Generic speech act annotation for task-oriented dialogues

Geoffrey Leech and Martin Weisser
University of Lancaster and Friedrich-Alexander-Universität Erlangen-Nürnberg

1. Introduction

The annotation of dialogues in terms of speech acts (or dialogue acts) has typically followed one of two paths. One path has been to tailor the speech act categories to a specific task or domain, as in the Edinburgh Map Task annotation scheme (Carletta et al. 1995). Another has been to aim at a more general coverage of dialogue, as in the DAMSL (Dialogue Act Mark-up in Several Layers) annotation scheme (Allen and Core 1997). We report here on a pilot project (supported by EPSRC grant GR/R37142/01) which has attempted to achieve a middle ground between the aims of genericity and specificity in data coverage. The SPAAC (SPeech Act Annotated Corpus) scheme has been developed to annotate a range of different kinds of dialogue, but within the general scope of telephone task-oriented dialogue between two people, as contrasted, for example, with general conversation. To try out the generic potential of the scheme, we have annotated different kinds of dialogue, especially British Telecom and The Trainline service-providing dialogues.[1] A set of 41 speech-act categories has been applied to these sets of corpus data.

As an efficient environment for annotation processing of the dialogues, an XML tool (SPAACy) has been developed by Martin Weisser (further details will be found in Weisser forthcoming 2003 and Weisser forthcoming). Although the 15-month duration of this pilot project did not allow time for a full-scale software development programme, SPAACy enabled us to develop a semi-automatic XML annotation procedure whereby the dialogues were assigned speech-act tags with gradually improving accuracy, before the dialogues were manually post-edited to achieve consistency with practices laid down in a speech-act annotation scheme manual.[2] The result was that 1219 dialogues were annotated, amounting to over 182,300 words.[3]

In addition to an initial assignment of speech-act values, the annotation tool also assigned four supplementary values for each speech act:

(a) *A form category*: declarative, yes-no question, wh-question, imperative, fragment, yes, no.
(b) *A polarity category*: positive, negative
(c) *A topic category*: e.g. (relating to train journeys) location, name, day, date, time, railcard

---

[1] We are grateful to both British Telecom and The Trainline for making this data available for the pilot project. The OASIS Corpus was supplied by British Telecom in a transcribed form. The transcription of The Trainline dialogues was undertaken, using a transcription scheme modelled on that of the OASIS corpus, by Paul Baker at Lancaster University. The initial segmentation and manual post-editing of the dialogues were undertaken at Lancaster by Jeremy Bateman, Jean Forrest, Costas Gabrielatos, David Hooper and Ursula Weinberger. Most dialogues went through a further manual post-editing by Geoffrey Leech, to improve consistency. We acknowledge the valuable advice of Jean Carletta, who with Amy Isard had previously undertaken a separate XML speech act annotation of part of BT's OASIS corpus (consisting of 325 dialogues). The intention is to combine the more domain-specific Edinburgh annotations with the generic Lancaster annotations, and to make this part of the corpus available as a separate entity.

[2] The Speech Act Annotation Scheme document is available on the UCREL website: http://www.comp.lancs.ac.uk/ucrel/. In more detail, the main procedures performed or enabled by SPAACy were:

(1) Automatic conversion of the text files containing the transcription to XML mark-up.
(2) Interactive segmentation of the dialogue into utterance-units (here called C-units).
(3) Automatic assignment of speech-act categories, together with other categories giving information on polarity, topic and mode.
(4) Manual post-editing and correction of speech-act tags.

[3] This set of dialogues was composed as follows:

BT operator calls: 643 dialogues, c. 85,840 words
BT customer services calls: 541 dialogues, c. 72,220 words
The Trainline dialogues: 35 dialogues, c. 24,246 words

The word estimates are based on a sample count.

(d) *A mode category*: e.g. semantic categories such as deixis, probability, reason.

These values supplied syntactic, semantic and pragmatic information which helped the tool to 'home in' on the correct speech act. However, with a few exceptions, these values have not been post-edited after the automatic assignment process has run, so that many errors and information gaps remain. It is likely that most of these supplementary values will be deleted before the corpus is made available to the research community.

2. The rationale for the project

Over the last decade there has been an intense interest in the development of spoken dialogue systems (Grice et al. 2000, Leech and Weisser 2003) with the objective of enabling humans to interact with a computer using the medium most natural and congenial to them: human speech. While existing working systems fall short of natural dialogue in many ways, it may be assumed that one of the means to the goal of human spoken interaction with machines will be achieved through the close study and modelling of real human-to-human dialogues. This is more likely to succeed to the extent that the dialogue is relatively constrained, task-oriented and transactional. A first step in this direction is the collection of corpora of such dialogues, e.g. calls made to a call centre or a telephone service provider by a member of the public. The dialogues we have annotated in the present project are of this kind. The BT OASIS corpus consists of dialogues from two rather different kinds of telephone calls: '100' calls to the operator, and '150' calls to BT residential customer services. The Trainline dialogues are of calls to a call centre providing railway timetable information, ticketing and seat reservation services. These three types of dialogue are sufficiently diverse to test the genericity of the annotation scheme to a reasonable extent, but we still intend to apply the scheme experimentally to a wider range of dialogue data: a task which has so far been only partially accomplished.

Once a suitable corpus has been collected and annotated, it can be used as a training corpus and/or test corpus for the modelling of dialogue behaviour using either statistical, rule/structure-based or hybrid language models. (On statistical dialogue modelling see for example Kita et al. 1996, Nagata and Morimoto 1994, and Reitlinger and Klesen 1997.) Some annotation schemes (e.g. Edinburgh HCRC Map Task scheme) already build into an annotated corpus a higher-level analysis of structure, analogous to the phrase markers of grammatical structure, or to the Sinclair and Coulthard (1975) framework for analysing classroom discourse. However, our pilot project confined itself to the more limited goal of speech-act annotation on a single level of sequential dialect acts, leaving to future research the formulation of grammars adding hierarchical levels of structure.

The paucity of existing corpus data for the training and testing of dialogue models was one of the motivations for our speech-act annotation project. Although call centres have burgeoned in recent years and carrying out service transactions by telephone dialogue is one of the growth industries of the present day, the fact remains that such datasets have been difficult to come by for research, particularly in the U.K. This is partly because of problems relating to confidentiality and the Data Protection Act. Another reason is that firms investing in dialogue system research and development do not particularly want to share their data with others. No doubt this data logjam will be overcome, but for the present, we are fortunate that British Telecom and The Trainline have agreed that their dialogue data may be made available to the research community. It seems important to give priority to the annotation of sufficient dialogue data for the development of robust dialogue models, including statistical models. In fact the SPAAC annotations have been applied to a large number of dialogues: 1219. This number compares very favourably with all but two of previous speech act annotation projects.[4]

While the development of computational models of dialogue was the primary purpose behind the project, we would emphasise the potential of the data for linguistic research in general – for example, to mention one topic, the study of politeness phenomena in these corpora has an obvious appeal. Other topics waiting to be investigated are grounding and dysfluency. Plans are now being made for the release of this data to researchers, and it is intended that these arrangements will be announced in the near future via the UCREL website (http://www.comp.lancs.ac.uk/ucrel/projects.html). The dialogues have been subjected to anonymization routines ensuring that no confidential information is preserved in

[4] The European MATE (Multilevel Annotation Tools Engineering) project supports 16 dialogue act coding schemes, of which the majority had been tested on less than 100 dialogues. The two major exceptions in this regard are SWBD-DAMSL and VERBMOBIL, which had been applied respectively to 1155 and 1172 dialogues (Klein et al. 1998).

the final form of the corpora. It should be pointed out, however, that no sound files accompany the transcribed record of the dialogues.

3. Exemplification of the corpus data

To illustrate the corpus data, mark-up and annotation, we reproduce the first two turns of one of the The Trainline dialogues. For clarity, in this sample, the speech act values and the actual words spoken are printed in *italic*:

```
<turn id="1" speaker="A">
<utt id="1">
<frag sp-act="greet" polarity="positive" topic="opening" mode="greet">
     good afternoon
</frag>
</utt>
<utt id="2">
<frag sp-act="identifySelf" polarity="positive" topic="intro" mode="greet">
     Virgin trainlines Sandra speaking
</frag>
</utt>
<utt id="3">
<q-wh sp-act="reqDirect" polarity="positive" topic="journey-preference" mode="open">
     for which journey do you wish to purchase a ticket
</q-wh>
</utt>
</turn>
<turn id="2" speaker="B">
<utt id="4">
<decl sp-act="direct" polarity="positive" topic="to-location-from" mode="closure">
    it's from London Euston to Birmingham International
</decl>
</utt>
</turn>
```

Within the structure of the overall dialogue, there is a hierarchy of three tags, one embedded in another. The outermost is the “turn” tag, showing the alternation between speakers A (the service provider) and B (the caller). Within this is the “utterance” tag. An utterance may contain a single speech-act unit, or more than one (if one or more of them are discourse-marker or yes or no tags, which may be considered peripheral to the utterance). The speech acts themselves are handled as attribute values of the C-unit, a grammatical unit which corresponds to a speech act in extent, and which is labelled by one of the syntactic class labels “decl” for declarative, “q-wh” for wh-question, etc.

Although this example may suggest a highly ritualised and predictable style of question-and-answer sequence which could be handled by a simple template-filling device, the following section from another of The Trainline dialogues illustrates the greater complexity which arises in a negotiating context with mixed initiatives, competing goals, and frequent dysfluencies. The mark-up has been stripped away from this example, to enable the key elements for this presentation (turns, words spoken, and speech acts,) to show up clearly. [Note: # indicates a pause, and #<*x*s> indicates a pause of *x* seconds.]

A1: … *for which journey do you wish to purchase a ticket* [reqDirect]
B2: *hi* [greet]
    i wanna buy em a ticket for Edinburgh [direct]
    to leave em going on the ninth of October [refer]
A3: *travelling to* [reqInfo]
A4: *Edinburgh* [answ]
B5: *from* [reqInfo]
A6: *well* {#} [init]
    em is there a train from Liverpool [reqInfo]
B7: *now* [init]
     do you hold a current credit or debit card [reqInfo]
B8: *yes* [answ]
A9: *and is it just for one person* [reqInfo]
B10: *it is* [answ]
     yes [answElab]
A11: *so it’s Friday the ninth of October* (*yes*) *from Liverpool to Edinburgh #*
      [confirm]
     do you have a railcard at all [reqInfo]
B12: *i don’t* [answ]
     no [answElab]
A13: *what time around would you like to depart* [reqDirect]

```
B14: well [init]
     saying actually i've got a couple of queries first # [raise-issue]
     me friends are going up and they're leaving from Wigan at 12 28 to Edinburgh
        [inform]
     but they said there wasn't # there was only one a day from Liverpool [inform]
     so that's why I wanna know what time the one from Liverpool is [inform]
A15: well [init]
     you'll be able to get the train from Liverpool Lime Street at 11 35 (yes)
        [inform]
     you have to change at Wigan Northwestern anyway (right) to arrive at 12 22
        [inform]
     then it's the 12 28 from Wigan to arrive in Edinburgh at 15 30 [inform]
B16: oh right [ackn]
     so [init]
     the train from Liverpool goes to Wigan anyway [confirm]
     yeah [init]
A17: yeah [ackn]
B18: alright [ackn]
     so [init]
     i may as well get on with them in Wigan then [express-possibility]
     how much is the ticket [reqInfo]
     can you tell me [reqInfo]
A19: you want a ticket from Wigan now [confirm]
B20: yeah [answ]
     on that train 12 28 [answElab]
A21: #<5s> i'll just check that for you [inform-intent]
B22: cheers thanks [thank]
A23: #<14> there's not any advance purchase tickets left for the ninth now #
        [inform]
     the cheapest fare that's going to be available is the saver return [inform]
     and that's 65 pounds 30 [inform]
```

4. The speech act categories

Previously, a variety of annotation schemes have been devised for different genres of dialogue data, varying from a wide range of non-task oriented dialogue (for example, the SWBD-DAMSL scheme – Jurafsky *et al* 1997), to a single specialized task, such as the HCRC Map Task (Carletta *et al* 1995). The varying order of complexity of these schemes is a reflection of the functional spread of dialogues they are meant to capture. For example, the SWBD-DAMSL scheme, intended for relatively unconstrained conversational dialogue, began with a set of 200 speech-act categories (subsequently narrowed down to 42 –Jurafsky *et al* 1997) whereas a much smaller set of categories (12) was devised for the 'laboratory' dialogue task represented by the HCRC Map Task Corpus. For a scheme designed to be generically applicable to task-oriented dialogue, we devised after experimentation a category set almost of the same size as the SWBD-DAMSL reduced set, of 41 tags: sufficiently rich to capture a wide range of dialogue types, yet sufficiently parsimonious to engender controllable and consistent annotation practices. Table 1 summarises the set of 41 categories:

Table 1 – SPAAC Classified List of Speech Acts [O = the other speaker; S = self]

label	***superordinate class***	***broadly characterised as***
accept	*mainly responding*	responding in an active positive way
ackn(owledge)	*mainly responding*	signalling decoding, understanding
answ(er)	*mainly responding*	answering a question
answ(er)Elab(orate)	*mainly responding*	elaborating the answer to a question
appreciate	*mainly responding*	expressing appreciation
bye	*interpersonal management*	saying farewell; closing a dialogue
complete	*dialogue control*	completing O's move
confirm	*responding / initiating*	repeating what O has said/implied
correct	*dialogue control*	correcting what O has said
correct-self	*dialogue control*	correcting one's own utterance
direct	*mainly initiating*	eliciting O's non-verbal response
directElab(orate)	*mainly initiating*	elaborating a "directive" move
echo	*dialogue control*	repeating O's words for verification
exclaim	*expressive*	expressing emotion
expressOpinion	*expressive*	expressing an opinion/evaluation
expressPossibility	*expressive*	expressing a possibility

expressRegret	*expressive*	expressing regret; apologizing
expressWish	*expressive*	expressing a wish or desire
greet	*interpersonal management*	greeting: *hello; hi; good morning* etc.
hold	*dialogue control*	signalling to O to hold the line
identifySelf	*dialogue control*	identifying S's name/institution
inform	*mainly initiating*	conveying information/awareness
informIntent	*mainly initiating*	indicating S's intention
informIntent-hold	*dialogue control*	(a combination of both moves)
init(ialize)	*dialogue control*	initiating a new phase of the dialog
negate	*mainly responding*	responding negatively
offer	*mainly initiating*	offering a service to benefit O
pardon	*dialogue control*	signalling the need for O to repeat
raiseIssue	*mainly initiating*	raising an issue (non-informative)
refer	*mainly initiating*	indicating a reference
refuse	*mainly responding*	responding negatively to an offer, etc
req(uest)Direct	*mainly initiating*	requesting a directive
req(uest)Info(rm)	*mainly initiating*	requesting verbal information
req(uest)Modal	*mainly initiating*	requesting permission, advice, etc
selfTalk	*external to dialogue goals*	speaking to oneself (S)
suggest	*mainly initiating*	proposing action by O (or O and S)
thank	*interpersonal management*	thanking
thank-bye	*interpersonal management*	thanking and saying farewell
thirdParty (talk)	*external to dialogue goals*	speaking to s.o. who is not S or O
unclassifiable	(unspecified)	(an unclassifiable move – e.g. a joke)
uninterpretable	(unspecified)	(because incomplete or incoherent)

[Note: Parentheses enclose what has been omitted from the abbreviated labels represented in the annotation schme.]

The second column of the table assigns the 41 moves to a set of 5 superordinate categories, including metadiscoursal categories containing speech acts concerned with the control of the dialogue and the management of interpersonal relations.

This set of categories owes a great deal to preceding proposed sets of speech act categories, particularly those of DAMSL (Allen and Core 1997) and VERBMOBIL (Alexandersson *et al*. 1997). The distinction between 'mainly initiating' and 'mainly responding' particularly applies to speech acts concerned with the goals of the dialogue itself, and recalls a similar classification in the DAMSL and HCRC annotation schemes. However, it is less absolute: for example, the apparently anomalous that the "confirm" speech act type is janus-faced, i.e. both initiating and responing. It confirms what has previously been said, and at the same time seeks an opportunity for further confirmation or acknowledgement from the other speaker. Another case is that of discourse markers such as *right, alright,* and *okay,* which are frequently used to demarcate the end of the main business of the phone call: in such a context, they appear both to respond (signalling the end of previous discussion) and to initiate (pointing forward to a new beginning). They could have been assigned both 'ackn' and 'init' labels, but in practice we decided to mark them as 'init': their demarcation function appears to apply to the previous transaction as a whole, rather than to a preceding individual turn.

5. Conclusion

This has been a pilot project, and therefore lacks conclusiveness: there are improvements and further research tasks to undertake. For example, should more than speech act label be assigned to the same piece of dialogue? We avoided this except in very clear cases. Also, we did not deal conclusively with the problem of indirect speech acts, where again one could argue that more than one label should be assigned to the same speech unit, one for the surface or 'literal' interpretation, and one for the underlying pragmatic intent (as in Stiles's annotation scheme – Stiles 1992) . Although there is more to be done on the development of the SPAACy tool, it is hoped that these can be added incrementally, either by ourselves or by other researchers working on the data. Meanwhile, these annotated corpora will already be a useful resource for research on task-oriented human-human dialogue.

References

Allen, J and Core M 1997 Draft of DAMSL: Dialog Act Markup in Several Layers
http://www.cs.rochester.edu/research/cisd/resources/damsl/

Alexandersson J, Buschbeck-Wolf B, Fujinami T, Maier E, Reithinger N, Schmitz B and Siegel M 1997 Dialogue Acts in VERBMOBIL-2. VM-Report 204, DFKI GmbH, University of Saarbrücken.

Carletta J C, Isard A, Isard S, Kowtko J, Doherty-Sneddon G, Anderson A. (1995). The coding of dialogue structure in a corpus. In J. A. Andernach, S. P. van de Burgt, & G. F. van der Hoeven (Eds.), *Proceedings of the Twente Workshop on Language Technology: corpus-based approaches to dialogue modelling* . Universiteit Twente, Enschede, The Netherlands.

Gibbon D, Moore R, Winski R (eds) 1997 *Handbook of standards and resources for spoken language systems.* Berlin, Mouton de Gruyter.

Grice M, Leech G, Weisser M, Wilson A 2000 Representation and annotation of dialogue. In Gibbon D, Mertins I, Moore R (eds) *Handbook of multimodal and spoken dialogue systems* Boston, Kluwer: 1-101.

Jurafsky D, Shriberg L, Biasca D 1997 Switchboard SWBD-DAMSL shallow-discourse-function annotation: coders manual. http://www.colorado.edu/ling/jurafsky/manual.august1.html.

Kita K, Yoshikanzu F, Masaki N, Tsuyoshi M 1996 Automatic acquisition of probabilistic dialogue models, *ICSLP-96,* Philadelphia pp196-199.

Klein M, Bernsen N O, Davies S, Dybkjaer L, Garrido J, Kasch H, Mengel A, Pirelli V, Poesio M, Quazza S, and Soria C 1998 *Supported coding schemes.* MATE Deliverable D1.1, LE Telematics Project LE4-8370.

Nagata M, Morimoto T 1994 First steps toward statistical modeling of dialogue to predict the speech act type of the next utterance *Speech Communication,* 15: 193-203.

Reitlinger N, Klesen M 1997 Dialogue act classification using language models. In *Proceedings of the 5th European Conference on speech communication and technology (Eurospeech '97),* Rhodes, Greece.

Sinclair J, Coulthard M 1975 *Towards an analysis of discourse: the English used by teachers and pupils.* London: Oxford University Press.

Stiles, W. B. 1992 *Describing talk: a taxonomy of verbal response modes.* Beverly Hills, Sage.

Weisser M forthcoming 2003 SPAACy – a semi-automated tool for annotating dialogue acts. *International Journal of Corpus Linguistics* 8(1).

Weisser M forthcoming Determining generic elements in dialogue.