy and added 11 more emotions: amusement, contempt, contentment, embarrassment, excitement, guilt, pride in achievement, relief, satisfaction, sensory pleasure and shame. Some emotions can be modified to form complex emotions.

In recent years several schemes for annotating emotions and emotion-related states have become available, e.g. (Craggs and McGee Wood, 2004), (Laskowski and Burger, 2005). (AMI Emotion Annotation Subgroup). Craggs and McGee Wood (2004) distinguish along with basic emotions like happiness, sadness also affection, dislike, and misery. Laskowski and Burger (2005) distinguish between the description of behaviour and feelings, because the authors noticed that annotators tend to describe how people behave rather than how they feel. To label emotions in participant’s behaviour they have labels like objecting, protesting, etc. Feelings are analysed in terms of valence: positive, negative and neutral.

In support of the design of the AMI annotation scheme (AMI Emotion Annotation Subgroup) experiments were carried out (Ordelman and Heylen, 2005), where subjects were provided with a list of 243 terms describing emotions and were asked to select the 20 most frequent ones occurring in AMI meetings. In this way 26 emotional and attitudinal terms were selected. After annotation experiments, the following emotional and attitudinal states are defined in the AMI scheme: neutral, curious, amused, distracted, bored, confused, uncertain, surprised, frustrated, decisive, disbelief, dominant, defensive and supportive. Inter-annotator agreement in terms of Krippendorff alpha was found to vary from 0.061 to 0.443 (Reidsma, Heylen and Ordeman, 2006).

To summarize, several taxonomies label emotional and attitudinal phenomena in dialogue with different levels of granularity: coarse (positive, negative and neutral); medium (basic emotions comparable to Ekman’s 6 emotions), and fine (labels for specific emotions like misery, annoyed, worry, etc., specific attitudes like criticism, impatient, agreeable, serious, curious, etc.). This suggests that it is sensible to leave this category open, choosing specific qualifiers according to different applications and tasks.

## 5 Conclusions and future research

Table 3 summarizes the qualifier attributes and values that we propose, indicating in the rightmost column the categories of communicative functions to which they may be attached.

For future work, we intend to investigate how...
well the proposed qualifiers are recognized by human annotators and how successful automatic recognition may be, measuring inter-annotator agreement in annotation experiments investigating the machine learnability of these qualifiers.

References

AMI Emotion Annotation Subgroup. Coding Guidelines for Affect Annotation of the AMI Corpus


Annex A: Decision trees for modal and conditional qualifiers

![Decision tree for modal qualifiers.

Figure 1: Decision tree for modal qualifiers.

The utterance is modalized and is about an action; is the speaker suggesting a potential action to the addressee?

- Yes
  - Is the performance of the future action contingent on the addressee’s ability, consent or need?
    - Yes
      - Attach conditional qualifier
    - No
      - Attach unconditional qualifier
  - No
    - Is the speaker committing himself to perform an action?
      - Yes
        - No action-commitment function is added
      - No
        - Is the commitment contingent on the addressee’s consent or need?
          - Yes
            - Attach conditional qualifier
          - No
            - Attach unconditional qualifier

Figure 2: Decision tree for conditional qualifiers.