Determining Generic Elements in Dialogue

Martin Weisser, Department of Computational Linguistics, Friedrich-Alexander-University Erlangen-Nuremberg, Germany

Abstract

This article reports on an attempt to identify generic elements in (mainly transactional) dialogue. By generic elements, I understand such elements that apply to a variety of different domains or may even be present in all types of spoken interaction. I will furthermore attempt to provide an appropriate classification and description framework for these elements and (where applicable) demonstrate strategies for determining them using corpus data.

Biography

Martin Weisser received his PhD in Phonetics/Corpus Linguistics from Lancaster University in 2001. He was recently involved in a project with Prof. G. Leech dealing with the creation of a speech-act annotated corpus for dialogue systems (SPAAC) in the Department of Linguistics and Modern English Language at Lancaster University. His main interests include all aspects of spoken language, including automatic processing and markup, general linguistics, and accents and dialects. He has recently co-authored (with G. Leech) the chapter on "Pragmatics and Dialogue" in the *Oxford Handbook of Computational Linguistics*, edited by R. Mitkov. Since January 2003, he has been working in the Department of Computational Linguistics at Erlangen.

Keywords: dialogue, generic elements

1. Background

The findings and ideas presented in this article are based on my recent work on the SPAAC (A SPeech Act Annotated Corpus for Dialogue Systems) project[1]. As the name indicates, the main aim of this project was to produce a corpus of pragmatically annotated dialogues that may be used as training data for dialogue systems, but 'secondary' aims were to develop a set of generic speech-act labels and increasingly also to determine other generic elements. In order to process a relatively large number of dialogues reliably and efficiently, our processing was done as far as possible automatically, a fact which has had a major impact on the ideas presented here.

The Data we were working with was deliberately varied: it consisted of Virgin Trainline bookings and timetable enquiries, more varied information seeking dia-

[1] I am greatly indebted to Geoffrey Leech for some invaluable comments on the first draft.

logues from a major telecommunications company and some spoken dialogues from the British National Corpus (BNC). One important feature to note is that the first two types of data are unpunctuated and we had no access to information about intonation.

2. Taxonomy of Generic Units

In order to provide the reader with a clearer idea what elements may be deemed generic in spoken interaction, I will first present a brief taxonomy of what I see as generic units, and which I will then discuss and expand on in the following sections. One of my basic assumptions is that it is possible to identify generic elements on the following four levels:

- structural units
- morphosyntactic (lexical) units (PoS)
- semantico-pragmatic units (*topics* & *modes*)
- speech acts

3. Segmentation Units in (Spoken) Language

The first question when trying to determine generic units in dialogue is to ask ourselves whether we can identify generic units at the level of structure. In the past, based primarily on a prevalent analysis of written texts and also ideas coming from a mainly Latin-based grammar model, spoken and written texts have often been analysed at the level of the sentence:

> "It is usually assumed that the SENTENCE is the highest-ranking unit of grammar, and hence that the purpose of a grammatical description of English is to define, […], what counts as a grammatical sentence in English. In this way, the terms 'grammar' and 'sentence' are mutually defining." (Quirk et. al., 1985: p. 47)

But:

> "The sentence is an indeterminate unit in the sense that it is often difficult to decide, particularly in spoken language, where one sentence starts and another begins […] To give a realistic account of English grammar, we therefore have to abandon neat boundaries, and to accept that there is a linguistic 'core' round which other aspects of linguistic organisation and usage are integrated." (ibid)

We therefore have to ask ourselves another question, namely whether traditional sentence types/categories are enough to adequately describe all aspects of general spoken (or written) interaction. To this, I would respond with a rather emphatic "No!", and for two main reasons.

a) They do not cater for short or syntactically incomplete utterances, such as: *yes (, please)*/*no (, thank you)* or *right/fine/well/aha*.

b) They are in most cases more concerned with describing form, rather than function.

So, which unit of analysis is then the correct and most useful one for annotating all types of language? The most appropriate unit, especially for annotating naturally occurring spoken language, should be deemed the *C-Unit*, described by Biber et. al., 1999 (p. 1070) as a unit comprising both "clausal and non-clausal units […] that […] cannot be *syntactically* integrated with the elements that precede or follow them.". However, one of the key issues here is not only their syntactic independence, but also that C-units form independent 'units of meaning', i.e. have a pragmatic function associated with them, even if they may just contain what could broadly be referred to as 'phatic' elements, such as interjections, certain types of discourse markers or conventionalised expressions, e.g. greetings or mutual introductions (cf. Stenström, 1994, p. 11). A further criterion – but one certainly related to the first one – is that a segmentation unit selected for largely automatic annotation and analysis has to be suitable for extracting sufficient information to determine not only its functional load, but also to relate it to other units around it, in order to identify higher-level discourse structures, such as *exchanges* or *transactions* (cf. Stenström, 1994, p. 30ff.). In other words, C-units should embody syntactically manageable 'free-standing' ideas or concepts. A tentative taxonomy of these concepts will be discussed in section below.

3.1. The SPAAC 'C-Unit' Type Taxonomy

In this section, I will present the C-unit type taxonomy used on the SPAAC project. The ordering within this taxonomy is at least partly determined by the order in which we perform our automatic analyses.

The first category is that of yes/no-like answers or statements, excluding backchannels. Here, we have to make a distinction between two different types:

a) those that signal an explicit acceptance or refusal, e.g. "Yes, please" or "No, thanks".

b) those that may also signal acceptance or refusal, but do so less explicitly, or, perhaps more frequently, simply express an acknowledgement or negation. In the latter case, it clearly depends on what they are in reponse to. If they are responding to a yes/no question, they may either acknowledge or negate the proposition expressed in the previous question. If they occur after a statement, they are most likely to simply signal that the information has been received, e.g. A: *There's the 7 33 from Birmingham New Street arriving at 9 05 in Euston. B: Yeah*.

Category number two encompasses all discourse markers. Here, again, we need to distinguish between two distinct types. Type a) includes such discourse markers as *aha*, *right*, *fine*, *ok*, etc., which tend to be rather similar to type b) yes/no-like c-units in that they are mainly used to express a form of acknowledgement and thus constitute responses. Type b), on the other hand, which comprises markers such as *well*, *now*, *so*, etc., fulfils a distinctly different function in that items of this category initiate or 'initialise' new turns and frequently also indicate a particular stage in topic management, e.g. ***Well**, em, is there a train from Liverpool?* or ***Now**, do you hold a current credit or debit card?*.

Categories one and two have another feature in common, which is that they tend to occur at the beginnings of turns if they have the functions described above.

The third category is that of questions, where we distinguish between Wh-questions and Yes/No-questions. These two C-unit types are the easiest to spot automatically because they may both contain inversions on the syntactic level or are also explicitly marked through the use of question words or auxiliaries.

Categories number four and five comprise declaratives and fragments (i.e. syntactically incomplete or unanalysable C-units). What both of these categories have in common is that we can sub-categorise them into 'pure', e.g. *you are able to get the next available train*, and 'subordinate', e.g. *if you miss the service i've reserved you on*, types. The reason why I have used the term *subordinate* in quotation marks here is that I believe it is misguided to suggest the kind of hierarchy implied by the using the Latin term – at least for spoken language. In other words, C-units of this type should be regarded as containing propositions pertaining to the same level of importance as the 'pure' ones because the have equal importance in influencing the 'flow' of the interaction. The fact that in most cases they begin with a conjunction, or possibly a prepositional phrase, merely reflects a cohesive function and not a hierarchical ordering. This is demonstrated by the fact that we can easily co-ordinate the two

examples given above in two different ways, so that the 'if-clause' either occurs before or after the pure declarative.

Finally, we also have the category of imperatives. This category is formally quite straightforward because it always begins with the base form of a verb, *don't* or *let's*.

4. From Global to Generic Lexicon – Motivation

The next level where we can identify generic elements is the lexical or, to be more precise, morphosyntactic one. My investigation into this area was originally guided by the need to include a lexicon in our analysis tool in order to improve segmentation. An initial pure pattern matching approach to segmenting utterances had proved to be relatively unsatisfactory once it came to determining the difference between declaratives and fragments, so that I started developing a minimal small-scale morphological lexicon, containing mainly function words, which increased the accuracy of the tool greatly. However, as we needed to analyse different types of more or less domain-specific data, i.e. switch between domains, and wanted to keep the processing overhead low, I began investigating the idea of producing a generic lexicon that can be augmented by domain-dependent data if and when necessary.

4.1. Main Ideas

In this section, I will describe and explain some of the main ideas behind my approach for developing a generic lexicon. Firstly, it should be relatively clear that there are some elements that always remain constant across domains, i.e. that there is a stock of essential vocabulary that is highly likely to occur in all domains. More details about the composition of this group of words will be given in the next section.

However, although words occurring in most domains may have the same orthographic shape, it is not enough to simply list these words in the lexicon because many of these words may also change their meaning in context. According to Thomas, 1995 (p. 4) "[…] there are occasions when we do quite genuinely experience difficulty in assigning contextual meaning and then we have to weigh up alternative interpretations. The likelihood of such problems occurring increases still further when there are rapid changes of topic […]"

This change in meaning, though, is often not only a change in function according to domain, e.g. *book* as 'reading material' vs. *book* as 'reserving a seat/ticket', but is often also associated with a change in word-class, something we could refer to as 'grammatical polysemy', e.g. the first *book* above is a N, and the second *book* a V.

However, it is also true that some meanings or usages tend to be more generic/prototypical than others and this is a feature that can be exploited in developing a generic lexicon.

4.2. Determination Strategies

So, how can we determine these prototypical meanings? Ideally through a mixture of theoretical linguistics and corpus-based methodology. The first step is to isolate 'pure' function words such as conjunctions, articles, pronouns, quantifiers, prepositions, question words, deictics, fillers and particles. Step two is to isolate other 'function words', i.e. auxiliaries and be-forms. Although these may also function as forms of full verbs, they are actually more likely to occur with auxiliary function. Next, we can determine high-frequency or 'everyday life' content words by making use of large-scale corpora, such as the BNC. However, such quantitative methods also ought to be supported by both empirical observation from materials under analysis and using intuition/linguistic knowledge, as pure frequency information may be skewed due to sampling methodologies involved in the creation of the corpus. The final step is to determine the most prototypical functions from tagged corpora by comparing frequencies of tag assignment and using intuition/linguistic knowledge.

4.3. Usage

Having determined which items may belong into our generic lexicon, how do we actually set it up in order to use it? As many words can have multiple part-of-speech (POS) categories, we can mark the default or most prototypical. This approach itself is certainly not a new one, as similar methodologies tend to be employed in probabilistic tagging systems, such as CLAWS (cf. Garside et al., 1987: p. 35), albeit in a somewhat different way. The way I suggest is by using a different type of tag-set that assigns only one prototypical tag per word, e.g. N for 'pure' nouns, n for words that tend to be nouns, etc. The distinction in the tagset between upper and lowercase tags not only signals the most likely word-class, but at the same time any lowercase tag indicates the potential for a different, though less likely, usage.

Once we have set up the generic lexicon in this way, we can then create domain-specific lexica and include not only domain-specific words, but also domain-specific PoS tags for each entry. During the data analysis phase, i.e. at run-time, the two lexica can then be combined, adding domain-specific words and overriding (or overwriting) generic PoS tags with domain-specific usage. Although this approach does not quite have the flexibility of specifying all potential PoS categories, as does a generalised PoS tagger such as CLAWS, it nevertheless helps to avoid many potential so-called

‘structural ambiguities’ before they even arise. In those cases where some ambiguity may still remain, it will then of course be necessary to look at the context in order to determine the real usage, provided that this is actually possible at all.

5. *Identifying Content*

When looking at the literature on dialogue systems, one often finds references to the terms *topic spotting* or *keyword spotting* when automatic means of identifying content are discussed (cf. Leech et al. 1998, p. 63). However, there are two problems inherent in these concepts. First of all, as many words only show their ‘true meaning’ in context, the idea of *keyword spotting* probably ought to be replaced by the notion of *key-phrase spotting* instead. Secondly, the notion of *topic spotting* also seems to be rather restricted in its focus on topics, rather than incorporating other elements that ‘drive’ a dialogue as well. I therefore propose a two-level model of content description, separated into:

a) domain specific content, i.e. content that is less likely to occur across different domains

b) generic content, i.e. content that reflects ‘everyday interaction’ or linguistic concepts

For labelling items in category a), I use the label *topics*, whereas for items of category b), I suggest the label *modes* since elements in this category tend to reflect the modus operandi of a given dialogue.

5.1. What Is ‘Generic Semantico-Pragmatic’ Content?

Generic semantico-pragmatic content, as I understand it, is content that can be related to, – or described in terms of – high-level categories of ‘aboutness’. It does, however, not necessarily describe the full pragmatic force of an utterance, but often only provides important clues towards its identification. So far, I distinguish between four different categories in my taxonomy. These categories represent four relatively distinct conceptual fields encompassing:

- grammatical concepts
- interactional concepts
- point-of-view concepts
- social concepts

I have used the term semantico-pragmatic to describe these categories as it is sometimes difficult to determine whether they ought to be considered semantic or pragmatic ones, although the first one seems to be more semantic, whereas the other three appear to be more pragmatic. I will discuss the properties of these individual categories in more detail in the following sections.

5.1.1. Grammatical concepts

Grammatical concepts are those that are associated more with function words, or combinations of these. These words/expressions, rather than having a high degree of content themselves, tend to be responsible for cohesive – in the broadest sense – aspects within the text (cf. Halliday/Hasan, 1976, especially chapter 5) or may express a form of modality. Grammatical concepts tend to be used in order to clarify the circumstances of the current situation by expressing choices, for example stating the existence of things, or conveying hypothetical information. Table 1 shows the expressions I have so far identified for this category, together with the labels used for them in the automatic analysis of dialogues.

label(s)	**expressions**
alternative	*either*, *or*
condition	*if*, *whether*, *unless*, *as long as*, *while*, etc.
constrain	*(al)though*, *but*, *only*, *have (got) to*, *must*, *need*, etc.
reason	*cos*, *because*, *that's why* …, etc.
exists	*there's*, *there are*, *is there*, etc.
poss1; poss2; poss3	*can*, *be able*, *might*, *may*, etc.
probability	*probably*, *probability*, *likely*, etc.

Table 1 – Grammatical Concepts

As far as expressions of modality are concerned, I have made the deliberate choice not to adopt commonly used labels such as *epistemic, deontic* and *dynamic* (Palmer, 1988 : p. 96ff.) as these already represent a form of interpretation which is not possible to achieve via an initial automatic analysis. There are actually three different labels in the possibility category (poss1, poss2 & poss3), depending on whether the C-unit is in the first, second or third person.

5.1.2. Interactional Concepts

Interactional concepts, on the other hand, are concepts that relate to the way that interlocutors interact with each other in order to keep the dialogue going. They include more phatic expressions, such as acknowledging backchannels, but also other expressions, such as of intent or offering, etc., that make it possible for the interlocu-

tors to negotiate turn-taking or to reach decisions. Table 2 below lists labels and potential expressions.

label	**expressions**
ackn	*m(h)m, etc.*
intent	*I'll just …, I'm (not) going to …, I'd like to …, etc.*
manage	*bear with me, hold the line, let me think, etc.*
offer	*I offer, etc.*
reassurance	*it's ok, that's fine, etc.*
report	*I'm told, I've been told, I'm trying to, etc.*
abandon	disfluency (no expression)

Table 2 – Interactional Concepts

Items under the label ***ackn*** refer to those elements in the dialogue that represent *backchannels* occurring in a separate turn, where the listener is simply acknowledging that they have heard or understood what the speaker has said and that they may agree with it. These act as important markers in the dialogue because they encourage the speaker to continue the subject for at least another turn.

Expressing an *intent* can be seen as interactional since it may signal to the interlocutor that the speaker is either going to do something that may interrupt the flow of the dialogue, e.g. *I'm booking from Euston to Birmingham New Street* or inform the interlocutor of a plan that is relevant to the present exchange, either as an indirect request or an issue that may be raised for discussion, e.g. *I'd like to leave on the 7 33 train*.

Under the label of *report*, we have to distinguish between two different types of reporting. The first one encompasses reports that we can describe under the general heading of reported speech, although the way it is currently being used in our analysis is only related to a caller giving the operator factual information – e.g. *sometimes I'm told that there's a fault* –, rather than to detect lengthier reporting passages, such as the ones we might expect to find in fiction. The second type is characterised by the occurrence of continuous forms in C-units with declarative content and reports mainly on ongoing attempts made by a caller, e.g. *I'm trying to get through to a number* or *I'm having problems*. Both of these types of reporting form characteristic parts of the interaction as they allow the participants to explain either what they are trying to achieve or refer to reactions or information they have had from third parties and which is relevant to the current interaction.

5.1.3. Point-of-view

The point-of-view category encompasses concepts that are usually discussed in the computational pragmatics literature under the headings of *belief* or *knowledge*, where it is often assumed that the beliefs of participants in a dialogue can be modelled and related to the participants' intentions (c.f. Bunt & Black, 2000: p. 14 or Jurafsky & Martin, 2000, p. 534ff.). To attribute belief or knowledge to one of the participants – or agent – in dialogue is, however, a serious issue, as there is often no way of verifying what an agent really believes or knows, although we can of course have certain assumptions about this, based on world knowledge and context. It is therefore far safer to restrict ourselves to annotating expressions that may signal a certain point-of-view, such as the categories listed in Table 3.

Label	**expressions**
awareness	*I (know/realise/understand) …, I'm aware …*, etc.
Opinion	*we think …, I suppose …, belief*, etc.
uncertainty	*I wonder (if) …*, etc.
Doubt	*I doubt …*, etc.

Table 3 – Point-of-view Concepts

The above three categories represent a kind of gradience from the speaker's signalling a conscious expression of realisation to a conscious expression of non-acceptance/disbelief. The key issues here are a) that we are not dealing with an inference about the beliefs of the speaker, but a direct expression of the latter and b) that we can assume that these beliefs are probably substantiated in the domain- or world knowledge of the speaker.

5.1.4. Social Concepts

The final category is that of social concepts, i.e. those concepts that signal a form of social interaction between two agents, summarised in Table 4.

Label	**expressions**
apology	*apolog(ise/y)*
appreciate	*no problem, that would be (brilliant/correct/fine/great/lovely/wonderful)*
expletive	*oh shit, damn*, etc.
Greet	*hi, hello, good afternoon, bye*
Insult	*you (bastard/idiot), (damn/blast) you*
Regret	*I'm (very) sorry, we regret*, etc

Table 4 – Social Concepts

In this category, we find the most common formulaic expressions, such as greetings, but also other expressions that signal interaction on a more personal, rather than structural level. For example, all the items listed under the label *appreciate* signal a form of acceptance on the part of the speaker, although the degree of commitment expressed by them may vary.

5.1.5. Residual questions

Of course, the taxonomy represented above should not necessarily be seen as complete yet, but rather represents a first attempt at establishing a generic framework for describing many different types of dialogues or texts, and there remain a few residual questions to be answered before this framework can be deemed anywhere near being complete.

The first of these questions no doubt ought to be whether deictic references belong into the realm of topics or that of modes. At first glance, deixis seems to be one of the most common features of any kind of interaction and therefore definitely qualifies for being incorporated into the category of modes, but on the other hand, not all types of texts include the same types of deictic reference, such as references to time or place.

The second one follows on from the first one. If there may be any doubt as to whether certain concepts belong more into the mode or the topic category, where should we set the boundary between topics and modes? Alternatively, we could of course ask ourselves whether there may be a gradient from topic to mode or maybe even an intermediate category.

The final question is one that really only becomes relevant if we want to aim at producing an analysis methodology that is maximally flexible, instead of being content with being able to handle a limited set of domains only, and is whether we can produce exhaustive sets of modes and topics to satisfy all possible genres and text-types?

5.2. What is 'Generic' Pragmatic Content?

Generic pragmatic content is such content that is, similar to *modes* on the semantic level, highly likely to recur in almost all types of dialogue. Although this notion seems to be a relatively straightforward one, as yet no generic set of speech-act labels exists which could properly be applied to a large variety of different domains. Most annotation schemes developed for use in dialogue systems still use very domain-specific tagsets, with probably the most generic scheme so far being DAMSL (Dialog Act Markup in Several Layers; Allen & Core, 1997) and its adaptation for use in the

Switchboard corpus (Jurafsky et al., 1997). However, both of these schemes still have some drawbacks with regards to the units they annotate and the nature of the tags used, although the Switchboard version provides some improvements. In DAMSL the underlying theoretical assumption is that the beliefs of the speaker or hearer can be modelled, so that e.g. the *Assert* tag is supposed to be used for coding "[…] utterances that make an explicit claim about the world." and "The key distinction for the Assert tag is that the speaker is trying to change the beliefs of the hearer." (Allen & Core, 1997, p. 10), but c.f. my earlier criticism of trying to ascribe belief in section 5.1.3 above. Other tags, such as the *Open-option,* seem to presuppose that we can always make a clear distinction between giving information and suggesting an option.

5.2.1. The SPAAC Scheme

In this section, I will discuss the generic scheme for representing speech acts developed during the (still ongoing) SPAAC project. This scheme represents an attempt to expand and clarify on the descriptions and recommendations made in Leech et al. 1998 (pp. 54-67).

The idea of generic pragmatic elements is based on a set of basic assumptions, the first being that there exists a set of high-level speech-act/interactional categories that can be employed to characterise the function of any C-unit in spoken – and perhaps also written – interaction. The second is that previous annotation schemes have often conflated too many C-units into single moves because their emphasis was on identifying the higher-level moves, rather than the more fine-grained acts contained in utterances, so that by increasing the granularity, we can produce a more accurate description. Assumption number three is that, given a particular C-unit, the range of possible speech-acts for this unit is limited by its type. Finally, assumption number four is that, given a particular C-unit type and sufficient information about its mode and topic attributes (and perhaps others, such as polarity), we can 'combine' all this information to deduce the speech-act expressed by it.

5.2.2. Analysis Steps

Assuming that our assumptions described above are correct, we can map out a basic strategy for determining the functions of c-units in a dialogue.

Step one is to determine C-unit types via syntactic analysis and to assign some default assumptions to each C-unit. We can automatically – and relatively accurately – determine the type of a C-unit by using an elimination procedure which first splits off and annotates initial yes/no like responses and discourse markers, then identifies wh-

and yes/no questions and finally imperatives, declaratives and fragments. As we are annotating the C-units syntactically, we can also fill in such default assumptions as e.g. questions tend to be requests for information (label: req-info), imperatives tend to be directives (label: direct) – unless they start with *let's* or *let me*, in which case they tend to be suggestions (label: suggest) –, etc.

We can then proceed to collect mode and topic information. This can be achieved by matching each C-unit against keyphrase 'thesauri' and counting the number of occurrence of each category, so that they can be ranked and written into the topic/mode attribute in the order of importance, assuming that the more frequently a category is represented, the more important it is for describing the content of the C-unit.

The next step is to find answers to questions and assign them an answer speech act attribute (label: answ). We can also assign or append the mode attribute closure because any answer effectively closes a question-answer exchange.

The final step then is to determine, or override previous, speech-acts. We can do this by first checking against mode attributes or a combination of mode + topic attributes, as the mode attributes tend to be the most reliable indicators of the function of the c-unit. If the mode attributes provide no clues, then, as a last resort, we can attempt to identify the function by using topic attributes only.

6. The Current Spaac Taxonomy of Generic Speech-Acts

Our current taxonomy of speech-acts aims at a level of generality that previous schemes have not been able to attain. Often, we can achieve this by being able to refer to some of the other attributes of a C-unit, such as polarity, etc. or by avoiding explicit labels signalling some form of commitment as in DAMSL. We also distinguish between high-level labels that are often automatically applied as default labels by our annotation routines, but may over-generalise, and low-level labels which tend to be more specific ones and may replace the default labels during the manual post-editing phase. Table 5 contains a list of our labels, each with a corresponding brief explanation.

label	**explanation**
accept	strong form of acknowledgement, not just confirming something the previous speaker has stated, but actually agreeing with it
ackn	weak form of acknowledgement, simply confirming or rejecting a previous proposition, depending on the polarity of the utterance

label	**explanation**
answ	answer to a request for information
answ-elab	elaboration to an answer, typically occurring in sequences such as Question: *do you have a credit or debit card* Answer: *yes* Elaboration: *a credit card*
bye	farewell at the end of a dialogue
complete	one speaker collaboratively completes an utterance of the previous speaker
confirm	speaker A summarises or presents information that has been discussed before, at the same time giving speaker B the chance to acknowledge or correct this information; usually occurs in something akin to a declarative question
correct	speaker A corrects something speaker B has said
correct-self	speaker A corrects him/herself after having been corrected by speaker B
direct	giving an instruction or making a request
echo	repeating all or part of the previous speaker's utterance
express-opinion	an explicit expression of an opinion, i.e. via such sequences as *i think, i believe*
express-possibility	any C-unit where a modal auxiliary can follows a non-1st-person subject and that cannot be interpreted as either a ***suggest***, ***direct*** or ***offer***
express-regret	any C-unit, apart from discourse markers, that contains words/phrases like *(I'm) sorry*
express-wish	a fallback category for a declarative that contains an expression of volition, but cannot be interpreted as a ***direct***
greet	greeting
hold	speaker A tells speaker B to 'hold the line' in order to signal that there may be some delay before the conversation can continue
identify-self	a formulaic introduction, like *this is X* or *X speaking*
inform	providing basic information; default label for declaratives or fragments if they do not fit into any other categories
inform-intent	an expression of intent in declarative form
inform-intent-hold	a special form of ***hold***, where speaker A also conveys information about what he/she intends to do, thus explaining what necessitates the ***hold***
init	initialising/initiating a turn (or utterance); more or less only applicable to discourse (turn-initial) markers; signalling a new (sub-)topic
negate	high-level label for a 'no-expression', either applicable if the proposition of the preceding utterance is negated or a *no* cannot be interpreted as either a ***refuse*** or ***answ***
offer	offering a service/favour
pardon	expressing non-understanding of the previous speaker's utterance, thus eliciting a repetition/clarification

label	explanation
refer	high-level label for a referential expression that does not constitute a full declarative, e.g. a stranded prepositional phrase like *on Monday*
refuse	refusing an offer
req-direct	request for a directive or instruction
req-info	request for information
req-modal	high-level label for a 'questioning' C-unit that contains an expression of modal possibility and which cannot be categorised according to any of the other categories
self-talk	the speaker talks to him/herself and does not actually contribute to the task at hand
suggest	making an explicit suggestion
thank	thanking someone
uninterpretable	applies to any C-unit that cannot be properly interpreted; usually an abandoned/disfluent C-unit

Table 5 – the SPAAC speech-act taxonomy

7. Conclusion

The methodology described above provides a framework for determining and identifying generic elements in interaction on various levels. It has so far proved very successful when applied to the automatic analysis of data from two different (albeit not radically so) domains. Our results for identifying speech acts for these two domains show a recall of up to 90-odd percent. We have yet to ascertain the real degree of precision through manual post-editing and feedback upon the categories from our post-editors, but cursory examination for at least the first of our domains seems to indicate that precision may well be in the 90+% range, and that it would simply be a question of improving and adapting both the mode and topic thesauri and the rules that determine the speech acts during the final analysis stage, plus adding domain-specific lexica in order to make the methodology usable in a large variety of different domains.

8. References

Allen, James and Mark Core. 1997. Draft of DAMSL: Dialog Act Markup in Several Layers. Version October 18, 1997. available in PostScript from: http://www.cs.rochester.edu/research/cisd/resources/damsl/ and in HTML format http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/RevisedManual.html [Allen & Core, 1997].

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan and Ed Finegan. 1999. Longman Grammar of Spoken and Written English. London: Longman. [Biber et. al., 1999]

Bunt, Harry and William Black. 2000. Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics. Amsterdam: John Benjamins. [Bunt & Black, 2000]

Garside, Roger, Leech, Geoffrey and Geoffrey Sampson. 1987. The Computational Analysis of English: a Corpus-Based Approach. London: Longman. [Garside et al., 1987]

Halliday, M.A.K. and Ruqaiya Hasan. 1976. Cohesion in English. London: Longman. [Halliday/Hasan, 1976]

Jurafsky, David and James Martin. 2000. Speech and Language Processing: an Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Upper Saddle River, New Jersey: Prentice Hall. [Jurafsky & Martin, 2000]

Jurafsky, David, Shriberg, Liz and Debra Biasca. 1997. “Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coder Manual”. available at: http:/stripe.colorado.edu/~jurafsky/manual.august1.html [Jurafsky et al., 1997]

Leech, G.eoffrey, Rayson, Paul and Andrew Wilson. 2001. Word Frequencies in Written and Spoken English. Harlow: Longman. [Leech et al. 2001]

Leech, Geoffrey, Weisser, Martin, Wilson, Andrew and Martine Grice. 1998. “Survey and Guidelines for the Representation and Annotation of Dialogue” in: Gibbon/Mertins/Moore (eds.). 2000. Handbook of Multimodal and Spoken Dialogue Systems. Dordrecht: Kluwer Academic Publishers. [Leech et al. 1998]

Palmer, F.R. 21988. The English Verb. London: Longman. [Palmer, 1988]

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey and Jan Svartvik. 1985. A Comprehensive Grammar of the English Language. London, Longman. [Quirk et. al., 1985]

Stenstöm, Anna-Brita. 1994. An Introduction to Spoken Interaction. London: Longman. [Stenström, 1994]

Thomas, Jenny. 1995. Meaning in Interaction: an Introduction toPragmatics. London: Longman. [Thomas, 1995]