Dialogue tagsets in oncology

Mary McGee Wood
Department of Computer Science
University of Manchester
Manchester M13 9PL U.K.
mary@cs.man.ac.uk

Abstract

Dialogue analysis is widely used in oncology for training health professionals in communication skills. Parameters and tagsets have been developed independently of work in natural language processing. In relation to emergent standards in NLP, syntactic tagging is minimal, semantics is domain-specific, pragmatics is comparable, and the analysis of cognitive affect is richly developed. We suggest productive directions for convergence.

1 Motivation

Dialogue analysis systems have been developed in oncology as a tool for assessing and improving the communication skills of health professionals. Rates of psychiatric morbidity (clinical anxiety and depression) in cancer patients are lowered when health professionals have adequate communication skills to discover and address the patients' concerns and worries. Health professionals interviewing patients sometimes exhibit negative behaviours, such as "blocking" a certain line of investigation rather than encouraging the patient to describe his or her problem. On the other hand, a skilled interviewer uses active interventions to direct the progress of the interview, as well as more passive responses. Several oncology research groups have demonstrated that these patterns can be detected and quantified through analysis of conversations between health professionals and patients. This in turn can form a basis for more effective training in communication skills.

The Psychological Medicine Group at Manchester (PMG), funded by the Cancer Research Campaign (CRC), is a leading group in dialogue analysis in oncology. This paper describes the parameters and tagsets (analogous to "Dialogue Act" tagging (Stolcke et al 2000)), which they and three other groups have developed for this highly specialised domain.

This domain offers an interesting contrast to the "instructional" or "service" dialogues commonly studied. The health professional is the "expert" in the conventional sense, and at times conveys medical information to the less knowledgeable patient in a conventional way. At other times, the patient should be seen as the "expert" with regard to his or her own perceived physical and mental condition, and the task of the health professional is effectively that of "knowledge elicitation" as understood in expert systems development. This flexible and dynamic shifting of participants' roles in a dialogue poses an interesting challenge, compared to the clearly defined and static roles assumed in much work in dialogue analysis.

2 Parameters for dialogue tagging

Complete and accurate tagging of dialogue must encode a number of independent aspects of each utterance. These are represented as "layers" in the DAMSL system (Core & Allen 1997). Form-based tags (question, statement) are supplemented with diacritics indicating other types of information, such as task-management or communication-management.

The four oncology dialogue tagging systems considered here all share this basic principle, although they differ in the specifics. Butow et al (1995:1115) cite the recognition as early as 1983 of "layers of meaning ... such as the content, the process, the emotion and the purpose". Their own CN-LOGIT system encodes three "dimensions": "source" (who is speaking), "process" (questions, responses, initiated statements), and "content". A complete dialogue can be mapped into a three-dimensional information space, and measures can be applied such as how much time was spent in each cell of the cube. Ong et al (1998) use the Roter Interaction Analysis System (RIAS). Each utterance in a dialogue is categorised, and also rated on five distinct "global

affect" scales. The Medical Interaction Process System (MIPS) of Ford et al (2000) also stresses the multi-dimensional nature of dialogue annotation, using fifteen "content codes" and eight "affective modes". PMG (Maguire & Faulkner 1988; Maguire p.c.) have separate tagsets for Form, Function, Content, Level, Cue, Cue Management, Blocking, and Focus.

One can see an implicit consensus here that (to use NLP terms) syntactic form, overt semantic content, pragmatic force, and cognitive affect are distinct and are all significant. The differing degrees of detail and prominence they receive in the different systems are discussed under those headings in the next section.

3 Dialogue tagsets

Not surprisingly, the actual tagsets developed in oncology reflect their domain more closely than the parameter sets do. In comparison with NLP work, syntactic classification is minimal and functionally oriented, while communication management and psychological / emotional loading receive prominent, fine-grained analysis.

3.1 Form

Although all four oncology systems encode the form of an utterance in some way, the classifications have a strong pragmatic bias. Questions are distinguished, not in traditional syntactic terms as yes-no or wh-, but according to their effect on the flow of the dialogue. The simplest set is that of Butow et al: Open Question, Closed Question, Response to Question, Statement, Other. PMG add Directive Question (open), Directive Question (closed), Screening Question, Leading Question, Multiple Question.

Ford et al distinguish "modes" from "content codes", but even the modes encode coarse-grained content information as well as affective classification. The form categories of Ong et al are "instrumental" (Directions, Question-asking, Information-giving, &c), and they specify that "if a decision must be made between categorizing an utterance in an instrumental or affect category, the affect category should be used" - quite reasonably, given the purpose of their analysis.

Even with a prior commitment to maintaining separate and independent levels of analysis, some leakage between levels can occur. (The set of forty-two Dialogue Act labels used by Stolcke et al (2000) shows some similar mixing of levels, including both purely syntactic tags (such as Declarative Yes-No Question) and affective tags (such as Appreciation).)

3.2 Content

The content of an utterance is also encoded in all four systems, and the tagsets on this level are the most domain-specific. Butow et al cite seven content categories: Treatment, Diagnosis, Prognosis, History, Other medical matters, Social matters.

Ford et al, with 15 content codes, and PMG, with 38, are the most fully developed. Both include Medical (further distinguished by PMG, with four categories for diagnosis and two for prognosis), Treatment, Psychological, Social, Lifestyle, &c. PMG are particularly detailed in their categories for psychological and emotional issues, shading into the affect level: Concerns, Feelings, Emotions, Religion, &c. Again, this is what one would expect, given that their reason for carrying out the analysis is to assess the health professional's success in getting the patient to talk about exactly these issues.

Both Ford and PMG also include the opening and closing of the interview under this heading, where it sits oddly. A separate level of communication management, as in DAMSL, would accommodate these and the open/ closed/ directive question distinction currently made in the Form tagsets, clarifying all three.

3.3 Pragmatics

As noted above, the Form classes used in the four coding schemes express more pragmatic than syntactic information. Ong et al's "instrumental clusters and categories" (Directions, Question- asking, Information-giving, Counselling) can be considered pragmatic. So can PMG's "Function" codes: eliciting, checking, acknowledgement (psychological, general, cognitions); reassurance, negotiation, information giving. These are similar to some of the Dialogue Act labels used in NLP work: Stolcke et al's (2000) agreement, response acknowledgement, summarize, or VERBMOBIL's suggest, confirm, clarify (Jekat et al 1995).

3.4 Affect

Cognitive affect - the psychological force, for a patient, of an utterance or a complete dialogue - is the focus of interest in oncology and thus the most highly developed area. Ford et al pick out eight of their "modes" as affective, including the expression of irritation, gratitude, apology, and concern.

Ong et al rate both doctor and patient, by coding their utterances, on five distinct "global affect" scales: Anger/ irritation, Anxiety/ nervousness, Dominance/ assertiveness, Interest/ engagement, Friendliness/ warmth. Their "affective clusters

and categories" comprise (with subheadings) social behaviour, verbal attentiveness, showing concern, and negative talk.

PMG do not represent affect as a separate parameter, as such. Their function codes include affective functions such as Empathy and Reassurance. Many of their content codes can also represent affect, as noted above. Topics such as Concerns, Feelings about health care, Religion / spiritual issues can be addressed at any level from simply factual to deeply emotional, blurring the picture: this would be clarified if the affect level were explicitly factored out. The most direct representation of affective level comes in the two codes Psychological explicit and Psychological implicit. Each utterance in a dialogue can be given several content codes, commonly including one of these two, as seen in the sample dialogue below.

Cognitive affect has barely been touched on by NLP research in dialogue tagging. It is clearly more subtle and difficult than syntactic, semantic, or pragmatic analysis, and also less significant in instructional or service dialogues than in the highly charged, life-critical domain of cancer care. It is, however, an important aspect of dialogue and speaker modelling, and of the design of appropriate responses. In this area, NLP could learn some valuable lessons from oncology.

4 An example

Here is a brief typical example from a PMG annotated dialogue. Notice the multiple and somewhat diverse content codes, and the classification of cue management (somewhat counter-intuitively attached to the cue utterance itself, not the response).

P26: I said there's only another thing that I hope I never have to have and that's selectron treatment.
Content: 23 (Psychological implicit)
   24 (Treatment)
Level: 1 (Hint)
Cue: 1 (Patient cue)
Cue management: 4 (Cue explored)

N27: Mmmm.
Form: 02 (Response)
Function: 06 (Acknowledgement - general)
Content: 24 (Psychological implicit)

P27: But I says … … if I have to I will do, I said whatever you say, I said.
Content: 24 (Treatment)
   14 (Information)

N28: Now why did you say that about selectron?
Form: 04 (Open directive question)
Function: 04 (Clarification - general)
Content: 24 (Treatment)

P28: No it were me that said that.
Content: 24 (Treatment)

N29: Right,
Form: 02 (Response)
Function: 06 (Acknowledgement - general)
Content: 24 (Treatment)

have you had it before?
Form: 05 (Directive question (closed)
Function: 04 (Clarification - general)
Content: 24 (Treatment)
   13 (History)

P29: I said, I, I don't mind what you do but I hope I've never to have selectron treatment again, but I said if I have to, if it's a necessity then I will.
Content: 23 (Psychological implicit)
   24 (Treatment)
   33 (The future)
Level: 1 (Hint)
Cue: 1 (Patient cue)
Cue management: 2 (Minimal acknowledgement)

N30: Right.
Form: 02 (Response)
Function: 06 (Acknowledgement - general)
Content: 24 (Treatment)
   14 (Information)

P30: But I hope I never have.
Content: 23 (Psychological implicit)
   24 (Treatment)
   33 (The future)
Level: 1 (Hint)
Cue: 1 (Patient cue)
Cue management: 4 (Cue explored)

N31: And why was that, because you were isolated or what was it….?
Form: 08 (Multiple question)
Function: 03 (Clarification - psychological)
Content: 23 (Psychological implicit)
   24 (Treatment)

5 Implementations

All the dialogue tagging work in oncology has been done manually. A few primitive software tools have been developed to support the tagging process and to analyse the results.

Ong et al (1998) have developed a Turbo Pascal "computerized version" of the RIAS coding system. The advantages they claim for it give one some idea of average state of its field:

"With this program, classification of utterances can be done directly on computer. As a result, the extra step of entering paper and pencil data into the computer is omitted. Also, sequential information is kept. Moreover, because the ten last classifications are constantly visible on the screen there is direct feedback about the ongoing conversation. This provides an important memory aid with respect to which utterance has to be coded next. As a consequence, the task becomes less attention demanding and therefore less error- prone. By giving the opportunity to save the content of the last coded utterance, an additional memory aid is provided for shorter and longer breaks." (Ong et al 1998:400)

Butow's group have developed a "computer-based interaction analysis system" with three parts: "(i) micro level analysis coded in real time and retaining the sequence of events, (ii) event counts and (iii) macro level analysis of consultation style and affect" (Butow et al 1995:1116). "At the micro level the aim is to break the consultation down into its components and to characterise, count and/or time them... At the macro level, the aim is to characterise the consultation in a more holistic way, such as patient-centred *vs* doctor-centred, authoritarian *vs* affiliative or friendly *vs* hostile." (ibid:1115) All three forms of analysis depend on counting and timing utterance-events classified according to the three-dimensional model described above, although Butow et al stress that they also "retain the sequence of events". "In future analyses we will explore sequential information effects" (ibid:1120). This is evidently a significant innovation in its field. The fundamental concept of a *grammar* of dialogue is simply missing from the oncology work. On the other hand, their techniques for "macro-level" analysis of dialogues may well have something to offer, especially in the subtle areas of modelling and adapting to speakers' attitudes and underlying intentions.

6 Prospects

All this work has been developed with care, in the light of experience, to serve a specific and unusual purpose. However, it shows no awareness of dialogue tagging work in NLP. Both fields can benefit from collaboration.

The author, together with Prof. Peter Maguire of PMG, has recently been granted support by CRC to develop practical software support for the PMG oncology dialogue annotators. This paper presents a preliminary analysis, part of a feasibility study for that project. An associated PhD studentship, awarded by the University of Manchester Department of Computer Science, ensures that the NLP perspective will be represented and the theoretical issues addressed. We look forward to presenting more detailed analyses, and original proposals, in the future.

References

Butow, PN, SM Dunn, MHN Tattersall, & Q. J. Jones. 1995. Computer-based interaction analysis of the cancer consultation. British Journal of Cancer 71:1115-1121.

Core, M. & J Allen. 1997. Coding Dialogs with the DAMSL annotation scheme. AAAI fall symposium on Communicative Action in Humans and Machines.

Ford, S, A Hall, D Ratcliffe, & L Fallowfield. 2000. The Medical Interaction Process System (MIPS): an instrument for analysing interviews of oncologists and patients with cancer. Social Science and Medicine 50:553-566.

Jekat, S, A Klein, E Maier, I Maleck, M Mast, & J Quantz. 1995. Dialogue acts in VERBMOBIL. Verbmobil-Report 65, Universitat Hamburg et al.

Maguire, P & A Faulkner. 1988. How to improve the counselling skills of doctors and nurses in cancer care. British Journal of Medicine 297, 847.

Ong, LML, MRM Visser, IPM Kruyver, JM Bensing, A van den Brink-Muinen, JML Stouthard, FB Lammes & JCJM de Haes. 1998. The Roter Interaction Analysis System (RIAS) in Oncological Consultations: Psychometric Properties. Psycho- Oncology 7:387-401.

Stolcke, A, K Ries, N Coccaro, E. Shriberg, R Bates, D Jurafsky, P Taylor, R Martin, C van Ess-Dykema, & M Meteer. 2000. Dialogue Act Modelling for Automatic Tagging and Recognition of Conversational Speech. Computational Linguistics 26(3):339-373.