Annotating emotion in dialogue

Richard Craggs
Department of Computer Science
Manchester University
craggs@cs.man.ac.uk

Mary McGee Wood
Department of Computer Science
Manchester University
mary@cs.man.ac.uk

Abstract

Communication behaviour is affected by emotion. Here we discuss how dialogue is affected by participants' emotion and how expressions of emotion are manifested in its content.

Keywords: Dialogue, Emotions, Annotation

1 Introduction

Dialogue annotation is a fundamental stage of much of the research conducted on both Human-Human and Human-Machine dialogue.

We are fortunate to have access to a valuable corpus of 37 dialogues between nurses and patients, each comprising 200-1200 utterances (Wood, 2001). These consultations contain genuine emotional speech and form the ideal basis for studies of realistic conversational dialogue.

The emotional state of participants affects the way in which the dialogue is conducted. I propose that annotating emotion in dialogue alongside currently annotated phenomena will reveal interesting and useful correlations that will improve our understanding of dialogue and benefit natural language applications. The overall aim of this research is to develop a scheme for annotating expressions of emotion, to create an annotated corpus of dialogue containing emotion and to study the effects that a participant's emotional state has on their communicative behaviour.

2 Effects of emotion in dialogue

This research is motivated by observations made on the consultation dialogues described above. These are naturally occurring conversational dialogues conducted under unusual circumstances, in which the consultant's goal is to elicit concerns from the patient. They therefore contain an unusually high level of emotional speech. When read with a dialogue analyst's eye it is apparent that certain phenomena, interesting to the dialogue analysis community, are affected by the changing level of emotion. For example, grounding behaviour is more protracted when a participant is discussing a subject about which they feel emotional. This is manifested in an increase in the number of clarification requests and repetitions. E.g. -

```
N. How do you feel when you look at your
scar?
P. Erm, it doesn't bother me that much
N Okay
P. But I still, When I'm washing and ev-
erything I still get a funny feeling
N. You get a funny feeling, in which way?
P. It just feels strange, hollow,
N. Physically?
P. Physically yes,yes
N. Yes
P. It feels really weird
```

Turn taking behaviour changes under these circumstances too. An emotional speaker will hold the floor for an increased length of time when discussing

a topic about which they feel, for example, anxiety or joy.

Although these are casual observations of a small amount of dialogue, other studies have benefited from investigation into a speaker's behaviour when emotional. For the Verbmobil project (Bub and Schwinn, 1996) it was recognised that anger in speakers changed the way in which they communicated (Fischer, 1999). Also applications such as automated call centres would benefit from recognition of human emotion so that humans could intervene when a customer becomes angry and frustrated (Petrushin, 1999). However, these insights are limited to the vocal expression of the speaker. An annotated corpus of emotional dialogue would allow us to study all aspects of a speaker's behaviour.

3 Annotating emotion in dialogue

I envisage that a scheme to annotate dialogue would constitute one or more layers augmenting an existing annotation scheme. There are plenty of other schemes developed for previous dialogue research, many of which are designed to investigate a particular phenomenon of communicative behaviour.

In this section we will look at some existing annotation schemes. We shall investigate if any of the layers may accommodate emotion and which may present interesting correlations with emotional tagging.

Of course when looking for possible indicators of emotional speech it is important to remember that people exhibit different behaviours from each other when they speak. For example some people are more expressive than others and so a large number of expletives from one person may be natural, and not indicative of their emotional state. Prosodic studies of emotion also suffer from this complexity and it would be interesting to see if language use, and dialogue behaviour are more robust indicators of emotion than prosody.

3.1 Task and conversational dialogue

Most dialogue research concentrates either on task based dialogue, where the participants converse in order to achieve some set goal (e.g. Maptask (Anderson et al., 1991) Coconut (Di Eugenio et al., 1998)), or on conversational dialogue, which is often less structured and contains a richer use of language (e.g. DAMSL (Core and Allen, 1997) and Chat (MacWhinney, 1998)). It seems likely that we would see more expressions of emotion in conversational dialogue where people are discussing topics of personal interest rather than the more mechanical process of achieving a goal through communication.

These differences are reflected in the types of phenomena that the schemes are designed to identify. Task based research may be more interested in the structure of the dialogue and the way that it represents the division of the task into sub-goals. Schemes to annotate conversational dialogue are more likely to require a greater breadth of dialogue acts to describe the wider range of illocutionary acts that may be performed in this type of speech.

3.2 Current dialogue annotation schemes

In order to learn how current annotation schemes accommodate emotion, we aligned the layers in a number of schemes. (Core and Allen, 1997; Di Eugenio et al., 1998; Traum, 1996; Walker et al., 1996; MacWhinney, 1996; Jekat et al., 1995; Anderson et al., 1991; Condon and Cech, 1996; van Vark et al., 1996; Walker and Passonneau, 2001). Layers from different schemes are grouped according to the similar phenomena that they label. Table 1 shows this alignment.

In this section we will look at these layers and discuss how they may relate to annotating emotion in dialogue.

Information level When analysing task dialogue, we may be interested in knowing whether an utterance pertains to the management of the communications channel, advancement of a task, discussing of a task etc. In the previous section we suggested that we are more likely to find emotional speech in conversational rather than task dialogue because the latter is more of a mechanical process than conversation. Perhaps we may consider this layer as an extension of that distinction where sub-dialogues are labelled according to how closely related to the task they are.

This may reveal a correlation where the more related to the task a sub-dialogue is, the less emotional speech becomes. There is evidence in our corpus that when one participant is attempting to achieve

a goal, often the elicitation of information, then the participants' behaviour becomes more business like and the language becomes more formal and less expressive –

```
N. Was that Dr Smith who you saw there?
P. Yes but it wasn't Dr Smith it was an-
other doctor.
N. Right.
P. But I was under Dr Smith.
N. Right.  So when did you actually have
those radiotherapy treatments?
P. I had the radiotherapy October 13th.
N. Right, thank you.
```

Communications status

Communications status indicates whether an utterance was successfully completed. It is used to tag utterances that are abandoned or unintelligible rather than whether the intention of a speech act was achieved.

Although failure to perform a successful utterance may be partly due to the emotional state of the speaker, annotating such utterances for their emotional content may be difficult, especially from the textual content alone. This and the multiplicity of reasons for unsuccessful communication means that using communications status as an indication of emotion in the speaker will produce unreliable results.

In Human-Machine dialogue failure on behalf of the machine to communicate can lead to frustration and anger in the user. In these cases communication status may signal behaviour that can result in emotion in the listener which is also applicable to Human-Human dialogue.

Speech acts

All of the schemes that we examined annotated the utterances for their illocutionary force. Since this is the layer that contains most information regarding the semantic content of an utterance, this is likely to be where we shall find the most interesting correlations. We have already seen that high levels of emotion in dialogue alters the frequencies of dialogue acts compared with the more impassive conversations conducted in the Switchboard corpus

(Wood and Craggs, 2002).

Forward communicative functions describe utterances that intend to evoke some response from the listener (such as believing a statement or answering a question), perform an action (such as committing to something) or similar dialogue advancing functions. These types of utterance are likely to be motivated by some intention or belief on behalf of the speaker, providing clues as to their cognitive state.

Forward communicative functions can play an important role in eliciting emotional responses from the listener. Open ended questions are more likely to produce an emotional response than a yes/no question (Maguire et al., 1996b). This is partly because open questions hand the initiative to the listener allowing them to express themselves. The relationship between questions, initiative and emotion is discussed further in (Wood and Craggs, 2003).

The following extract from our corpus show how an open question elicits an emotional response from the listener.

```
N. How were you coping with that yourself?
P. Oh mentally I've never been down men-
tally
```

Backward communicative functions are used to label utterances that respond to something that has been said to them. Some responses are required by the previous utterance, for example an answer following a question. In these cases the utterance wasn't motivated by a desire on behalf of the speaker but rather an obligation to adhere to the rules of engagement for communication. This of course doesn't mean that a response can not be emotional. When faced with a proposal, question or offer the listener is free to react as they wish and this includes emotional responses.

Here a backward communicative act responds to appreciation with sympathy.

```
N. Good okay.  Well thank you as I say for
filling me in and...
P. poor girl you've got to listen to all
that
```

However, from observations of our emotional

dialogues it appears that short Question-Answer, Offer-Acceptance exchanges tend to be formal. Emotion tends to build though a sub-dialogue on a topic that speakers find funny, feel anxious about etc.

Dialogue grammars are used to exploit the expected sequences of speech acts. These can be used in dialogue act classification to predict the next act in a series of utterances (Stolcke et al., 2000). It may be possible that a complementary approach may be used to automatically identify emotional utterances.

One way would be to develop grammars based on patterns discovered in emotional sections of dialogue where a particular sequence of acts may indicate the proceedings have become emotional. Another may be to apply established grammars to dialogue so that deviations from the grammar may highlight interesting or emotional passages.

Topic

Several annotation schemes contain a layer that labels the topic discussed in an utterance. This is usually in task domains where there is a finite number of subjects that will be discussed. For example in the Alparon scheme for transport dialogues (van Vark et al., 1996), the topic layer (called 'coding of information'), labels utterances according to whether they relate to topics such as *timetable, price, time* and *locations*.

For our corpus of cancer consultations it is apparent that certain topics are more likely to invoke emotion in people. However topic annotation is only usually performed in the restricted domains of task dialogues, where the range of topics that may be discussed is limited. However it is in these types of dialogue that we expect the levels of emotion to be low, and topics are chosen because of their necessity for the task. Because of this we may not get to see the correlation between topic and emotion that we expect.

Topics may play a further role in identification of emotion in dialogue since in our corpus, patients tend to remain on the same topic for longer when they emotional about it. Length of a topic, or returning to a previously discussed topic are indications of emotion.

Phases

Some schemes distinguish between dialogue phases such as *opening, negotiation* and *query*. Emotion in dialogue also goes through phases and it is possible that there are boundaries between the phases of emotion that correspond to those tagged using the phase layer.

An interesting area of research would be to identify how boundaries between the phases of different levels and types of emotion are manifested in the use of language. For instance psycho-oncology research states that open ended questions are more likely to elicit emotional responses than yes-no questions (Maguire et al., 1996a). This may cause a correlation between forward-looking functions and the onset of phases.

Surface form

Surface form tagging is used in David Traum's adaptation of the TRAINS annotation scheme (Traum, 1996) and the Coconut scheme to tag utterances for certain special features such as cue words or negation.

It has been shown that certain syntactic features of an utterance may be indicators of emotion. For example in German use of modal particles such as '*eben*' and '*denn*' colour the utterance with a particular emotional attitude.

Although the surface form of utterances is dependent on the style of the speaker, it does sometimes contain indications of emotion.

```
P. Oh no, no no no no, I'm not in any dis-
comfort
```

Relatedness and Information relations

The relatedness layer is used to show how utterances relate to one another, usually by tagging an utterance with the distance to the antecedent to which it refers.

Information relation describes the relationship between utterances, for instance that one utterance presents information in support of its antecedent.

These layers are more concerned with the structure of the dialogue than the semantic content and are therefore less likely to correlate well with emotional tags.

It would be interesting to see if the emotional

level of the dialogue or its participants has an affect on the dialogue's structure. In our corpus it appears that discussion of emotional topics is often more protracted, with speakers answering questions with successive statements, each adding more detail to their answer. This type of behaviour may show up in the relatedness and information relations layers.

Grounding

Grounding describes the process by which common ground between the participants is established. As with relatedness and information relations, emotion in the dialogue may be manifested in this layer by protracted grounding behaviour as people reiterate points about which they feel emotional. In our highly emotional corpus this resulted in four times as many summaries and five times as many repetitions than in the Switchboard corpus.

Besides the layers listed here there are other layers included in schemes that do not fit into any of these categories. For instance Verbmobil (Jekat et al., 1995) includes a layer for annotating the propositional content of an utterance, and content relevance in the Penn multi-party coding scheme (Walker et al., 1996). Investigation on dialogue annotated for emotion will show whether there are any interesting correlations with these layers.

4 Emotional speech corpora

One of the difficulties in analysing emotion in communication is in obtaining the material to study. For studies into task dialogues, researchers can simply record speakers performing the tasks. However capturing conversational dialogue in general and especially emotional dialogue is a much more difficult task.

Studies into emotional speech based on acoustic features use three approaches to attain their data. Ideally it is preferable to use genuine speech taken without the speaker's knowledge since you can be confident that the resulting data will faithfully represent human behaviour. An example of research using this type of data is (Scherer and Ceschi, 2000). This approach isn't commonly adopted, partly because of the ethical issues concerned with recording people without their consent and also because of the

difficulty in controlling variables such as recording quality or establishing age, sex, etc. of the speaker.

For dialogue studies this would also be the desired type of data. If we are interested communicative behaviour such as turn-taking and language use rather than the acoustic features of the speech then we need not be so concerned with the acoustic quality. If it were possible to obtain recordings of police interviews, legal trails or calls to emergency services then these would provide suitable material to study. Our corpus of oncology consultations is a good example of this type of dialogue.

A more common type of data used in speech studies is that of acted emotions. Actors deliver lines expressed with different emotions (e.g. (Dellaert et al., 1996)). The quality of this data is reliant on the accuracy with which the emotion is acted. This is suitable for establishing the prosodic features associated with various emotions but not for dialogue studies. It would be much more difficult to recreate the communicative behaviour of an emotional person through acting than to simply sound emotional.

Finally, induced emotion, where participants are provoked into an emotional state so that their speech can be recorded (e.g. (Huber et al., 2000)) . This provides natural emotion within a laboratory setting. It is conceivable that this process could be adapted to obtain induced emotional dialogue. One participant may try to conduct a conversation during which the other may behave emotionally. However it is likely that the data derived from this would be unlike real conversations.

It is apparent that when studying emotion in dialogue it would be desirable to obtain genuine conversations that contained some degree of emotion. Attempting to induce emotion is likely to cause the communicative behaviour to become unnatural. The preferable option would be to use natural conversation in unusually emotional circumstances such as those described above.

5 Toward an emotion annotation scheme

In developing an annotation scheme our first step will be to decide on the facets of emotion which we would like to identify. Emotion is a very vague word and so it is important that we polarise it into clear and understandable aspects of human cognition. In

Layer	DAMSL	Coconut	Traum	Penn	Maptask
Info level	Info Level	Info Level		Info Type	
Comm status	Comms Status				
Topic		Topic		Topic	
Speech act	Dialogue acts	Comm function	Illocutionary func	Speech acts	Moves
Info relations	Info relations	Info relations	Argumentation		
Relatedness	Antecedent	Link	Relatedness	Initiative	
Grounding			Grounding	Info status	
Surface form		Surface features	Surface form		
Phases					

Layer	Verbmobil	Chat	Condon & Cech	Alparon	Date
Info level		Interchange type	Metalanguage		Domain
Comm status					
Topic				Info coding	Subtasks
Speech act	Dialogue acts	Illocutionary force	Move function	Moves	Speech acts
Info relations					
Relatedness					
Grounding					
Surface form					
Phases	Phases			Phases	

Table 1: Annotation schemes and their layers

order for the annotation to be useful these aspects must have some influence on their communicative behaviour. They must also be identifiable from the language of the dialogue. This will mean that the scheme may consist of several layers each describing a different aspect of human emotion.

One of the differences between these types of layers and those current schemes is that rather than discrete categories such as those used to label speech acts we can observe varying levels of emotion. A precedent for this type of annotation exists in the labelling of expressions of concern in the oncology consultation coding scheme of Psychological Medicine Group at Manchester University (Heaven and Green, 2001). where these cues are rated 0–3. If this approach was adopted then we would have to decide on the number of levels to chose from based on a trade-off between ease of performing annotation with getting a fine enough distinction between different levels.

This would allow us to draw conclusions about communicative behaviour under different levels of emotion (e.g. *"The length of utterances becomes longer under increasing levels of anxiety"*) and correlations with other layers (e.g. *"People ask open questions when relaxed but closed questions when agitated"*). It would also allow us to plot the quantitative level of emotions throughout the dialogue, investigate the way in which this changes and identify the language phenomena that signal these changes.

If only for pragmatic reasons, it would be wise to choose utterances as the basic unit for annotation. By utterances here we refer to the common understanding described as 'a sequence of communicative behaviour bounded by lack of activity' (Allwood, 1996). This would not only allow us to apply other schemes to emotionally annotated dialogue, but also to use tools that have been developed to work on utterances. It would therefore be necessary to chose dimensions of emotion that can be applied to utterances.

There is an interesting question of whether emotion is a property of the participants or the dialogue. Obviously two or more people participating in a dialogue will react differently to the proceedings and will therefore exhibit different emotions. However

it is apparent from our corpus that the dialogue itself has its own levels of emotion. For instance, the conversation may go through a phase of solemnity during which the participants may exchange a joke. The mood of the dialogue outlives this perturbation and remains serious. It would appear that it may be useful to track the emotional state of the dialogue as well as the speakers since one will clearly have an effect on the other. Quantitative annotation and analysis of the flow of these levels would therefore be useful here too.

6 Future work

Our next step will be to design an annotation scheme based on the observations and principles stated throughout this paper. We could then start annotating our corpus for the emotional dimensions that we had chosen.

In order to assess the correlations that we proposed might exist in section 3, we would have to annotate these dialogues with the layers of other schemes. Since none of schemes contain all of the layers, we would have to combine individual layers based on our beliefs about which could be most useful and the ease with which we would annotate the dialogue. It would make sense to select layers from schemes which have comprehensive coding manuals, which have been shown to be reliable and which would be accommodated by annotation tools.

Before any claims about the effects of emotion in dialogue can be made, the reliability of the scheme must be established. Once this has been achieved than analysis of the results can begin.

An annotated corpus would present us with the opportunity to investigate correlations and attempt to identify the effects the various types of emotion on the behaviour of the participants. It is likely that along with the possible effects that we have proffered in this paper there will be other interesting patterns that become apparent from the results of our annotation. This will improve our understanding of behaviour in dialogue and benefit dialogue applications.

References

J Allwood. 1996. On dialogue cohesion. In *Papers from Thirteenth Scandinavian Conference of Linguistics*.

A Anderson, M Bader, E Bard, Boyle E, G Doherty, S Garrod, S Isard, J Kowtko, J McAllister, J Miller, C Sotillo, H Thompson, and R Weinert. 1991. The HCRC Map Task corpus. *Language and Speech*, 34:351–66.

T Bub and J Schwinn. 1996. VERBMOBIL: The evolution of a complex large speech-to-speech translation system. In *Proc. ICSLP '96*, volume 4, pages 2371–2374, Philadelphia, PA.

S Condon and C Cech, 1996. *Manual for Coding Decision-Making Interactions*.

Mark G Core and James F Allen. 1997. Coding dialogs with the damsl annatation scheme. In *AAAI Fall Symposium on Communicative Action in Humans and Machines*.

F. Dellaert, T. Polzin, and A. Waibel. 1996. Recognizing emotions in speech. In *Proc. ICSLP '96*, volume 3, pages 1970–1973, Philadelphia, PA.

B Di Eugenio, P.W Jordan, and L Pylkkanen, 1998. *The COCONUT project: Dialogue Annotation Manual*. ISP Technical Report 98-1.

K Fischer. 1999. Annotating emotional langauge data. Technical report, Verbmobil - report 236.

C Heaven and C Green, 2001. *Medical Interview Aural Rating Scale, CRC psychological medical group*, October.

R. Huber, A. Batliner, J. Buckow, E. Nöth, V. Warnke, and H. Niemann. 2000. Recognition of Emotion in a Realistic Dialogue Scenario. In *Proc. Int. Conf. on Spoken Language Processing*, volume 1, pages 665–668, Beijing, China, Oktober.

S Jekat, A Klein, E Maier, I Maleck, M Mast, and J Quantz. 1995. Dialogue acts in Verbmobil. Technical Report 65, DFKI Saarbrucken, Universitat Stuttgart, Technische Universitat Berlin, Universitat des Saarlandes.

B MacWhinney, 1996. *The CHILDES System*. Carnegie Mellon University.

B MacWhinney, 1998. *The CHILDES Project: Tools for Analysing Talk*. Carnegie Mellon University.

P Maguire, K Booth, C Elliott, and B Jones. 1996a. Helping health professionals involved in cancer care work shops acquire key interviewing skills – the impact of workshops. *European journal of cancer*, 32A(9):1486–1489.

P Maguire, K Faulkner, K Booth, C Elliot, and V Hillier. 1996b. Helping cancer patients disclose their concerns. *European Journal of Cancer*, 32(9):1486–1489.

V. Petrushin. 1999. Emotion in speech: Recognition and application to call centers.

K.R Scherer and G Ceschi. 2000. Criteria for emotion recognition from verbal and nonverbal expression: studying baggage loss in the airport. *Personality & Social Psychology Bulletin*, 26(3):327, March.

A. Stolcke, 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*, 26:339–373.

D Traum, 1996. *Coding Schemes for Spoken Dialogue Structure*.

R.J van Vark, J.P.M de Vreught, and L.J.M Rothkrantz, 1996. *Analysing OVR dialogue coding scheme 1.0*. Delft University of Technology.

M Walker and R Passonneau. 2001. DATE: a dialogue act tagging scheme for evaluation of spoken dialogue systems. In *Proceedings: Human Language Technology Conference*, San Diego, March. AT&T Shannon Labs.

M Walker, E Maier, J Allen, J Carletta, S Condon, G Flammia, J Hirschberg, S Isard, M Ishizaki, L Levin, S Luperfoy, D Traum, and S Whittaker, 1996. *Penn multiparty standard coding scheme, Draft annotation manual*.

M.M Wood and R Craggs. 2002. Rare dialogue acts in oncology consultations. In *Submitted to SIGdial3*.

M.M Wood and R Craggs. 2003. Initiative in health care dialogues. In *Submitted to DiaBruck 7th workshop on the semantics and pragmatics of dialogue*.

M.M Wood. 2001. Dialogue tagsets in oncology. In *Proceedings of Sigdial2*.