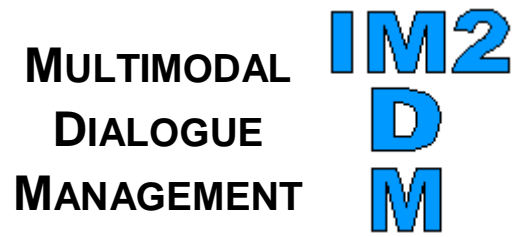




<http://www.im2.ch>



<http://www.issco.unige.ch/projects/im2/mdm/>

ISSCO/TIM/ETI, University of Geneva

Dialogue act tagsets for meeting understanding:
an abstraction based on the DAMSL,
Switchboard and ICSI-MR tagsets

Andrei Popescu-Belis

IM2.MDM Report 09 – September 2003

Version 1.2 (December 2004)

Dialogue act tagsets for meeting understanding: an abstraction based on the DAMSL, Switchboard and ICSI-MR tagsets

Plan of the report *

Chapter 1. Theoretical considerations on dialog acts

- 1.1. Motivations for defining a set of dialog acts
- 1.2. Useful terminology: utterances, sentence-types, and functions
- 1.3. Functions of utterances in several dimensions
 - 1.3.1. Speech act dimension
 - 1.3.2. Turn management dimension
 - 1.3.3. Dimension of adjacency pairs
 - 1.3.4. Dimension of overall organization
 - 1.3.5. Other dimensions

Chapter 2. Critical overview of DAMSL, SWBD, ICSI-MR

- 2.1. DAMSL
- 2.2. Discussion of DAMSL
- 2.2. Switchboard-DAMSL (SWBD)
- 2.3. Formal analysis and comments on SWBD
- 2.4. Final SWBD tagset based on empirical analysis

Chapter 3. Analysis of ICSI-MR annotation guidelines

- 3.1. Overview of ICSI-MR guidelines
- 3.2. Rule-induction for ICSI-MR labels
- 3.3. Criticism of the tagset
- 3.4 Empirical analysis of six ICSI dialogues: some facts and inaccuracies

Chapter 4. An abstraction of ICSI-MR dialog act tagset

- 4.1. Proposal: MALTUS - BIS
- 4.2. Empirical study: conversion of ICSI-MR labels to MALTUS
- 4.3. Correspondence with ICSI-MR, DAMSL and SWBD
 - 4.3.1 ICSI dialogue acts and their mapping to the IM2.MDM abstraction
 - 4.3.2 Mapping with DAMSL, SWBD and other tagsets

Chapter 5. Towards a multidimensional dialog act tagset based on pragmatic theories: PRIMULA

* The author would like to thank Sandrine Zufferey and Alexander Clark (ETI, University of Geneva) for valuable discussions and insights, as well as Liz Shriberg and Barbara Peskin from ICSI for their help with the ICSI-MRDA corpus

Chapter 1. Theoretical considerations on dialog acts

1.1. Motivations for defining a set of dialog acts

There is not much agreement, within the NLP community, on the definition of a *dialog act*. It seems that the term is used to denote some “function” of an utterance in a dialog. One of the main inspiration sources are “speech acts”, which are a well-defined though controversial concept in pragmatics. But the original list of speech acts (as formulated for instance by Searle, Vanderveken, etc.) has been considerably enriched with other “functions”, depending on the goals of those who set up their list of possible dialog acts.

In this report, our goal is to abstract from existing sets of dialog acts, based on the following considerations or constraints:

1. **Theory:** the DA set should be somehow related to a theory (descriptive, explanatory, etc.) of the phenomena or “functions” it pretends to grasp (annotate).
2. **Empirical validation:** the DA set should be reliably tagged by human annotators – inter-annotator agreement, e.g. measured by the kappa coefficient, should be reasonably high (at least potentially).
3. **Insights from the data:** apart from theoretical considerations, the DA set should be motivated also by looking at the functions of actual utterances, in a given domain.
4. **Possibility of detection:** the DA set should not be too remote from the present capabilities for automatic detection of the DAs (this is of course a requirement for NLP applications).
5. **Role of the application:** since there are numerous possible “functions” of utterances, the DA set should be designed depending on the targeted NLP application. The tagset should mark the “functions” that are important instead of trying to mark all “functions”.
6. **Mapping to existing DA sets:** the DA set should be shown to be reasonably compatible with previous proposals (or at least compared to them) so that useful insights are preserved, and data can be reused.

The general goal behind DA sets is therefore to reliably extract some useful information from dialogs, where the information is not at the level of syntax or semantics, but at a higher level, which is related to the dialog structure and to the intentions of the speakers, falling thus broadly under the scope of *pragmatics*. So, our first question is: what kind of “useful information” has already been theorized within the field of pragmatics? Some suggestions follow—for constraint (1), keeping in mind constraints (2), (3) and (4).

One of the general observations that can be made is that analyses of dialog are often located on different, and somehow independent, planes. For instance, Schiffrin (1987) posits an exchange structure, an action structure, an ideational structure, a participation framework, and an information state. In Levinson (1983), one finds an analysis in terms of speech acts (which is quite criticized by Levinson), followed by findings from conversation analysis and related fields regarding turn-taking, adjacency pairs, openings, etc. We will follow these very general distinctions in what follows.

1.2. Useful terminology: utterances, sentence-types, and functions

Some terminology can be useful at this level. The *utterance* of a *sentence* (or sub-sentential fragment) occurs in a given context, and bears a certain “function” (or several functions). Utterances are also signalled by syntactic and/or prosodic means. Utterances seem to be the atomic subparts of a turn that bear one or more “functions”. A *turn* is the interval when a speaker is active, between two pauses in his/hers flow of speech, occasioned by the intervention of another speaker.

An important distinction must be made from the start between the *sentence-type* and the *function* of an utterance, especially in a speech-act framework. “The three major sentence-type in English are the imperative, interrogative and declarative” (Levinson 1983). These appear even to be cross-language universals (Sadock & Zwicky 1985). Sentence-types are somewhat related to the *mood* of the main verb phrase, i.e. indicative, imperative, subjunctive, conditional, but also to the syntax and the other markers of the sentence. So, one of the problems of dialog acts is to relate sentence-types to functions of the utterances.

Now, how are the “functions” specified? Surely, this depends on the theory that is used and on the aspect of dialog that is analyzed. However, a certain number of functions appear naturally, and therefore quite frequently, in the study of conversation: question, order, greeting, etc. Many of these terms receive a formal definition within the speech act theory (this being one of its strengths) even if such definitions in terms of felicity conditions have often been criticized. But does the use of these terms entail an implicit acceptance of speech act theory? Levinson (1983: chap. 6.5) is worth quoting here:

“... the terms *request*, *invitation*, *greeting*, and so on, are not the inventions of speech act theory, but rather part of a rich (if largely unexplored) natural language metalanguage [...]. It does not follow from the existence of such terms [...] that such categories are properly explicated by providing sets of necessary and sufficient conditions for speech act category membership [...]. ...categories like *request* should in fact be backed up by at least (i) a full sequential explication in terms of the range of expectable responses (like refusals, deferrals, compliances, etc.), (ii) an account of the way that requests are typically formulated in order to obtain the desired responses”.

So it seems we have to live for the moment with some kind of ambiguity regarding the various possible “functions”. Many analyses point out at the variability of the functions of, say, “questions”, that is, utterances whose sentence-type is a question. Their functions depend strongly on the activity in which they are used. Levinson (1979, 1992) shows for instance that depending on the setting (which he calls the *speech event*), there is huge variability in the roles of questions. Questions fulfil quite different functions in exams, in courts, in greeting sequences, in openings, etc. But these roles are explained by Levinson only using verbal explanations in each case—without defining formally a systematic repertoire of roles. However, in each of the planes or dimensions described below, there are more specific definitions of the *functions* of utterances.

1.3. Functions of utterances in several dimensions

An analysis of the various fields of pragmatics suggests that theoretical progress has been made in several dimensions, such as the role of utterances in conveying *speech acts*; their role in conversational *turn-taking*; their function as members of *adjacency pairs*; their role in the general organization of conversation (in openings, closings, topic shifts). Of course, as mentioned above, there is almost infinite variation in what the “function” of an utterance can be; semantic aspects, or emotional ones, could also count as functions. In the remaining of this chapter, we will review the main dimensions we just mentioned. But insights from other dialog act sets, or from the application DAs are used for, could increase the number of dimensions.

1.3.1. Speech act dimension

According to Austin and Searle (and many others), there are a number of actions that are performed by speaking. The function of an utterance according to this dimension is to perform one of these actions (ideally one, but maybe more). One of the advantages of this theory is the formal definitions it provides for these actions—of course, the definitions were criticized for being too restrictive. Here is a list of acts, or types of *illocutionary force*, grouped in five categories, with their paradigm cases (Searle 1976 via Levinson 1983):

1. **representatives** (commit the speaker to the truth of the expressed proposition): assertion, conclusion, etc.
2. **directives** (attempts by the speaker to get the addressee to do something): request, question, suggestion, etc.
3. **commissives** (commit the speaker to some future course of action): promise, threat, offer, etc.
4. **expressives** (express a psychological state): thanks, apologize, welcome, congratulation, etc.
5. **declarations** (effect immediate changes in the institutional state of affairs; tend to rely on elaborate extra-linguistic institutions): excommunication, declaration of war, christening, firing from employment, etc.

One of the main difficulties in associating the right speech act to an utterance is the problem of *indirect force*. Utterances, apart from being of a certain sentence-type, have also a *literal illocutionary force* (one of the speech acts above), as well as, quite often, an *indirect force* that seems to represent the “real” or “dominant” function of the utterance—much better in this case than the literal force. For instance, when uttered at table, “Can you pass the salt?” is in the form of a question (sentence-type), its literal force is a question, but its indirect force is a request to pass the salt, and it obviously prevails over the literal force. That is, an answer such as “Yes, I can” followed by no action would obviously count as a wrong reaction. Note however that the literal force is not totally absent, since the reply can be: “Yes, of course I can, here it is”. Many solutions have been proposed to the problem of indirect speech acts (ISA, see Levinson 1983: chap 5.5), i.e. to the problem of form-to-force correlation, but none of them seems to be satisfactory (Lycan 2000:199-202).

But the problem if ISAs is merely synonymous to the difficulty to *find* the proper force (and in particular to explain how speakers manage to find it), not to the fact that there *is* some kind of

“dominant speech act” related to an utterance. To simplify, we have mentioned three aspects for an utterance:

- the form of the utterance or its sentence-type
- the literal force, quite closely related to the sentence-type
- the indirect force (possibly, not always) which can be somehow inferred from the utterance and the context.

What really matters is of course the “correct” or “dominant” force, that is either the literal one (if no indirect force is detected), or the indirect one. Of course, this doesn’t prevent the existence of secondary forces. This dominant force but must be somehow computed. So, the problem of ISAs is not so much a problem about a repertoire of speech acts, but a problem about associating a label (or more) from the speech act repertoire to an utterance.

Here are some more cues to determining the right force in some simple cases (from Levinson 1983:263-264). Literal force could be computed in this way: “(i) explicit performatives [i.e. with an explicit performative verb] have the force named by the performative verb in the matrix clause [APB: even if, in English, they are in declarative format]; (ii) otherwise, the three major sentence-types in English, namely the imperative, interrogative and declarative, have the forces traditionally associated with them, namely ordering (or requesting), questioning and stating respectively (with the exception of explicit performatives in (i)).” The proper indirect force must be inferred from all elements available.

1.3.2. Turn management dimension

Findings from conversation analysis (especially work by Schegloff) show that some utterances, or fragments of utterances, serve to manage the complex mechanisms of turn-taking, along with expressive devices embedded in each utterance. Of course, it might therefore be said that each utterance embeds turn-taking control marks, but some utterances function exclusively as tools for turn-taking control, like for instance *backchannels*. Following insights from existent DA sets, one could distinguish four functions: *backchannel*, *floor holder*, *floor grabber*, *hold*. The definitions will be discussed below, but note that a speaker can easily “grab the floor” without a specific “floor grabber”, just by asking a question in a loud voice, for instance. So, given that virtually all utterances also serve turn-taking purposes, the idea would be to mark in this dimension of turn-management those utterances that serve *exclusively* these purposes. This provides backchannels with a “function”, since otherwise they seem semantically empty. Also, being able to detect such a function prevents an attempt for semantic interpretation of those utterances that serve only for turn-taking.

There seem to be three types of “exclusively turn management” functions: backchannel (let the other speaker continue), floor-holder (let the speaker continue) and floor-grabber (stop other speaker and let utterer talk). Logically, a fourth type should be present, a signal for the other speaker to take his/turn (equivalent of “I stop and you speak”), but it seems a stop (silence) is the device *par excellence* to do this.

1.3.3. Dimension of adjacency pairs

Adjacency pairs (APs) are fundamental units of conversational organization. They are “sequences of two utterances that are: (i) adjacent; (ii) produced by different speakers; (iii) ordered as first part and second part; (iv) typed, so that a particular first part requires a particular

second (or range of second parts)” (Levinson 1983:303). These defining constraints should probably be relaxed a bit, in particular to allow insertions between the first and the second part.

This structure provides us with an essential distinction, between utterances that are *first part* (1) vs. *second part* (2) of an AP. This distinction is not *a priori* linked to the dimensions mentioned above, though there are speech acts that function preferably as (1) or (2). The question arises whether an utterance can play both roles, i.e. be a second part to a previous utterance, and a first part to a further utterance. Comparing this distinction with the independent forward-function and backward-function from the DAMSL tagset (see below), the answer seems to be affirmative: some utterances could serve both as (1) and (2). But on the other hand, it is likely that such an utterance would be made of two subparts, each with a different function. In situations such as Q1 → Q2 → R2 → R1 (imbricated APs), we wouldn't say that Q2 is an answer or a second part to Q1, but rather an “unexpected second”, which does not cancel the expectation of an answer to Q1 (finally R1). So we could hypothesize that an utterance is either a first part or a second part (see also the empirical evidence from SWBD). A supplementary piece of information in this dimension would be to link effectively first and second parts, and assign a label to the link (is in the ICSI AP task).

Refinement of adjacency pair information is based on commonly observed types of first/second pairs. Here is a list from (Levinson 1983:336)—the second part is written as “preferred / dispreferred second”:

- request → accept / refuse
- offer → accept / refuse
- invite → accept / refuse
- assess → agree / disagree
- question → expected answer / unexpected answer
- blame → denial / admission

We may also add less structured set:

- apology → downplay (minimization)
- thank → welcome
- greeting → greeting

Are adjacency pairs related to dialog grammars or sequencing rules? Such normative structures for dialogues, often proposed by discourse analysts (e.g., Sinclair and Coulthard 1975; Geneva school), were criticized by Levinson (1983:288-294) and others. The idea is first to associate speech acts to utterances (but the algorithm is seldom specified by these grammars, and the speech act framework seems too restrictive to be used here), then, second, to find sequencing rules that constrain the utterances in a dialog (but it seems there are no such constraints, only preferences). Such models are thus very far from providing a full theory of dialog. But this does not mean that they cannot contribute with something to our understanding.

According to Levinson (1983:293-4), “... sequencing constraints in conversation could in any case never be captured fully in speech act terms. What makes some utterance after a question constitute an answer is not only the nature of the utterance itself but also the fact that it occurs after a question with a particular content – ‘answerhood’ is a complex property composed of

sequential location and topical coherence across two utterances, amongst other things; significantly, there is no proposed illocutionary force of answering. The [dialog grammar models] skirts the puzzling issue of constraints on topical coherence...”

1.3.4. Dimension of overall organization

The structure of a conversation depends on a previously agreed plan (as registered in social norms for instance). General conversations, such as telephone calls, do not seem to have more structure than: < opening + first_topic (+ next_topics)* + closing >. Openings and closing have been studied a lot in conversation analysis. This generic structure could lead to the following functions for an utterance:

- is_part_of_opening
- is_part_of_closing
- is_part_of_topic (_development)

The last one could be further subdivided as:

- topic_changer
- topic_continuer (default)

These functions are mutually exclusive. It does not seem useful to distinguish the first topic from the others. In fact, it is not obvious to individuate the episodes and topics of a conversation. In more structured interactions such as meeting recordings, if an agenda has been agreed upon for the meeting, then utterances can also be classified based on it.

1.3.5. Other dimensions

Among other functional dimensions, we could first mention the emotional dimension (does the utterance have highly emotional content or not), or speaker involvement. Politeness is another dimension, and could be formalized using the face-saving / face-threatening model developed by Brown and Levinson. This would generate another set of tags. Another set could capture the humoristic function of utterances. Rhetorical relations (*à la* Mann & Thompson) could provide another dimension. In fact, a close look to existing sets of dialog acts could provide more ideas for new dimensions. Some properties of utterances could serve as functional dimensions, but are best conceived as phonetic / syntactic / semantic properties; one of them is intonation.

Chapter 2. Critical overview of DAMSL, SWBD, ICSI-MR

We attempt now to analyze three related set of dialog acts. Our analysis uses the annotation guidelines and additional publications related to these sets, and is based in part on the previous framework (Chapter 1). We have not found yet an extensive theoretical grounding of the three tag sets: the notion of dialog act, in particular, is seldom discussed. The three tag sets are (with their references):

- **DAMSL** – James F. Allen & Mark G. Core (1997), DAMSL: Dialog Act Markup in Several Layers (Draft 2.1) Multiparty Discourse Group, Discourse Research Initiative, Sept./Oct., 1997: <ftp://ftp.cs.rochester.edu/pub/packages/dialog-annotation/manual.ps.gz>, <http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/RevisedManual.html>.
- **Switchboard-DAMSL (SWBD)** – Daniel Jurafsky, Elizabeth Shriberg & Debra Biasca (1997), Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation (Coders Manual, Draft 13), Technical Report University of Colorado, Institute of Cognitive Science, 97-02: <http://www.colorado.edu/linguistics/faculty/jurafsky/pubs.html#Tech>, <http://www.colorado.edu/ling/faculty/jurafsky/manual.august1.html>; Jurafsky, D., Shriberg, E., Fox, B., and Curl, T. (1998), “Lexical, prosodic, and syntactic cues for dialog acts”, *ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers*, Montréal, p.114-120; Daniel Jurafsky (in press), “Pragmatics and Computational Linguistics”, in *Handbook of Pragmatics*, ed. Laurence Horn and Gregory Ward, Blackwell, Oxford, UK.
- **ICSI-MR** – ICSI Meeting Recorder Project: Sonali Bhagat, Rajdip Dhillon, Hannah Carvey, (Ashley Krupski) & Elizabeth Shriberg (2003), Labeling Guide for Dialogue Act Tags in the Meeting Recorder Meetings, Report ICSI (International Computer Science Institute), Berkeley, August 2003: <http://www.icsi.berkeley.edu/Speech/mr/docs/Draft7.pdf> (previous version August 2002). See also the website.
- The interested reader should also consult: David R. Traum (2000), 20 Questions for Dialogue Act Taxonomies, *Journal of Semantics*, **17**:1, p.7-30: <http://www.ict.usc.edu/~traum/Papers/amstel.pdf>.

One of the key aspects that we are interested in is the *multidimensionality* of the tags, or, conversely, their *mutual independence*. In some documents, dialog act tags (i.e., types) are simply compared on a one-to-one basis. However, the rules that govern their organization are essential, and we will focus on them below. The number of tags, the number of possible combinations, and the number of classes of tags are also related issues. All these are motivated by the desire to better understand the possible range of functions for an utterance, and consequently *to reduce the search space* for a system that attempts to detect the “functions” of utterances. Also, we try to give a theoretical, coherent grounding to the tagset. And we need to relate the tag set to an application, which is here the IM2.MDM meeting analysis and retrieval system (more about it in section 3).

2.1. DAMSL

In the DAMSL tagset, “the notion of utterance [is based] on an analysis of the intentions of the speaker (the speech act). [...] The utterance tags all indicate some particular aspect of the

utterance unit itself, summarizing the intention of the speaker [...] and the content of the utterance”. Theoretical justifications from the DAMSL guidelines end here. We learn from Jurafsky (in press) that the tagset draws on Allwood’s work (1992, 1995), but this work seems to be far less specific than the DAMSL tagset itself. The task that was proposed to the speakers recorded in the TRAINS corpus, to which DAMSL is related, is collaborative planning in the domain of merchandise transportation.

Each utterance can be annotated with zero, one or more labels, from four dimensions. The authors acknowledge (correctly, to my mind) that “generally, the dimensions are orthogonal, and you can find examples of any possible combination of labels. We will explicitly point out a few places where this does not seem to hold”. There are not many such places, unfortunately, in the DAMSL guidelines: section 3 gathers on one page about half a dozen examples. Probably, this was thought to be a matter of empirical study, as we will see with Switchboard.

The four **dimensions**, *sub-dimensions*, and possible dialog acts are summarized below. Dimensions are in bold, sub-dimensions in italics, and the dialog acts themselves are underlined. We also tried to figure out from the guidelines which sets of dialog acts are mutually exclusive (we marked this as “A *xor* B”, for exclusive or) and which are not (we marked this “A and/or B”).

1. **Communicative Status:** uninterpretable and/or abandoned and/or self-talk

Note: “the features are independent of each other... [and] most utterance units have no features marked [on this dimension]”

2. **Information Level:** task *xor* task-management *xor* communication-management *xor* other-level

Note: this level “provides an abstract characterization of the content of the utterance”. Rather, to my mind, it reflects a distinction between the task and the conversation itself, given that there *is* a task in the TRAINS exercise, unlike in the Switchboard conversations. Only one of these dialog acts can apply, hence the ‘xor’ (exclusive or).

3. **Forward Looking Function (FLF)**

This dimension codes the “effect an utterance has on the subsequent... interaction”. Eight different aspects can be coded. The guidelines state that constraints could be set for domain-specific annotation, but no example is given in the guidelines. The instruction given for annotation of this dimension is: “as a default,... code all aspects that are applicable”.

Forward Looking Function: *statement* and/or *influencing-addressee-future-action* and/or *info-request* and/or *committing-speaker-future-action* and/or *conventional* and/or *explicit-performative* and/or *exclamation* and/or *other-ff*

- *statement* (a claim made by the speaker): assert *xor* reassert *xor* other *xor* nothing
- *influencing-addressee-future-action* (Searle’s “directives”): open-option(weak suggestion or listing of options) *xor* action-directive (actual command) *xor* nothing
- *info-request* (a question by the speaker): yes *xor* no (sometimes a check question has been added)

- *committing-speaker-future-action* (Austin’s commissives): offer (speaker offers to do something, subject to confirmation) xor commit (speaker is committed to doing something) xor nothing

As the authors of DAMSL note, “the remaining functions are relatively rare”.

- *conventional*: opening xor closing xor nothing
- *explicit-performative* (examples: thanks, apologize, etc.) : yes xor no
- *exclamation*: yes xor no
- *other-ff*: yes xor no

4. Backward Looking Function (BLF)

This dimension “indicates how an utterance relates to the previous discourse”. There was even a project within DAMSL to link utterances that have a BLF to the utterance that is their first part (of the adjacency pair) using a ‘Response-To’ label, but it seems that this was never implemented. There are three sub-dimensions for BLF, which are said to be “semi-independent” (not clear what this means). A plan for a fourth sub-dimension (Information-relation) was not implemented. Note that the answer may contain a pointer to the index of the utterance which it answers.

Backward Looking Function: *agreement* and/or *understanding* and/or *answer*

- *agreement* (speaker’s response to previous proposal): accept xor accept-part xor reject-part xor reject xor hold (putting off response, usually via a sub-dialogue) xor nothing
- *understanding*: signal-non-understanding xor *signal-understanding* xor correct-misspeaking xor nothing
 - *signal-understanding*: acknowledge (via backchannel or assessment) xor repeat-rephrase (via repetition or reformulation) xor completion (via collaborative completion)
- *answer* (complying with an antecedent info-request action, always marked as ‘assert’ too): yes xor no

2.2. Discussion of DAMSL

To better grasp the allowed combinations of dialogue functions, we abstracted the syntactic rules that generate the DA tags. Each utterance can have functions in four dimensions, and in each dimension a certain number of functions can be simultaneously achieved. We simplified some of the notations above. We use the caret ‘^’ as an append sign (as in the following tagsets), ‘?’ means ‘zero or one’, and ‘|’ means ‘exclusive or’ (i.e. pick only one of the list). Underlined functions are terminal tags and subcategories are in italics. A tag for an utterance must be made only of terminal tags (functions).

| | |
|--|-----|
| DA → CS? ^ IL? ^ FLF? ^ BLF? | (1) |
| CS → <u>uninterpretable?</u> ^ <u>abandoned?</u> ^ <u>self-talk?</u> | (2) |
| IL → <u>task</u> <u>task-management</u> <u>communication-management</u> <u>other-level</u> | (3) |
| FLF → <i>statement?</i> ^ <i>influencing-addressee-future-action?</i> ^ <u>info-request?</u> ^ <i>committing-speaker-future-action?</i> ^ <i>conventional?</i> ^ | (4) |

| | |
|---|-----|
| <u>explicit-performative?</u> ^ <u>exclamation?</u> ^ <u>other-ff?</u> | |
| <i>statement</i> → <u>assert</u> <u>reassert</u> <u>other</u> <i>influencing-addressee-future-action</i> → <u>open-option</u> <u>action-directive</u> <i>committing-speaker-future-action</i> → <u>offer</u> <u>commit</u> <i>conventional</i> → <u>opening</u> <u>closing</u> | (5) |
| BLF → <u>agreement?</u> ^ <u>understanding?</u> ^ <u>answer?</u> | (6) |
| <i>agreement</i> → <u>accept</u> <u>accept-part</u> <u>maybe</u> <u>reject-part</u> <u>reject</u> <u>hold</u> <i>understanding</i> → <u>signal-non-understanding</u> <u>acknowledge</u> <u>repeat-rephrase</u> <u>completion</u> <u>correct-misspeaking</u> | (7) |

The DAMSL tagset is an attempt to maintain annotation flexibility and expressivity, by allowing a multidimensional scheme: many of the previous functions (or rather subcategories) could a priori occur in one utterance (except for some infrequent mutually exclusive values). DAMSL does not seem to commit itself to a specific theory which would rule out a certain number of combinations based on a priori (and infeasible) conclusions. An analysis of the DAMSL dimensions reveals some differences with our theoretical dimensions in Chapter 1. Regarding the communication status (CS) and information level (IL) dimensions, we would not think of them as “functions” within a dialog. Regarding forward and backward looking functions, we notice first that these are merely two categories that group lower level “functions”, which are often expressed in terms of traditional speech acts. But FLF and BLF represent two aspects of the adjacency pair dimensions (first part vs. second part), and as such they should be more than just “class names” for sets of speech acts. It is true that there is certainly some correlation between speech acts and adjacency pairs (e.g., questions occur in general as first part), and this correlation is visible in the contents of the FLF and BLF categories.

Obviously, it would be too restrictive to state that an utterance has either a FLF or a BLF (e.g., because assert and reject could appear together, cf. DAMSL section 2.4.3). But if the functions within these categories are not mutually exclusive, then the FLF / BLF distinction is merely a presentation one, as appears from the representation above: FLF and BLF could be replaced directly, in rule (1), by their right-hand part from the respective rewriting rules (4) and (6), and they would thus disappear. So it seems that, at least concerning the FLF and BLF dimensions, the DAMSL tagset resembles a taxonomy of speech acts. But a closer look shows that not all of the hypothesized functions are speech acts: there are problems with exclamation, with opening and closing, with backchannels (in acknowledge), and with answer (as discussed in section 1). Of course, since these are not mutually exclusive, one can pick one or more functions depending on the utterance.

The fact that the tagset has so little mutually exclusive functions (or tags) generates another problem: the number of possible DAMSL tags, i.e. of combinations of all non-mutually-exclusive functions, is very large. It is the product of the number of possible combinations in the four dimensions; the forward and backward looking functions give rise to a huge number of tags. In other words, the DAMSL tagset is subject to very few constraints. The total number of combinations is: $(2^3) \times (4) \times (4 \cdot 3 \cdot 2 \cdot 3 \cdot 3 \cdot 2 \cdot 2 \cdot 2) \times (6 \cdot 6 \cdot 2) = 3,981,313$ possible

combinations (almost 4 million). This makes it quite difficult for a computer program to assign DAMSL functions to a dialog.

It seems nevertheless that behind the FLF / BLF distinction lies the interesting idea that in reality, an utterance could have for instance *at most* one FLF, and/or *at most* one BLF. Or that, at least within FLF and BLF, subclasses could be defined that would be mutually exclusive. This would provide indeed an interesting finding about the combinations of the AP x SA dimensions—i.e. combinations of {first_part, second_part} with {... *speech_acts*...}. This is in fact one of the main contributions of the Switchboard-DAMSL tagset, based mainly on an empirical study of the co-occurrence of DAMSL functions.

2.2. Switchboard-DAMSL (SWBD)

SWBD was inspired from the DAMSL tagset—in fact the two were more or less contemporary, designed by people who participated in the Discourse Research Initiative (DRI). Switchboard is in fact a large corpus of telephonic conversations, on a free topic. The corpus was transcribed and annotated, and the “dialog act” annotation will be here referred to, metonymically, as SWBD. While the annotators started using DAMSL-like functions, it occurred at the end of the experiment that only very few of the DAMSL combinations of functions occurred in reality. Therefore, these *mutually exclusive, mono-dimensional* tags were listed, and the less frequent ones were merged into the most frequent ones, thus arriving at a list of 42 classes, or simplified dialog acts. This is, in our view, a major empirical contribution.

In a first stage of SWBD, the DAMSL set was somewhat simplified, and short names were given to its tags. We quote then details from (Jurafsky et al., 1998:115):

“The SWBD-DAMSL dialog act tag set (Jurafsky et al., 1997) was adapted from the DAMSL tag-set (Core and Allen, 1997), and consists of approximately 60 labels in orthogonal dimensions (so labels from different dimensions could be combined).” [The corpus was then segmented into utterances, though this was not without some theoretical and practical difficulties.] “The average conversation consisted of 144 turns, 271 utterances, and took 28 minutes to label.” [Then, seven annotators labelled the Switchboard corpus,] “resulting in 220 unique tags for the 205,000 SWBD utterances ... each utterance receiv[ing] exactly one of these 220 tags. [...] The labelling agreement was 84% (k = .80).”

This is a very strong empirical result, showing that most of the DAMSL “dialog functions” are in reality mutually exclusive. Still, “the resulting 220 tags included many which were extremely rare, making statistical analysis impossible. [...] We thus clustered the 220 tags into 42 final tags. [...] The 18 most frequent of these 42 tags are shown [below]”.

| | | |
|-------------------------|------------------------|-----------------|
| Statement 36% | Continuer 19% | Opinion 13% |
| Agree/Accept 5% | Abandoned/Turn-Exit 5% | Appreciation 2% |
| Yes-No-Question 2% | Non-verbal 2% | Yes answers 1% |
| Conventional-closing 1% | Uninterpretable 1% | Wh-Question 1% |

| | | |
|-------------------------|-----------------|--------------------------|
| No answers 1% | Response Ack 1% | Hedge 1% |
| Declarative Question 1% | Other 1% | Backchannel-Question 1%. |

The goal of the SWBD project, with respect to dialogue acts, should be reminded here, since its influence on the tagset is very important. The main idea of the project was to try to improve automatic speech recognition methods, by recognizing something called “utterance-types”, related to dialogue acts. As (Jurafsky et al. 1997) put it:

“Our goal is... to build a number of different ‘utterance-type detectors’, based on different sources of evidence for utterance-type: prosodic, acoustic, lexical, and discourse sequence. Given an utterance from the test-set, we will use the predicted utterance-type to select the appropriate utterance-type-specific language model for the utterance.”

So, strictly speaking, SWBD tried to classify utterances into some “types” in order to use them as an intermediate level of representation for an ASR device. The utterance-types are implicitly related to the dialog function of the utterance, but include other elements that the authors believed to be useful. Unfortunately, we have not found evidence that the interesting research program sketched here has been fulfilled.

2.3. Formal analysis and comments on SWBD

SWBD has introduced important changes in the DAMSL set, in several stages. The most straightforward one is the refinement or change of some of the DAMSL functions, in several dimensions. A hidden change is the introduction of (implicit) exclusivity constraints, partly based on theory, partly on the empirical observations quoted above. The hidden change renders the whole tag set “one-dimensional”, i.e. each utterance is supposed to have one and only function (more or less complex).

Let us first represent below the DAMSL tag set with the straightforward modifications introduced by SWBD, in SMALL CAPITALS, followed by the SWBD abbreviation; deleted categories or tags are ~~barred~~. Terminal tags are still underlined and subcategories in italics. There is a special class of SWBD tags which are preceded by a caret (^). This is not really explained in the guidelines, which talk about “secondary caret-dimensions”, but in general they seem to be independent tags and they can be appended to many other tags. Some can also appear alone, e.g. ^q, ^2^g, etc. In the representation below, we already introduced some mutual exclusiveness constraints, based on our interpretation of the SWBD-DAMSL guidelines (there are more ‘|’ symbols instead of the ‘^’ symbols).

| | |
|---|-----|
| DA → CS (IL? ^ FLF? ^ BLF?) | (1) |
| CS → <u>uninterpretable</u> (%)? ^ <u>abandoned</u> (%-)? ^ <u>self-talk</u> (t1)? ^ <u>THIRD-PARTY-TALK</u> (t3)? ^ <u>NON-VERBAL</u> (x)? | (2) |
| IL → task <u>task-management</u> (^t) <u>communication-management</u> (^c) other level | (3) |
| FLF → <i>statement?</i> ^ <i>influencing-addressee-future-action?</i> ^ <i>info request?</i> ^ Δ | (4) |

| | |
|---|------|
| <i>committing-speaker-future-action</i> ? ^ <i>OTHER-FF</i> ? | |
| <i>statement</i> → <u>STATEMENT-NON-OPINION</u> (sd) <u>STATEMENT-OPINION</u> (sv) <u>assert</u> <u>reassert</u> <u>other</u> <i>influencing-addressee-future-action</i> → <u>open-option</u> <u>action-directive</u> <i>INFO-REQUEST</i> <i>committing-speaker-future-action</i> → <u>offer</u> (co) <u>commit</u> (cc) <i>OTHER-FF</i> → <u>conventional-opening</u> (fp)? ^ <u>conventional-closing</u> (fc)? ^ <u>explicit-performative</u> (fx)? ^ <u>exclamation</u> (fe)? ^ <u>other-ff2</u> (fo)? ^ <u>THANKING</u> (ft)? ^ <u>YOU'RE-WELCOME</u> (fw)? ^ <u>APOLOGY</u> (fa)? | (5) |
| <i>INFO-REQUEST</i> → (<u>YES-NO-QUESTION</u> (qy) <u>WH-QUESTION</u> (qw) <u>OPEN-QUESTION</u> (qo) <u>OR-QUESTION</u> (qr) <u>OR-CLAUSE</u> (qrr)) ^ <u>DECLARATIVE-QUESTION</u> (^d)? ^ <u>TAG-QUESTION</u> (^g)? | (5') |
| BLF → <i>agreement</i> ? ^ <i>understanding</i> ? ^ <i>ANSWER</i> ? | (6) |
| <i>agreement</i> → (<u>accept</u> (aa) <u>accept-part</u> (aap) <u>maybe</u> (am) <u>reject-part</u> (arp) <u>reject</u> (ar)) ^ <u>hold</u> (^h)? <i>understanding</i> → <u>signal-non-understanding</u> (br,br^m) <i>SIGNAL-UNDERSTANDING</i> <i>SIGNAL-UNDERSTANDING</i> → <u>acknowledge</u> (b,bh) <u>ACKNOWLEDGE-ANSWER</u> (bk) <u>repeat-phrase</u> (^m) <u>completion</u> (^2) <u>SUMMARIZE</u> (bf) <u>APPRECIATION</u> (ba) <u>SYMPATHY</u> (by) <u>DOWNPLAYER</u> (bd) <u>correct-misspeaking</u> (bc) <i>ANSWER</i> → (<u>YES-ANSWERS</u> (ny) <u>NO-ANSWERS</u> (nn) <u>AFFIRMATIVE-NON-YES-ANSWERS</u> (na) <u>NEGATIVE-NON-NO-ANSWERS</u> (ng) <u>OTHER-ANSWERS</u> (no) <u>DISPREFERRED-ANSWERS</u> (nd)) ^ <u>EXPANSIONS-OF-Y/N-ANSWERS</u> (^e)? | (7) |

The changes introduced in the DAMSL tagset are not without difficulties, and we indicate here some of them.

- Whereas tags in DAMSL were explicitly non-mutually-exclusive, here the names prompted us (in the table above) to hypothesize some dimensions where the tags are mutually exclusive, for example for *INFO-REQUEST* and for *ANSWER*. The SWBD guidelines aren't in general explicit about this (with some exceptions), but the overall result of SWBD, that is, a one-dimensional tagset, suggests that the general approach is to use tags parsimoniously (one per utterance is best).
- Communicative-status (CS): in SWBD, if an utterance is marked with one of the tags in CS, then it will not be marked with any other tags (guidelines, §3)—hence our modification of rule (1).

- Information-level, communication-management (§4.2 of the guidelines). The DAMSL definition is particularly obscured by the fact that other SWBD tags, such as greetings/closings (fp, fc) are said to be also part of communication management. Here is the ambiguous comment:

“The SWBD-DAMSL **^c** tag is only a subset of the DAMSL Communication-Management tag. Communication-Management [APB: *in general*] includes a number of other things which SWBD-DAMSL does not code with **^c**. Following is a paragraph from Allen and Core (page 6), split out on separate lines together with the SWBD-DAMSL tag which corresponds with each SWBD function: ... fp greetings..., fc closings..., b acknowledgements..., b^m repeating parts of what the speaker said, ^h stalling for time..., ?? speech repair..., [...] br, **^c** to address communication problems. [...] So when mapping from SWBD-DAMSL to DAMSL, the tags **fp**, **fc**, **b**, **^h**, and **br** can be mapped automatically to Communication-Management”.

This is quite strange, since these tags can also be mapped, with good reasons, to the respective BLF from DAMSL.

- Info-request: in DAMSL, this was a separate dimension, at the same level as *influencing-addressee-future-action*; here, info-requests are at the same level as *open-option* and *action-directive*, which is OK if everything is mutually-exclusive, but not otherwise. The info-requests cover several types of questions, plus two additional (second-level?) tags, **^d** and **^g**. Now, these can also occur alone, so their status (1st or 2nd level) is not completely clear.
- Another rearrangement has occurred among the “other-forward-function” class, which gathers a diverse set of tags (compare the DAMSL and the SWBD tables).
- In the answer class, the ‘**^e**’ tag functions in a particular way, not completely rendered in our table above, in which it appears as an optional addition. It can be used only for no-plus-expansion (nn^e) or yes-plus-expansion (ny^e) or statement-expanding-y/n-answer(sd^e,sv^e). These mark “only the first utterance after the yes/no answer” (section 6.3.5). In fact, the reason to use this tag has more to do with the segmentation of utterances (the “slash units”). If the utterance has been segmented in two, such as “no, // I live alone”, then it will be tagged “nn // sd^e”, whereas if it has been left as “yeah, I do”, then it will be tagged “ny^e”. So, ‘**^e**’ has nothing much to do with a principled analysis of dialog functions.
- More generally, it seems that the “secondary caret dimensions” share a specific property that is not really explained. They are somehow secondary, and most of them are factored out in the reduced version of the tagset (42 tags).
- The ‘**^q**’ tag, for quotes, not only does not belong in any previous dimension, but the information it encodes is of a very specific nature, not really related to dialog structure: “we code this because we suspect this may effect [sic] pitch and other prosodic features of the utterance” (section 7.1). There is also a (**^q**) tags which denotes partial quoting.
- Another SWBD tag not integrated in the previous DAMSL dimensions is ‘h’ for hedge.

2.4. Final SWBD tagset based on empirical analysis

One of the main contributions of the SWBD tagset is the empirical analysis of the combinations of tags that occur in reality. The theoretical tagset, which was initially highly multidimensional,

was finally reduced to set of 42 *mutually exclusive* tags. As quoted above, based on the guidelines only, the annotators used only “220 unique tags for the 205,000 SWBD utterances [...] [These were clustered] into 42 final tags.” In particular, the “secondary caret dimensions” were removed in the reduced tagset. It is not clear if these caret dimensions were independent from the start, and intended to be some kind of complement to the main set; but in this case they shouldn’t occur alone, as they do.

These results are on one side compelling, and on the other a bit worrying. This is a nice empirical proof that very few combinations of tags can really occur: an utterance is unlikely to be at the same time a statement, a question, and a positive answer (though such examples could be construed). Among the 42 final SWBD tags, only 5 are combined tags from two DAMSL tags: qy^d, qw^y (declarative + question), b^m (signal-understanding + mimic), nn^e (negative non-no answer, grouped with ng) and ny^e (affirmative non-yes answers, grouped with na).

The main problem is that the resulting set of tags seems to label properties that are sometimes remote from dialog *functions*, resulting rather from some tinkering with the DAMSL set. The SWBD tagset gathers very different types of “functions”, and one may wonder why they must be mutually exclusive. For instance, in the SWBD tagset we find:

- speech acts (statements, several types of questions, etc.)
- conversation flow markers (Continuer, Abandoned/Turn-Exit, Backchannel-Question)
- low-level comprehensibility features (Non-verbal, Uninterpretable)
- adjacency-pair information (various types of answers)
- semantic descriptions (yes-answers, i.e. “Yes”, or no-answers, such as “no”)
- conversation structure features (conventional-closing)

Why should an utterance receive a label on only *one* of these types? How could for instance Agreement (accept, maybe, etc.) and Answer be mutually exclusive? The most meaningful interpretation we can give is the following: these tags could indicate the *main function* of the utterance, leaving aside other implicit functions. For instance, a yes-answer is probably also a statement, albeit a very elliptical one, since it could be paraphrased with a full declarative; but it has also a function of continuing the conversation, e.g. some kind of backchannel. Conversely, an utterance tagged as a ‘backchannel’ should be, according to SWBD (and even more to ICSI-MR), an utterance that functions *exclusively* or *mainly* as a backchannel. But in this case, the mapping provided between SWBD and DAMSL is quite misleading, since in DAMSL the various dimensions of “dialog function” are quite independent.

The theoretical stance of DAMSL is therefore diluted, with SWBD, by the empirical studies of the occurrences of tags. But SWBD has also an even more applied dimension: its tagset is designed also in a manner such that it could be assigned automatically by a computer program, especially one that is based on machine learning. In this case, a set of mutually exclusive tags is very convenient, since it reduces a lot the search space. It is not clear whether the SWBD “functions” have been used with profit in a computer application. As mentioned above, the initial motivation of SWBD-DAMSL was to improve spoken dialog recognition. The main results that were published are those of DA recognition itself (Stolcke et al. 2000). The incorporation of DAs into language models slightly improves WER for speech recognition.

Chapter 3. Analysis of ICSI-MR annotation guidelines

3.1. Overview of ICSI-MR guidelines

The ICSI Meeting Recording project focuses on multi-party dialogs. The project defined guidelines for “dialog act labelling”, of which we have seen two versions (August 2002 and August 2003)—one of the questions is, of course: how will resources annotated according to the first version be converted to the second version. The need for a specific tagset, different from SWBD or DAMSL, is motivated by the different nature of the data: multiparty meetings of about 1 hour, as opposed to the SWBD 6 minute dialogs over the phone. The ICSI-MR tagset is based on SWBD-DAMSL, and a “mapping” is provided between the two.

Our interpretation of the guidelines is the following—it should however be supplemented by studying the actual tags assigned by annotators¹. Each utterance has a *label* composed of one or more *tags*, separated by special characters (diacritics).

The first diacritic to understand is the vertical bar (or pipe bar, “|”) which separates both an utterance and its label into two sub-utterances with distinct functions. As in SWBD, this seems mostly due to the technical inability of DA annotators to further segment utterances of the corpus, so a pipe is inserted. But conceptually, the two sides of the pipe seem to behave as separate utterances.

Otherwise, a label is made of exactly one *first tier tag* (or *general tag* in 2003 vocabulary) and zero or more *tier 2 tags* (or *specific tags* in 2003 vocabulary). The second level tags are separated by carets ‘^’. The period ‘.’ is used to add *disruption* or incompleteness tags at the end of a label.

The tags are classified into groups according to their similarity of function, probably with the idea that tags in a group are mutually exclusive, but this is not specified in the guidelines. There are 11 groups. The *general tags* are those from groups 1, 2, 5, and one tag from group 4 (‘b’ for backchannel). Disruption tags are in group 4, and the *specific tags* are in the remaining groups. The tags are listed in section 4.3.1, together with their symbol, their group, etc. The second level tags are ordered alphabetically in a label—that is, their order is not relevant in the label, i.e. the initial ones are not supposed to be more significant than the final ones.

It appears that disruption marks have a more specific behavior than the one we just outlined (H. Carvey, personal communication). The five tags in this group are “%-” (interrupted), “%--” (abandoned), “x” (non-speech), “%” (indecipherable) and “z” (impertinent). They are divided in three subgroups: %- and %--, x and %, and z. The combination rule seems to be the following: .(%-|%-?)?.(x|%)?.z? —that is, possibly one from the first subgroup, possibly one from the second, and possibly z. Moreover, such a label can follow a tier1+tier2 label, but it can also occur alone. For instance, “%-.%” denotes an indecipherable utterance that was interrupted; “%-“ alone denotes an utterance that was interrupted and has almost no content/words (Carvey, personal communication). We believe that these combination rules could be specified a little. In

¹ Thanks to Hannah Carvey from ICSI for her updates and explanations related to the ICSI-MR tagset.

particular, it does not seem coherent to put the ‘%’ and ‘x’ tags in the same category as ‘%-’ and ‘%--’. We have not found examples of ‘z’ yet. We will further analyze these tags below.

In the 2002 ICSI-MR guidelines, a star “*” was used to separate two labels among which the annotator could not decide which one was correct. It seems this has disappeared in the 2003 guidelines. Either the annotator is forced to choose one (*general tag*), or he/she writes the two with a caret between them (*specific tags*); in the second case, that means that two incompatible tags, such as *aap* and *am*, can appear in one label.

3.2. Rule-induction for ICSI-MR labels

Our interest in dialog acts stems from the overall IM2 goal, which is a meeting browser. Therefore, we need to process a meeting by extracting relevant features, in order to be able to answer queries about meetings. Dialog act information is one level of dialog processing that could be useful to answer queries about meetings.

One of the main difficulties in dialog act recognition is to limit the size of the search space, that is the number of possible “dialog acts”, in order to ease the task of the corresponding computer program. There are two ways to limit this size:

- (a) reduce the number of tags, and/or
- (b) reduce the number of possible combinations of tags.

In any case, one needs to formalize the syntax of dialog act labels.

This is what we attempt in this section on the ICSI-MR tagset, going beyond the formula given in the ICSI-MR guidelines (2003), which is:

$$“gen_tag [\wedge spec_tag_1 \dots \wedge spec_tag_n] . disruption”$$

(regardless of the pipe “|”). We first outline a theoretical analysis based on the guidelines, while the empirical analysis follows below. In the formal representation we use rewriting rules (such as the ones for context-free grammars); the pipe or vertical bar below means “or” in the syntax of the rules. Terminals tags are in lower case letters. We presupposed in general that tags within a group are mutually exclusive, unless this is obviously not the case.

| Rule | Explanation |
|----------------------------------|--|
| DA → DA' ^ %? ^ x? | An utterance can be marked as unintelligible (%), even if it has a transcription (!), and/or can be marked as containing non-speech noises (x). |
| DA' → T1 (^ T2)* (.T3)? | A full label is made of a tier_1 label, plus a series of tier_2 labels separated by carets, plus 0 or 1 tier_3 labels. |
| T1 → s Q b FM | A tier_1 tag can be a “statement” (s, group 1, previously contained also sj), a question tag (6 possibilities), a “backchannel” (b), or a floor mover. |
| Q → qh qo qr qrr qw qy | There are six types of questions (group 2). |
| FM → fg fh h | There are three types of “floor movers” (group 5). |
| T3 → %- %-- | Tier 3 is only for disruptions (group 11). In fact, ‘%’ is |

| | |
|--|---|
| | also included in this group, though it has quite a different semantics. ‘%-’ is abandoned, ‘%--’ interrupted. |
| T2 → RE? ^BA? ^CK* ^RI? ^SF? ^PO? ^FD* | All types of tags inT2 are optional, i.e. there can be zero or one of them, except for CK and FD, of which there can be zero, one, or more. |
| RE → RP RN RU | A response (group 3) can either be positive, or negative, or uncertain. |
| BA → bh bk | Backchannels/acknowledgements in tier 2 (group 4). Not really backchannels, after all. The typical ones are marked ‘b’ in tier_1. |
| AC → cc co cs fx | Actions (group 6). Maybe the old ‘oo’ (open-option, kind of suggestion) could come here. |
| CK → br bu f g | Checks (group 7). It seems that these <i>are not mutually exclusive</i> : there are examples such as ‘qy^g^f’. Also, note that ‘g’ often follows a ‘bu’, ‘bu’ often follows an ‘s’, while ‘f’ and ‘g’ seem to follow questions. |
| RI → RIC RIR RIC → bc bsc RIR → r bs m | Restated information (group 8) can be a correction or a repetition. All these tags are mutually exclusive. |
| SF → ba df fe 2 e | Supportive functions (group 9) |
| PO → bd by fa ft fw | Politeness (group 10). These tags seem to be mutually exclusive. Note that ‘by’ is often followed by a ‘bd’; ‘by’ resembles in form ‘fa’; ‘bd’ often follows a ‘ba’ (appreciation, in SF) |
| FD → rt t t3 tc tl d j | “Further descriptors” (group 12) <i>do not seem to be mutually exclusive</i> (hence the FD* in the rule for T2). This is a very diverse set. |

3.3. Criticism of the tagset

One of the main problems with ICSI-MR is having a multidimensional set (several tags allowed per utterance) which is derived from SWBD, which is one-dimensional. Even though a “mapping” was provided in the ICI-MR guidelines, in fact this is rather a loose correspondence: new tags have been introduced, the semantics of some existing tags has been altered, and, especially, the new set is (again) multidimensional. So, talking about the semantics of each tag is quite vague if on one side the tags must appear alone (SWBD), and on the other, they can be combined. It is as if the effort of building a mutually-exclusive tagset (SWBD) has been cancelled by the need to combine those tags, acknowledging that utterances have functions in several dimensions. But in this case, using DAMSL to derive the ICSI-MR tagset would have been more coherent than using SWBD.

The combinations of tags into labels are not completely free, but the number of tag combinations is still huge. In fact, the only explicit case of exclusiveness is in tier 1 (or “general tags”), which must get one and only one tag from a set of ten tags. But within tier 2 (the special

tags) there is no explicit instruction stating that some tags cannot or should not be used at the same time. Now, obviously, a response cannot be positive, uncertain and negative at the same time: so tags in group 3 *are* probably mutually exclusive, even if the guidelines do not say so. In fact, the only groups in which tags are not mutually exclusive (i.e. several of them *can* be used at the same time) seem to be checks (group 7) and further descriptors (group 12). This is why the rewriting rule for tier 2 (specific tags) looks like this: T2 → RE?^BA?^CK*^RI?^SF?^PO?^FD* (there are 2 stars and 5 question marks). Moreover, there are also dependencies between tier 1 and tier 2 tags that do not appear in the guidelines: for instance, it is unlikely to have a question ('q...') in tier 1 followed by a response in tier 2. Rather, a response in tier 2 will always follow a statement. (Of course these claims could be also tested empirically).

Regarding the vertical bar, we believe that this separation of an utterance in two should in fact be made at a previous stage, so that when we talk about annotating dialog acts, each utterance has only one label, without vertical bars. In our proposal, we will therefore consider that each utterance receives one or more tags *as a whole*. If some kind of specialization can be found in the utterance, then the utterance has to be divided in two (or more) parts, and each one annotated separately.

Here are some other details, some derived from observations on the data we kindly received from ICSI:

- 'b' (backchannel) is the only tag in group 4 which is a tier 1 tag (general tag): the other two ('bk' and 'bh') are tier 2 tags. We believe these should be separated. Note that 'bh' seems to be preceded by a QUESTION tag, whereas 'bk' seems to be preceded by an 's'. In fact, 'bk' denotes a case when the speaker speaks directly to the previous speaker, not just "mumbles" (p.27).
- most of the time the 'b' tag is *alone*, and we noted from the data that it can be only followed by 'rt' (rising tone!?), i.e. 'b^rt'. So it seems a 'b' in the first tier prevents the appearance of other tags in the second tier.
- in the data we kindly received, 'h' is never a tier 1 tag as it should be according to the guidelines (which say "Holds occur at the beginning of an utterance, floor-holders in the middle or at the end").
- in principle, 'oo' and 'sj' have disappeared from the 2002 guidelines (in the 2003 version). We hope the data was cleaned accordingly. We noticed also an 'sd' tag in an example from the 2002 guidelines (entry for "s", page 17), reminiscent of SWBD.

3.4 Empirical analysis of six ICSI dialogues: some facts and inaccuracies

This section is now superseded by Annexes A and B.

Within the IM2 project, the DA-tagged data was received by ISSCO from ICSI as part of the IM2/ICSI agreement². We received in November 2002 six one-hour dialogs (Bmr001, 007, 012, 014 and Bro011, 017), tagged with dialog act information according to the ICSI-MR guidelines. The analysis below attempts to check whether the observed tags conform to the guidelines, and

² We are grateful to Liz Shriberg (ICSI and SRI) and Barbara Peskin (ICSI) for providing us with the data, and for their related explanations.

to provide an estimate of their frequencies. We extracted only the tags from the dialogs, so the following observations do not indicate exactly from which of the six dialogues a tag comes from. Note first that there are about 56 tags that can be combined into labels, and even with the minimal constraints set in the previous section, there is still a huge number of possible combinations (remember the 4 million DAMSL combinations).

There are about 7130 utterances tagged with dialog acts labels in the six dialogues we received. There are strictly speaking 570 different labels (combinations), but some do have a vertical bar. After separating the left and right sides of the bar as different labels (1 label → 2 labels), we get 7578 labels (tokens) with now only 432 different combinations of tags. The most frequent ones are the following (more than 10 occurrences):

| | | |
|---------------|--------------|-------------|
| 2057 s | 53 s^ar | 20 qy |
| 1178 b | 51 s^df | 19 s^t |
| 425 fh | 48 s.x | 16 s^t1 |
| 323 s^aa | 46 s^j | 15 s^t3 |
| 266 s.%-- | 45 s^no | 15 s^bs |
| 247 s^bk | 42 qy^d^g^rt | 15 qh |
| 212 fg | 36 s^nd | 14 s^ft |
| 165 s^cs | 36 % | 13 b^rt |
| 150 s.%- | 35 s^2 | 12 sj^fe |
| 144 sj | 35 qy^d^rt | 12 s^r |
| 129 %- | 33 s^cc | 12 s^arp |
| 95 sj^ba | 32 b.% | 12 s^aap |
| 78 s^rt | 31 s^bu | 12 qy^g^rt |
| 77 qy^rt | 28 s^tc | 11 qy.%- |
| 76 %- | 27 s^ng | 10 s^fa |
| 71 h | 26 qy^g | 10 s^cs.%-- |
| 68 qw | 26 qw^rt | 10 s^am |
| 55 s^co | 25 fh.%-- | 10 qy.%-- |
| 54 s^e | 24 s^fe | |
| 53 s^na | 23 fg.%- | |

We emphasized the occurrences of ‘%-’, ‘%--’ and ‘%’ alone. As we discussed, the meaning of these tags is not completely clear, since ‘%-’ alone must mean not only ‘interrupted’ but also ‘indecipherable’ since no other tag is added. Obviously, these tags are quite frequently used alone as labels.

In fact, ‘%’ appears alone only 36 times. Note that ‘%--’ appears alone 76 times, while normally it is not supposed to appear alone; it is even worse for ‘%-’ which appears alone 129 times. Normally, only ‘%’ is supposed to appear alone, so either some tag should be inserted before all isolated ‘%-’ and ‘%--’, or they should be converted to ‘%.%-’ or ‘%.%--’ if we want to record that the corresponding utterances were interrupted. There are only two utterances marked ‘%.%-’

‘, which makes a bit more sense: “undecipherable and interrupted/abandoned”. On the other hand, undecipherable (%) should suffice alone. There is also one utterance marked ‘%.-’ which has no meaning.

All in all, disruption tags (or other tags) are appended with a dot to 812 utterances. If we separate these disruption marks we find that there are 95 x, 65 %, 258 %-, and 384 %--. Of course, the 65 occurrences of ‘%’ after a full stop are again a little puzzling: how can an indecipherable utterance be labelled with a dialog act? According to (Harvey, personal communication):

"x" and "%" CAN go at the end of ANY tag. Sometimes the transcription is there but for some reason we can't hear the words when we are annotating, so in these cases we label according to the official transcript and then add "%." to the end (this can lead to things like "s.%--.%). In cases where the sound file is distorted, we put our best guess as to what was there and add ".x" to the end. If there is a transcription, but for some reason the annotator cannot hear the utterance, then it is marked ‘%’.

If this is true, ‘%’ would be just a technical mark, not really a dialogue function. Anyway, the 65 occurrences of ‘%’ at the end of a label appear as follows: 32 times b.%, 6 times s.%, 3 sj^fe.%, 3 fh.%, 2 sj.%, 2 s^no.%, 2 s^co.%, 2 s^aa.%, 2 qy^g.%, 2 qy.%, and one each of: sj^ba.%, s^t1.%, s^fa.%, s^cs.%, s^ar.%, qy^r.%, qy^g^f.%, qy^d^rt.%, qw.%. Some of the tags used with ‘%’ are indeed quite specific, so why one may wonder how they are assigned if the utterance is undecipherable.

Other tags can follow a full stop too, but these are *mistakes*: ‘%_’ and ‘.-’ (once). We find twice the label ‘b.fg’, and three times ‘fg.qw’ and ‘fg.s’, which are again *mistakes* (two tier 1 tags cannot appear); probably the dot stands here for a vertical bar? Note that there are all in all 10 labels containing two dots, eight of them having ‘.%--.x’ at the end (correct) and two being ‘fg.qw.%-’ (*mistaken*).

If we now disregard the disruption forms and look only at the ‘tier_1(^tier_2)*’ combinations, there are at most *four* second level tags in a label. There are 13 occurrences of labels (out of 7578) in this case: qy^bk^t3^rt^d, qy^co^d^rt^bu, qy^co^rt^t^tc, qy^d^bu^m^rt, qy^d^f^g^rt, qy^d^g^rt^f, qy^d^rt^bu^bsc, qy^d^rt^bu^t3, qy^f^g^d^rt, **qy^g^d^rt^f**, **qy^g^d^rt^f**, qy^t^tc^cs^d, s^tc^t^rt^co. Note that #10 and #11 are the same (**bold**), and #5 and #6 and #9 are the same too (*italics*), but have different tag orders (in the 2003 guidelines, they are supposed to be ordered alphabetically). There are 112 occurrences of labels with three tier 2 tags, etc.

Let us now look at tier_1 tags only, i.e. labels before a full stop or isolated. There are 6525 labels that are not pure disruption forms, and among these, there are 4209 labels with no second tier (no caret ^). We still find here one ‘fh*’ (strange), and two s* (strange again). No other ‘*’ appear elsewhere in the labels, so we remove in what follows these three characters. Here are the numbers of tier 1 labels among the 4209 labels with no second tier:

| | | | | |
|--------|------------|--------|--------|--------|
| 2059 s | 1178 b | 426 fh | 212 fg | 144 sj |
| 71 h | 68 qw | 20 qy | 15 qh | 5 qr |
| 4 qo | 3 x | 3 qrr | | |

It is strange to see the label ‘x’ three times as a first tier tag, since our understanding is that it can only terminate a label. But its definition does not prevent it from being an isolated tag, though in this case the utterance should be also marked ‘%’. Otherwise, we find in this list the tier 1 labels only: ‘s’, ‘sj’ (which disappeared in ICSI 2003), the six questions ‘q...’, the backchannels ‘b’, and the floor holders/grabbers ‘fh’, ‘fg’, ‘h’.

Let us now see what kind of tier_2 tags (specific) can follow the tier_1 (general) tags. We remove first the disruption indications , remove all the isolated %, %-, %-- and therefore keep 7329 labels, then select only the tier 2 labels regardless of the tier 1 label. We are left with 2468 occurrences, among which we find 229 different types of labels (the order being relevant). The most frequent types (i.e. the ones that appear more than 10 times) are listed below, with the number of tokens for each one.

| | | | | |
|--------|---------|----------|------------|-----------|
| 339 aa | 248 bk | 213 rt | 200 cs | 98 ba |
| 62 ar | 61 na | 61 co | 59 j | 58 e |
| 55 df | 52 no | 43 nd | 43 d^rt | 42 d^g^rt |
| 41 bu | 39 tc | 39 fe | 35 cc | 35 2 |
| 28 ng | 28 g | 24 t | 19 t1 | 18 t3 |
| 18 r | 16 bs | 14 ft | 14 fa | 14 arp |
| 13 aap | 12 g^rt | 12 bu^rt | 11 d^bu^rt | 11 am |

This list shows mainly that most tier_2 tags appear alone (as in the SWBD empirical study), except some frequent combinations such as: ‘d^__^rt’, i.e. declarative (question) with rising tone, and in place of __ either a tag question (‘g’) or an understanding check (‘bu’). Above all, these frequencies show that the use in ICSI-MR of the SWBD tagset, which was designed to be mutually exclusive, still yields a mutually exclusive set (i.e. most tags appear alone). Note however that we are talking here about tier_2 tags, which are *always* combined with a tier 1 tag. So combinations of two or three tags are in fact quite frequent.

Chapter 4. An abstraction of ICSI-MR dialog act tagset

In this chapter, we propose an attempt to abstract (generalize) the ICSI-MR tagset by grouping some of the tags together, mostly along the groups that were already defined in the ICSI-MR guidelines. The main goal is to reduce the number of possible combinations of tags, in order to reduce the search space for a dialog act tagger (especially one based on machine learning). Also, having in mind a meeting retrieval application, we figure that users would not query a database of meeting recordings using very specific tags, but queries would typically refer to more general concepts such as “question”, “order”, “promise”. The matter of finding exactly how users would search for particular utterances or episodes in a meeting is of course complex, and is the object of ongoing research.

In this chapter, our proposal remains close to the ICSI-MR tagset, which it abstracts; experiments on the ICSI-MR data are analyzed. In the next chapter, we propose a more principled tagset, departing from the ICSI-MR one. In both cases, since user studies are only incipient, we do not have strong constraints on the “usefulness” of our tagset.

4.1. Proposal: MALTUS - BIS

The table below is derived from our analysis of the ICSI-MR tagset, and contains our proposal for a simpler, more general tagset. The rewriting rules contain three types of categories or tags, starting from ‘DA’, which is the label itself: non-terminal ones in non-bold capitals, terminal ones for our proposal in bold capitals, and possible refinements (i.e. the ICSI-MR tags) in lower case non bold letters. So, only the terminal tags (bold capitals) can compose a dialog act label (refinements are quoted here only as a perspective). Some of the ICSI-MR groups are not represented in this abstraction: they have been considered too remove from “dialogue function” to appear here (e.g., group 9, “Supportive functions”, and group 12, “Further descriptors”).

| Rewriting rule | Explanation |
|----------------------------------|---|
| DA → (U T1 (^T2)?) (.D)? | An utterance is either unintelligible (U) or it is marked by a “full” label. A full label is made of a tier 1 or general tag (T1), plus an optional tier 2 label (T2) which can be composed of several labels. Both can be followed by a disruption mark D , which is a terminal label that groups abandoned and interrupted speech (the difference between the two is not very clear) and non speech noises. |
| T1 → S Q B H | A tier 1 tag can be a statement , a question , a floor-holder or a backchannel . |
| Q → qh qo qr qrr qw qy | <i>Optional</i> : Q could be refined in 6 types of questions (ICSI group 2). |
| H → fg fh h | <i>Optional</i> : H could be refined in 3 types (ICSI group 5). |
| T2 → RE?^AT?^DO?^RI? ^PO? | A tier 2 (secondary or special) label is composed of a series of tags, grouped for convenience in five categories. In fact, three are terminal tags: attention tag (AT), action tag, and politeness (PO); they could be refined based on ICSI-MR. The two other tags are in fact non terminal categories, and <i>must</i> be refined: response tag (RE) and restated-info tag (RI). |

| | |
|---|--|
| RE → RP RN RU | A response RE (group 3) can either be positive, or negative, or uncertain. |
| AT | 'AT' for attention groups backchannels/acknowledgements (ICSI group 4 except the 'b') and checks (ICSI group 7). It seems that these <i>are not mutually exclusive</i> , so several AT could appear in the conversion process, but only one of course should be left in the end. |
| DO → cc co cs fx | "DO" corresponds to Actions (group 6) plus the old 'oo'. It means someone is told to do something (includes commitment, other performatives). <i>Optional</i> : DO could be refined in 4 or 5 types (ICSI group 6). |
| RI → RIC RIR RIC → bc bsc RIR → r bs m | Restated information (group 8) can be a correction or a repetition. <i>Optional</i> : could be refined as shown. |
| PO → bd by fa ft fw | Politeness (ICSI group 10). <i>Optional</i> : could be refined as shown. |

The tagset proposed here has thus the following features:

- It is based on the ICSI-MR set, which is itself a multidimensional version of the SWBD tagset; the main categories in ICSI-MR and SWBD are themselves inspired from DAMSL.
- It is an abstract set, in the sense that the assigned tags encompass broad meanings, and could be refined further on.
- It is a layered set, in the sense that labels have a principal component (tier_1 tag) followed by a number of secondary components.
- It is a "constrained multidimensional" set: some properties of utterances are mutually exclusive (ME), and some are not. In particular, the possible values of tier_1 are ME; then, the sub-dimensions of tier_2 are not ME, but their values (when there are several options) are ME.

Therefore, we will refer to this tagset as: "multidimensional abstract layered tagset for utterances, based on ICSI-MR and SWBD", or shortly MALTUS-BIS³.

The number of possible tags is greatly reduced. The number of possible combinations is: $(1 + (4 \times 4 \times 3 \times 2 \times 2 \times 2)) \times 2 = 770$. This number is divided by two if one is not interested in annotating disruption marks. Also, if one of the tier_2 dimensions is removed, the number is divided again by 2 or more (the number of options for that dimension plus one).

We now summarize the formal description as a set of guidelines for human annotators—or for someone designing an automatic dialog act tagger.

| |
|---|
| <u>MALTUS Guidelines</u> |
| <p>1. Is the utterance intelligible?</p> <ul style="list-style-type: none"> - yes: use a tier_1 tag and one or more tier_2 tags - no: use U |

³ Multidimensional Abstract Layered Tagset for UtteranceS – Based on ICSI-MR and SWBD: MALTUS – BIS. The connection with Malthus appears if you note that this tagset attempts to reduce the number of possible labels.

2. Whatever the answer to (1) is, you can add a disruption tag **D** to note that the utterance has been interrupted or abandoned, or if it contains significant non speech noises.

3. For the tier_1 tag, use *exactly one* of the following:

- **S** if the utterance is a statement
- **Q** if the utterance is a question (not merely in its form, but in its “function”)
- **B** if the utterance is only a backchannel
- **H** if the utterance is a floor-holder, floor-grabber, or hold

4. For the tier_2 tag, answer *all* of the following questions—tier_2 can contain zero, one or more tags:

- 4a. Is the utterance a response? If yes, use *exactly one* of the following, if not, do nothing:
- **RP** for a positive response
 - **RN** for a negative response
 - **RU** for other responses (undecided)
- 4b. Does the utterance restate (in a very similar wording) previous information? If yes, use *exactly one* of the following, if not, do nothing:
- **RIC** if the restated information is a correction
 - **RIR** if it is only a repetition
- 4c. Is the utterance related to an action such as an order, a suggestion, a promise, etc.? Does it suggest doing something? If yes, use a **DO** tag, if not, do nothing.
- 4d. Is the utterance related to attention management in discourse? (such as tag questions, acknowledgements, etc. but excluding pure backchannels which are marked as **B** in tier 1). If yes, use an **AT** tag, if not, do nothing.
- 4e. Is the utterance strongly related to politeness? (examples include “thanks”, “you’re welcome”, downplayers, etc.) If yes, use a **PO** tag, if not, do nothing.

Formally, the rules can be expressed simply as follows. Note that we keep the ‘^’ and ‘.’ signs from the ICSI-MR tagset, though there is no reason to maintain a difference between the two, since D cannot be mistaken with other tags. Again, symbols in bold are terminal.

- DA → (U | T1 (^T2)?) (.D)?
- T1 → **S** | **Q** | **B** | **H**
- T2 → (**RP** | **RN** | **RU**)? ^AT? ^DO? ^(RIC | RIR)? ^PO?

Or in only one rule, the general form of a MALTUS label is:

(U | ((**S** | **Q** | **B** | **H**) (^**RP** | ^**RN** | ^**RU**)? (^**RIC** | ^**RIR**)? (^**DO**)? (^**AT**)? (^**PO**)?)) (.D)?

In addition, these labels can be glossed using keywords. The keywords could be used in a meeting retrieval application to match user queries (in this case, some variants should probably be added). The following is a summary of the abbreviations:

U = undecipherable (unclear, noisy)

| | |
|-----|---|
| S | = statement |
| Q | = question |
| B | = backchannel |
| H | = hold (floor holder, floor grabber, hold) |
| RP | = positive answer (or positive response) |
| RN | = negative answer (or negative response) |
| RU | = other answer (or undecided answer or response) |
| RIC | = restated information with correction |
| RIR | = restated information with repetition |
| DO | = command or other performative (includes: command, commitment, suggestion, open-option, explicit performative) |
| AT | = the utterance is related to attention management (acknowledgement, rhetorical question backchannel, understanding check, follow me, tag question) |
| PO | = the utterance is related to politeness (sympathy, apology, downplayer, “thanks”, “you’re welcome”) |
| D | = the utterance has been interrupted or abandoned |

4.2. Empirical study: conversion of ICSI-MR labels to MALTUS

We defined a procedure to convert the ICSI-MR dialog act labels to our abstract set, based on the definition of the MALTUS-BIS tagset from the ICSI-MR groups (see table in section 4.3.1). The conversion procedure is mainly a simple replacement of ICSI-MR tags with more general ones, while preserving the ‘^’ and ‘.’ signs. The conversion cannot operate on ill-formed ICSI-MR labels, for instance those that have a ‘.’ sign twice. We believe these labels should be corrected prior to conversion (e.g. in a further release of the annotated data).

The first pass conversion generates the following labels from the six initial ICSI-MR dialogues received by MDM. The simplified MALTUS tagset appears in capitals (as above), and where lower case letters appear (highlighted), it is because the ICSI-MR label did not conform to the original guidelines. Some tiny mistakes were corrected before this conversion, such as ‘x’ alone, 3 times ‘s.fg’, etc. It is not impossible that the mistaken labels (in bold) that appear below are due to conversion errors; this will be checked later. The list below is the list of all tags, so there are now 53 different types of labels (60 minus those in bold).

| | | | | |
|-----------------|---------------|--------------|--------------|--------------|
| 2783 S | 1226 B | 724 H | 453 S.D | 409 S^RP |
| 328 Q | 307 S^DO | 295 S^AT | 245 U | 181 Q^AT |
| 141 S^RN | 65 S^RU | 58 H.D | 57 S^RI | 52 Q.D |
| 38 S^PO | 28 S^DO.D | 22 Q^AT^AT | 18 S^RP^RI | 16 Q^DO |
| 15 S^RN.D | 13 S^RP.D | 9 S^AT^RI | 7 S^RI^AT | 7 S^DO^RP |
| 6 S^RI^RP | 6 S^RI.D | 6 S^AT.D | 6 Q^AT.D | 5 S^RU.D |
| 4 S^RU^RI | 4 H^RI | 3 S^RN^DO | 3 Q^RI | 3 Q^AT^RI |
| 3 H.qw | 2 U.D | 2 S^PO^RI | 2 S^DO^RN | 2 Q^RN |
| 2 H.qw.D | 2 B.fg | 1 S^n | 1 S^h | 1 S^RU^RN |
| 1 S^RP^DO | 1 S^RN^RN | 1 S^RN^RI.D | 1 S^PO.D | 1 S^DO^DO^RP |
| 1 S^DO^DO | 1 Q^y | 1 Q^RP | 1 Q^DO^AT | 1 Q^DO.D |
| 1 Q^ | 1 H^RP | 1 H^RI.D | 1 H^AT | 1 B.D |

Given that each label can appear, in theory, with or without a final ‘.D’, we will remove this terminal information when present and find out the new number of tags (removing the previous errors):

| | | | |
|-------------------|------------|------------|-----------|
| 3236 S | 1227 B | 782 H | 422 S^RP |
| 380 Q | 335 S^DO | 301 S^AT | 247 U |
| 187 Q^AT | 156 S^RN | 70 S^RU | 63 S^RI |
| 39 S^PO | 22 Q^AT^AT | 18 S^RP^RI | 17 Q^DO |
| 9 S^AT^RI | 7 S^RI^AT | 7 S^DO^RP | 6 S^RI^RP |
| 5 H^RI | 4 S^RU^RI | 3 S^RN^DO | 3 Q^RI |
| 3 Q^AT^RI | 2 S^PO^RI | 2 S^DO^RN | 2 Q^RN |
| 1 S^RU^RN! | 1 S^RP^DO | 1 S^RN^RN | 1 S^RN^RI |
| 1 S^DO^DO^RP | 1 S^DO^DO | 1 Q^RP | 1 Q^DO^AT |
| 1 H^RP | 1 H^AT | | |

In this list, there are several duplicates, due to different ordering of the tier_2 tags. Since the order is not relevant in tier_2, we can reorder the tags and conflate duplicate labels, so that the final list of occurring labels, apart from disruption marks, is:

| | | | |
|---------|---------|----------|---------|
| U | S^RN | S^RI^PO | Q^DO^AT |
| B | S^RN^DO | S^PO | Q^AT |
| S | S^RN^RI | | Q^AT^RI |
| S^RU | | Q | |
| S^RU^RI | S^DO | Q^RP (1) | H |
| S^RP | S^AT | Q^RN (2) | H^RP |
| S^RP^DO | S^AT^RI | Q^RI | H^AT |
| S^RP^RI | S^RI | Q^DO | H^RI |

Only 29 combinations of MALTUS tags appear in this data—remember that there are ca. 7500 utterances in the six ICSI dialogues we analyzed. Comparing this figure with the 60 SWBD tags, it appears that a more abstract set allows a reduction of the number of possible labels, even when several dimensions are used. Of course, the limited size of our data also accounts for the limited number of tags. It is quite likely that more combinations will occur when more data is available to us. However, the number of occurring labels will probably remain in the SWBD range, which is *much* lower than the number of possibilities, and even the number of occurring tags, with the ICSI-MR tagset.

4.3. Correspondence with ICSI-MR, DAMSL and SWBD

The correspondence between MALTUS and ICSI-MR is quite clear: we summarize it in the first subsection below (4.3.1). Questions arise about the mapping to and from DAMSL and SWBD-DAMSL. We provide tables (4.3.2) that show for each DAMSL or SWBD tag its correspondent MALTUS tag. This is quite simple since the authors of ICSI-MR provide a correspondence from SWBD, and the authors of SWBD did the same from DAMSL. We also use below the MATE report (see beginning of 4.3.2).

However, it is obvious that such mappings are imperfect for two reasons: first, since MALTUS is a rather abstract tagset, the “mapping” works only in one direction, from the more specific

(ICSI-MR / SWBD / DAMSL) to the more abstract tagset (MALTUS). Indeed, a more abstract tagset cannot be mapped towards a more detailed one. Second, the problem of dimensionality makes a mapping incomplete, if one does not state which tags are mutually exclusive according to the guidelines. For instance, a conversion from SWBD to MALTUS would generate for each utterance only one tag (or sometimes two) from the abstract set, while up to six tags can be combined for an utterance. On the other hand, a conversion from DAMSL would probably use more often several of the dimensions in the abstract set.

4.3.1 ICSI dialogue acts and their mapping to the IM2.MDM abstraction

The following table provides an overview of all the ICSI-MR dialog acts, with an index number (#), the level or tier (1, 2, or 3), their name and group in the ICSI-MR 2002 and 2003 versions. We provide then the formal label that was assigned to the dialog acts in our formalization (“form. label”, cf. section 3.2), the final symbol in the IM2.MDM abstraction (IM2.MDM) and a brief gloss or keyword. The glosses sometimes appear even for tags that have not been selected for the abstract tagset, as a suggestion for possible refinements, if they are needed by an application. The format is: “full_gloss > *refinement*”.

| # | T | Abbr | ICSI Name | ICSI 2002 Group | ICSI 2003 Group | Form. label | IM2.MDM | IM2.MDM gloss |
|----|----|------|---------------------------------|----------------------------|----------------------------------|-------------|---------|--------------------------------|
| 2 | 1 | sj | Subjective Statement | statements | <disappeared> | s | S | statement |
| 1 | 1 | s | Statement | statements | 1 Statements | s | S | statement |
| 33 | 1 | qo | Open-ended Question | Forward Functions | 2 Questions | Q | Q | question |
| 3 | 1 | qr | Or Question | questions | 2 Questions | Q | Q | question |
| 4 | 1 | qrr | Or Clause After Y/N Question | questions | 2 Questions | Q | Q | question |
| 5 | 1 | qw | Wh-Question | questions | 2 Questions | Q | Q | question |
| 6 | 1 | qy | Y/N Question | questions | 2 Questions | Q | Q | question |
| 18 | 1 | qh | Rhetorical Question | Short Utterances | 2 Questions | Q | Q | question |
| 12 | 1 | b | Backchannel | Short Utterances | 4 Backchannels / Acknowledgments | b | B | backchannel |
| 14 | 1 | fg | Floor Grabber | Short Utterances | 5 Floor movers | FM | H | attention |
| 15 | 1 | fh | Floor Holder | Short Utterances | 5 Floor movers | FM | H | attention |
| 17 | 1 | h | Hold | Short Utterances | 5 Floor movers | FM | H | attention |
| 7 | 1* | % | Indecipherable | Disruption Forms | 11 Disruption | % | U | unclear |
| 32 | 2 | oo | Open-Option | Forward Functions | <disappeared> | AC | DO | action > <i>suggestion</i> |
| 21 | 2 | aap | Partial Accept | Responses | 3a Response: positive | RP | RP | positive answer |
| 26 | 2 | na | Affirmative Answer | Responses | 3a Response: positive | RP | RP | positive answer |
| 16 | 2 | aa | Accept | Short Utterances | 3a Response: positive | RP | RP | positive answer |
| 24 | 2 | ar | Reject | Responses | 3b Response: negative | RN | RN | negative answer |
| 22 | 2 | arp | Partial Reject | Responses | 3b Response: negative | RN | RN | negative answer |
| 23 | 2 | nd | Dispreferred Answer | Responses | 3b Response: negative | RN | RN | negative answer |
| 23 | 2 | ng | Negative Answer | Responses | 3b Response: negative | RN | RN | negative answer |
| 25 | 2 | am | Maybe | Responses | 3c Response: uncertain | RU | RU | other answer |
| 27 | 2 | no | Other | Responses | 3c Response: uncertain | RU | RU | other answer |
| 19 | 2 | bh | Rhetorical Question Backchannel | Short Utterances | 4 Backchannels / Acknowledgments | BA | AT | attention > <i>acknowledge</i> |
| 13 | 2 | bk | Acknowledgement | Short Utterances | 4 Backchannels / Acknowledgments | BA | AT | attention > <i>acknowledge</i> |
| 30 | 2 | cc | Commitment | Forward Functions | 6 Actions | AC | DO | action > <i>commitment</i> |
| 28 | 2 | co | Command | Forward Functions | 6 Actions | AC | DO | action > <i>command</i> |
| 31 | 2 | cs | Suggestion | Forward Functions | 6 Actions | AC | DO | action > <i>suggestion</i> |
| 29 | 2 | fx | Explicit-Performative | Forward Functions | 6 Actions | AC | DO | action > <i>performative</i> |
| 39 | 2 | br | Repetition Request | Backward-Looking Functions | 7 Checks | CK | AT | attention > <i>check</i> |
| 41 | 2 | bu | Understanding Check | Backward-Looking Functions | 7 Checks | CK | AT | attention > <i>check</i> |
| 49 | 2 | f | Follow Me | Descriptive Tags | 7 Checks | CK | AT | attention |

| | | | | | | | | |
|----|---|-----|--------------------------|----------------------------|-----------------------------|-----|-----|-------------------------|
| 50 | 2 | g | Tag Question | Descriptive Tags | 7 Checks | CK | AT | attention |
| 36 | 2 | bc | Correct-Misspeaking | Backward-Looking Functions | 8a Restated_info_correction | RIC | RIC | r. i. correction |
| 37 | 2 | bsc | Self-Correct Misspeaking | Backward-Looking Functions | 8a Restated_info_correction | RIC | RIC | r. i. correction |
| 53 | 2 | r | Repeat | Descriptive Tags | 8b Restated_info_repetition | RIR | RIR | r. i. repetition |
| 40 | 2 | bs | Summary | Backward-Looking Functions | 8b Restated_info_repetition | RIR | RIR | r. i. repetition |
| 52 | 2 | m | Mimic | Descriptive Tags | 8b Restated_info_repetition | RIR | RIR | r. i. repetition |
| 35 | 2 | ba | Assessment/Appreciation | Backward-Looking Functions | 9 Supportive functions | SF | - | - > <i>appreciation</i> |
| 42 | 2 | df | Defending/Explanation | Backward-Looking Functions | 9 Supportive functions | SF | - | - > <i>defending</i> |
| 55 | 2 | fe | Exclamation | Conventionalities | 9 Supportive functions | SF | - | - > <i>exclamation</i> |
| 45 | 2 | 2 | Collaborative Completion | Descriptive Tags | 9 Supportive functions | SF | - | - |
| 46 | 2 | e | Elaboration | Descriptive Tags | 9 Supportive functions | SF | - | - |
| 38 | 2 | bd | Downplayer | Backward-Looking Functions | 10 Politeness | PO | PO | politeness |
| 44 | 2 | by | Sympathy | Backward-Looking Functions | 10 Politeness | PO | PO | politeness |
| 43 | 2 | fa | Apology | Backward-Looking Functions | 10 Politeness | PO | PO | politeness |
| 56 | 2 | ft | Thanks | Conventionalities | 10 Politeness | PO | PO | politeness |
| 57 | 2 | fw | Welcome | Conventionalities | 10 Politeness | PO | PO | politeness |
| 54 | 2 | rt | Rising tone | Descriptive Tags | 12 Further descriptors | FD | - | - |
| 47 | 2 | t | About-Task | Descriptive Tags | 12 Further descriptors | FD | - | - |
| 10 | 2 | t3 | 3rd-Party Talk | Disruption Forms | 12 Further descriptors | FD | - | - |
| 34 | 2 | tc | Topic Change | Forward Functions | 12 Further descriptors | FD | - | - > <i>topic_change</i> |
| 20 | 2 | t1 | Self Talk | Short Utterances | 12 Further descriptors | FD | - | - |
| 48 | 2 | d | Declarative Question | Descriptive Tags | 12 Further descriptors | FD | - | - |
| 51 | 2 | j | Joke | Descriptive Tags | 12 Further descriptors | FD | - | - > <i>humor</i> |
| 11 | 3 | x | Nonspeech | Disruption Forms | 11 Disruption | D | D | disruption |
| 8 | 3 | %- | Interrupted | Disruption Forms | 11 Disruption | D | D | disruption |
| 9 | 3 | %-- | Abandoned | Disruption Forms | 11 Disruption | D | D | disruption |

4.3.2 Mapping with DAMSL, SWBD and other tagsets

We use here mainly the MATE synthesis of dialogue acts, from MATE Deliverable 1.1, but the final MATE proposal (which does not however propose a new set of dialogue acts) is also useful.

- Marion Klein & Claudia Soria 1998, Dialogue Acts, in *MATE Deliverable 1.1: Supported Coding Schemes*, ed. Marion Klein, Niels Ole Bernsen, Sarah Davies, Laila Dybkjær, Juanma Garrido, Henrik Kasch, Andreas Mengel, Vito Pirrelli, Massimo Poesio, Silvia Quazza and Claudia Soria, MATE (Multilevel Annotation, Tools Engineering) European Project LE4-8370.
<http://mate.nis.sdu.dk/about/D1.1/> or <http://www.dfki.de/mate/d11/chap4.html>
- Andreas Mengel, Laila Dybkjaer, J.M. Garrido, Ulrich Heid, Marion Klein, V. Pirrelli, Massimo Poesio, S. Quazza, A. Schiffrin & Claudia Soria 2002, MATE Dialogue Annotation Guidelines, Deliverable MATE Telematics Project LE4-8370, D2.1, 8 January 2000.
<http://www.andreasmengel.de/pubs/mdag.pdf> or <http://www.ims.uni-stuttgart.de/projekte/mate/mdag/>

As explained above, the following table provides only a loose correspondence, since it does not define the meaning of tags, and it does not represent multidimensionality issues.

| MALTUS | MALTUS gloss | DAMSL | SWBD-DAMSL | Traum's | Chat |
|--|--------------|--|---|--------------------------------|---|
| - | - | <i>Forward looking function</i> | <i>Forward Communicative - Function</i> | <i>Illocutionary Function</i> | <i>Categories of Illocutionary Force</i> |
| S | Statement | Statement Assert Reassert Other | Statement Statement-no-opinion Statement-opinion | Inform Supp-Inf Supp-Sug | Statement: AC, CN, DW, ST, WS Declarations: DC, DP |
| Q Q Q Q Q AT AT - | Question | Info-Request | <i>Influencing-Addressee-Future-Action (1)</i> Yes-No-Question Wh-Question Or-Clause Declarative-Yes-No-Question Declarative-Wh-Question Tag-Question Backchannel-in-Question Rhetorical-Question | YNO WHQ | <i>Questions:</i> AQ, AA, AN, EQ, NA, QA, QN, RA, SA, TA, TQ, YQ, RQ |
| DO | Action > | <i>Influencing-</i> | <i>Influencing-Addressee-</i> | Request | <i>Directives (1):</i> |

| | | | | | |
|---|---|--|--|---------------------------------------|--|
| | <i>suggestion</i> | <i>Addressee-Future-Action</i> Action-directive Open-Option | <i>Future-Action (2)</i> Open-Question Action-Directive | Suggest | RP, RQ |
| DO | Action > <i>commitment</i> Action > performative Exclamation | <i>Committing-Speaker-Future-Action</i> Offer Commit Explicit-performative Exclamation | <i>Committing-Speaker-Future-Action</i> Offers Options Commits | Offer | <i>Commitments:</i> FP, PF, SI, TD <i>Directives (2):</i> CL, SS |
| | | - | - | Promise | PD |
| | | <i>Backward looking function</i> | <i>Backwards-Communicative-Function</i> | - | - |
| RP RN RP RN RU - | Positive answer Negative answer Positive answer Negative answer Other answer | Answer | <i>Answer</i> Yes Answer No Answer Affirmative non-yes-answer Negative non-no answer Other answer Dispreferred answers | Eval | <i>Evaluations:</i> AB, CR, DS, ED, ET, PM <i>Directives (3):</i> AC |
| RP RP RU RN RN H | Positive answer Positive answer Other answer Negative answer Negative answer Attention | <i>Agreement</i> Accept Accept-part Maybe Reject Reject-part Hold | <i>Agreement</i> Agree/Accept Maybe / Accept-part Reject Hold before answer/agreement | Accept Reject Check | <i>Directives (4):</i> AD, AL, CS, RD, GI, GR, DR <i>Declarations (2):</i> ND, YD |
| | | <i>Understanding</i> | <i>Understanding</i> | <i>Grounding</i> | - |

| | | | | | |
|------------|--------------------------|---|------------------------------------|----------------|---|
| | | - | - | RequestAck | - |
| AT | Attention | <i>Signal- understanding</i> Acknowledge | Response- Acknowledgement | Acknowledge | Speech Elicitations: CX, EA, EI, EC, EX, RT, SC |
| RIR | Restated info repeat | Repeat-rephrase | Repeat-phrase | | |
| -- | -- | Completion | Collaborative- Completion | | |
| AT | Attention | | Acknowledge | | |
| RIR | Restated info repeat | | Summarize/Re- formulate | | |
| PO | Politeness | | Appreciation | | |
| PO | Politeness | | Downplayer | | |
| AT | Attention > check | Signal-Non- Understanding | Signal-non- understanding | Request-Repair | <i>Demands for clarification:</i> RR |
| RIC | Restated info correct | Correct- Misspeaking | | Repair | <i>Text editing:</i> CT |
| | | - | <i>Other-forward- function</i> | Greet | - |
| | ? | | Conventional-opening | Apologise | |
| | ? | | Conventional-closing | | |
| PO | Politeness | | Thanking | | |
| PO | Politeness | | Apology | | |
| | | - | - | - | - |
| | | - | <i>Other</i> | - | <i>Vocalisation:</i> YY, OO |
| | | | Quotation | | |
| | | | Hedge | | |
| | | - | - | - | <i>Markings:</i> CM, EM, EN, ES, MK, TO, XA |
| | | - | - | - | <i>Performances:</i> PR, TX |

Chapter 5. Towards a multidimensional dialog act tagset based on pragmatic theories: PRIMULA

We propose here the draft of a multidimensional tagset for dialog acts. Our goal is to annotate the *function* of utterances with respect to dialogue, based on existing theoretical analyses of these functions. Therefore, we should not annotate as “dialog acts” semantic or phonetic properties of the utterances. It is quite often the case that an utterance has a *function* in several dimensions or planes: for instance, turn management is not only a matter of pure backchannels or floor holders, but every utterance plays in fact a role in turn management. Therefore, when looking for the “function” of an utterance, several dimensions should be explored and annotated for an utterance, as in the DAMSL tagset. If it appears that an utterance has virtually no function in one of the dimensions, then a null sign should be somehow annotated for that dimension or plane. A more constraining point of view, akin to SWBD, would be to annotate only the most salient function of an utterance, i.e. only one tag from one dimension. This would reduce the number of possible combinations of tags for one utterance, but would also contain less information about the dialog.

We propose therefore an approach and a tagset, which can be defined as “Principled Multifunctional Annotation” of utterances in dialog, or in short PRIMULA⁴.

The PRIMULA tagset for utterances in dialog features the following dimensions:

1. Speech acts
2. Turn management
3. Adjacency pairs
4. Overall organization and topics
5. Politeness management
6. Rhetorical role



Each of these dimensions is based on a particular theory, which describes the possible list of functions or roles that can be assigned to an utterance according to each dimension. We have already discussed some of these theories in Chapter 1.

1. **Speech acts** are defined in the speech act theory (Austin, Searle, and others). Although there are criticisms directed to this theory as an attempt to describe “the” function of utterances, it is nevertheless accepted that speech acts grasp an important aspect of utterances. What has been criticized in particular is the strict definition of speech acts in terms of pre- and post-conditions (Levinson 1983).
2. **Turn management** is better thought of as a series of findings from conversation analysis (mainly Schegloff). These analysts have attempt to model the processes that govern turn-taking (and turn-giving, turn-holding, etc.), and classify the functions utterances may serve.
3. **Adjacency pairs** are links between a first-part and a second-part utterance such as question and answer (see Chapter 1). Although the definition of “question” and “answer” in such terms is not

⁴ *Primula* is in fact a name of a flower and of a botanical genus, more commonly known as the Primrose (*Primevère* in French). See image above. *Primula vulgaris* is the English Primrose. The name is related to the Latin word for spring.

completely obvious, adjacency pairs grasp important functions of utterances that cannot be modelled through speech act theory (such as the concept of “answer”).

4. **Overall organization and topics** — again in conversation analysis, it has been established that conversations have conventional openings and closings, and are often structured as a series of topics. An utterance can have a function in this dimension if it serves as an opening or closing, or serves to introduce or to close a topic.
5. **Politeness management** — politeness marks appear of course in almost every utterance. One would say that an utterance always serves some kind of function in this dimension. A theory that could help to specify this function is the face-threatening / face-preserving model put forward by Brown and Levinson (1984). In this framework (to be developed), an utterance could serve a face-threatening or a face-preserving role. From a less specific point of view, an utterance could be simply marked as “politeness” if it serves *mainly* or *obviously* a politeness function.
6. **Rhetorical role** — Rhetorical Structure Theory, or other theories of rhetorical relations, could be used to assign to each utterance a rhetorical role. Such theories are less often applied to dialogue, and it is easier to assign rhetorical roles in a text than in a dialog.

Here are the possible tags, i.e. concrete functions of an utterance, according to each of the previous dimensions. Terminal tags are in bold. Sometimes, for clarity, these functions (or tags) are grouped into several non-terminal categories or classes.

1. Speech acts (from Searle 1967)
 - representatives: **assert, conclude**, ...
 - directives: **suggest, request, question**
 - commissives: **promise, threaten, offer**, ...
 - expressives: **thanks, welcome, apologize, congratulate**, ...
 - declaratives (in speech-act sense): **excommunicate, christen, declare war**, etc. (or, as a class: **explicit-performatives**);
 - **nothing** (it is not clear whether an utterance can have no speech act role, but this option should in any case be left open).
2. Turn management: **backchannel, floor holder, floor grabber, hold, nothing**.
3. Adjacency pairs—note that there is some overlapping in the following terms with those from speech acts; they should not be confused, and should received different symbols when used for annotation. A possible list could be:
 - **request / accept, refuse**
 - **offer, invite / accept, refuse**
 - **assessment / agreement, disagreement**
 - **question / answer, non-answer** (or **unexpected answer**)
 - **blame / denial, admission**A simpler tag set for this dimension could be: **first-part / second part / both / none** or, otherwise expressed, **forward-looking / backward-looking / both / none**.
4. Overall organization and topics: **opening, closing, change-topic** (better: **start-topic, end-topic**), **continue-topic** (default), **none**.

5. Politeness management: the simple version has only **politeness** / *none*. But a more complex tag set based on Brown and Levinson could be: **face-threatening** / **face-saving** / *none*.
6. Rhetorical role: many sets of rhetorical roles have been defined, some with a lot of possible functions. But most of them regard “texts”. Also, in RST (Mann & Thompson) for instance, rhetorical relations are relations between the nucleus and the satellites, and it is the relations that bear names, not the utterances (or sentences). It would be a little abusive to transfer the name of the RST relation to the satellite utterance only. Here are some of the relations that were defined, according to the original RST proposal (21 + 3 types): **elaboration, circumstance, solutionhood, volitional cause, non-volitional cause, volitional result, non-volitional result, purpose, condition, otherwise, interpretation, evaluation, restatement, summary, evidence, antithesis, concession, motivation, enablement, justify, background** (plus possibly: **contrast, joint, sequence**). In dialogue however, these relations seem to be restricted to relations between utterances within the *same* turn.

The present proposal, PRIMULA, should of course be refined, for instance using the following course of action:

- the theoretical bases of each dimension should be stated more clearly, as well as the list of tags in each dimension;
- an experiment in annotation using the PRIMULA scheme should prove that the inter-annotator agreement is high (until now, *kappa* was about 80% using the previous tagsets; it should be at least 90% for highly reliable conclusions);
- conversion procedures should be provided from other tagsets, so that existing resources can be reused;
- an experiment in automatic detection of dialogue acts with the PRIMULA tagset could be useful too;
- finally, it would be interesting to show that the PRIMULA tagset is relevant to the meeting recording application targeted in IM2.MDM.

Appendix A – Notes on the 25-10-2003 release of the ICSI-DA corpus

The following observations originate in an attempt to derive some statistics from the dialog acts that occur in the ICSI-DA corpus (released October 25, 2003) and to convert the ICSI-MR dialog acts to the MALTUS set. Although the anomalous dialog act annotations appeared gradually through our analysis, we summarize them at the beginning (section 1). In the following sections, we remove the incoherent annotations from the various figures that we give, rather than correct them, and compute some characteristic figures. We hope the anomalous dialog acts will be corrected in a future release of the corpus.

As an initial estimate, the 50 dialogs (meetings) in this release total 64282 “utterances”, or rather entries in the DADB format, since some of them are clearly made of two utterances, separated by a vertical bar. Overall, the correctness of the annotations in such a large database is remarkable, as are the documentation and the statistics that accompany the release. Still, some points should be revised, at least in our view.

A.1. Incoherent annotations of the dialog acts

The following anomalies appear here more or less in the order in which they were detected.

A.1.1. Lines without labels

There are 72 lines (in the whole set of 50 DADB files) that have no dialog act label at the corresponding place in the comma-separated format. Most of them are “bleeped” lines, but not all of them, as shown here.

One line has a space instead of a label:

```
../icsi_da_corpus_251003/data/Bed006.dadb:864.261,864.581,A,864.261+864.581+right, ,Bed006-c0, , ,FBH,right .,right .
```

The 71 others have no space between the commas in place of a dialog act. They are detected with the following UNIX command:

```
> egrep '^.*,.*,.*,.*,.*,.*,.*,.*,.*,.*,.*$' ../icsi_da_corpus_251003/data/*
```

The lines that have not been bleeped and that should have a DA label (at least ‘z’, unmarked) are the following (total 16):

```
../icsi_da_corpus_251003/data/Bed003.dadb:1714.33,1715.72,A,1714.33+1714.66+but|1714.66+1714.96+uh|1715.12+1715.35+it's|1715.35+1715.72+free,,Bed003-c3,,,but uh - it's free .,but uh - it's free .
../icsi_da_corpus_251003/data/Bed006.dadb:3725.89,3726.35,A,3725.89+3726.35+o_k,,Bed006-c4,,b,,okay .,o_k .
../icsi_da_corpus_251003/data/Bmr010.dadb:1379.99,1380.35,A,1379.99+1380.35+uh-huh,,Bmr010-c8,,,FBH,uhhuh .,uh-huh .
../icsi_da_corpus_251003/data/Bmr012.dadb:481.348,484.438,A,481.348+481.648+because|481.648+481.748+it|481.748+482.048+sounded|482.048+482.348+like|482.348+482.548+even|482.548+482.608+the|482.608+482.808+ones|482.808+482.868+it|482.868+483.128+got|483.128+483.488+wrong|483.488+483.598+it|483.598+483.808+sort|483.808+483.868+of|483.868+484.038+got|484.038+484.098+it|484.098+484.438+right,,Bmr012-c4,,,because it sounded like even the ones it got wrong it sort of got it right .,because it sounded like even the ones it got wrong it sort of got it right .
../icsi_da_corpus_251003/data/Bmr014.dadb:2537.37,2537.6,A,2537.37+2537.6+yeah,,Bmr014-c1,,,yeah .,yeah .
../icsi_da_corpus_251003/data/Bmr023.dadb:3002.44,3004.43,A,3002.44+3002.71+and|3003.48+3003.76+and|3003.76+3003.86+and|3003.86+3003.92+it|3003.92+3004.09+helped|3004.09+3004.14+a|3004.14+3004.43+lot,,Bmr023-c5,,,and - and - and it helped a lot .,and - and - and it helped a lot .
../icsi_da_corpus_251003/data/Bmr024.dadb:2941.78,2942.12,A,2941.78+2942.12+o_k,,Bmr024-cB,,,okay .,o_k .
../icsi_da_corpus_251003/data/Bmr031.dadb:15.39,16.25,A,15.39+15.62+close|15.62+15.68+the|15.68+15.82+door|15.82+16.16+behind|16.16+16.25+him,,Bmr031-c5,,3b+, ,close the door behind him .,close the door behind him .
../icsi_da_corpus_251003/data/Bro003.dadb:5208.72,5211.72,Y,5208.72+5208.84+right|5208.84+5209.04+but|5209.04+5209.25+during|5209.25+5209.31+the|5209.31+5209.81+training|5210.18+5210.4+we|5210.4+5210.
```

```

49+would|5210.49+5211.06+train|5211.26+5211.72+on|XXXX+XXXX+{sigmoid}|XXXX+XXXX+{x},,Bro003-
c2,,,right but during the training we would train on sigmoid x .,right but during the training we
would train on {sigmoid} {x} .
./icsi_da_corpus_251003/data/Bro003.dadb:5212.6,5217.39,A,5211.72+5212.6+<sigmoid-
x_>|5212.6+5212.82+and|5212.82+5213.03+then|5213.78+5213.97+at|5213.97+5214.06+the|5214.06+5214.35+e
nd|5214.35+5214.61+just|5214.61+5214.84+chop|5214.84+5215.15+off|5215.15+5215.34+the|5216.12+5216.47
+final|5216.47+5217.39+nonlinearity,,Bro003-c2,,,and then at the end just chop off the final
nonlinearity .,<sigmoid-x_> and then at the end just chop off the final nonlinearity .
./icsi_da_corpus_251003/data/Bro005.dadb:3776.46,3780.02,A,3776.46+3776.57+if|3776.57+3776.83+peopl
e|3776.83+3776.93+are|3776.93+3777.3+interested|3777.3+3777.53+in|3777.53+3777.78+in|3777.78+3778.01
+getting|3778.01+3778.28+jobs|3778.28+3778.73+running|3778.73+3778.94+on|3778.94+3779.11+that|3779.1
1+3779.34+maybe|3779.34+3779.38+i|3779.38+3779.53+could|3779.53+3779.73+help|3779.73+3779.84+with|37
79.84+3780.02+that,,Bro005-cB,,,if people are interested in - in getting jobs running on that maybe
i could help with that .,if people are interested in - in getting jobs running on that maybe i could
help with that .
./icsi_da_corpus_251003/data/Bro012.dadb:3578.76,3579.06,A,3578.76+3579.06+mm-hmm,,Bro012-
c3,,,uhhuh .,mm-hmm .
./icsi_da_corpus_251003/data/Bro013.dadb:2045.53,2045.78,A,2045.53+2045.78+o_k,,Bro013-
c3,,,okay .,o_k .
./icsi_da_corpus_251003/data/Bro022.dadb:76.57,76.85,A,76.57+76.85+o_k,,Bro022-c0,,10b,,okay .,o_k .
./icsi_da_corpus_251003/data/Bro022.dadb:894.876,895.296,A,894.876+895.296+mm-hmm,,Bro022-
c4,,b,,uhhuh .,mm-hmm .
./icsi_da_corpus_251003/data/Bro022.dadb:1159.03,1161.75,A,1159.03+1159.19+but|1159.19+1159.58+that
|1159.58+1159.98+um|1161.07+1161.3+didn't|1161.3+1161.5+work|1161.59+1161.75+either,,Bro022-
c4,,40a,,but that um didn't work either .,but that um didn't work either .

```

A.1.2. Occurrences of disruption marks

Although most occurrences are coherent with the guidelines, the following cases must be noted:

| | | |
|-------|---------|-------------|
| %--.% | appears | once |
| %.%-- | appears | seven times |
| %-.% | appears | three times |
| %.%- | appears | five times |

It seems that the first two and the last two labels have the same meaning, that is, indecipherable because abandoned, indecipherable because interrupted. It seems that the correct order has ‘%’ in the first place.

It was found also that ‘x.%’ occurs twice in the whole corpus. Normally, ‘x = non-speech’ implies that the utterance is also indecipherable, so maybe ‘x.%’ should not be accepted.

A.1.3. Reported speech

There are 521 occurrences of dialog act labels that contain a colon ‘:’ for reported speech. It is not always clear that they really separate two utterances of which the second is reported speech. But the main problem is that unlike the vertical bar ‘|’, the colon is not marked on the transcription, for no obvious reason. It could be useful to mark the beginning of the reported speech on the transcription too.

A.1.4. Blank spaces in DADB files

Although there is no obligation, the general rule is that the DA labels are comprised between two commas in the DADB file with no extra space before or after the label. There are very few exceptions to this rule, probably less than seven. They must be taken into account when processing the data, since ‘fh ’ is not identical to ‘fh’. Here are four that we identified:

```

Bro005.dadb:1518.4,1519.22,A,1518.4+1518.63+so|1518.63+1519.22+um,fh ,Bro005-c4,fh ,,,so um ==,so um
==
Bro005.dadb:2197.03,2203.78,M2,2183.87+2184.47+<mm-hmm>|2197.03+2197.47+uh|2198.17+2198.23+<mmm>|219
8.23+2198.39+<uh>|2198.39+2198.54+i|2198.54+2198.73+mean|2198.73+2198.86+<that>|2198.86+2198.96+the|
2198.96+2199.18+the|2199.18+2199.43+fact|2199.43+2199.63+that|2199.63+2199.73+<s->|2199.73+2199.95+w
ell|2199.95+2200.16+for|2200.34+2200.78+for|2201.02+2201.66+t_i-digits|2202.1+2202.47+the|2202.55+22
02.88+timit|2202.88+2203.09+net|2203.2+2203.35+is|2203.35+2203.45+the|2203.45+2203.78+best,h|s ,Bro0
05-c0,h|s ,17b,,uh | i mean the - the fact that well for for t i digits the timit net is the best .,
<mm-hmm> uh | <mmm> <uh> i mean <that> the - the fact that <s-> well for for t_i-digits the timit ne
t is the best .
Bro012.dadb:2349.87,2350.46,A,2349.87+2350.46+and,fh.%-- ,Bro012-c5,fh.%-- ,,and ==,and ==

```

Bro018.dadb:2268.65,2270.84,A,2268.65+2268.81+yeah|2268.81+2268.96+see|2268.96+2269.23+one|2269.23+2269.51+thing|2269.51+2269.73+that's|2269.73+2269.76+a|2269.76+2270.02+little|2270.02+2270.41+bit|2270.41+2270.84+um,fg|s.%-- ,Bro018-c5,fg|s.%-- ,,,yeah | see one thing that's a little bit um ==,yeah | see one thing that's a little bit um ==

A.1.5. Incoherent combinations of tags

The following occurrences were detected:

- fh.s in Bmr026 (use of the dot)
- 36a+ in Bro003 (AP indication in place of DA label)
- qy^^d^f^g in Bro005 (two carets)
- qy^^rt in Bmr014
- s^^ba^rt in Bed004
- s^^ba^rt^tc in Bmr013
- s^df? in Bmr013 (question mark)
- w.%- in Bro018 (there is no 'w' tag in ICSI-MR; maybe qw?)
- qh^s in Bmr023 (two general tags)
- qy^h^... in Bed003 (two general tags followed by something else, noted here with '...')
- s^h^t1 in Bed003 (two general tags) – appears *twice* in Bed003
- s^fe^h in Bmr001 (two general tags)
- s^h^m.%-- in Bro012
- s^bsc^fg in Bro012
- qy^b^bu^rt in Bro008
- s^h^t1 in Bro005
- s^fg^tc in Bro018

A.1.6. A tag that never occurs: fw

This tag means “you’re welcome” but it is never used in the 50 meetings. A way to check this:

```
$ egrep 'fw,' *
[nothing]
$ egrep 'fw' * | wc -l
6
$ egrep 'fw' * | egrep -v 'halfway'
[nothing]
```

A.2. Figures regarding the ICSI-MR labels

If we discard the DADB lines with no DA label at all, there are **64210** lines with labels. All in all, there are 1650 different labels (or combinations of labels with ‘|’ and ‘:’). The most frequent ones are s = 14678 times, b = 8721 times, fh = 3448 times, s^bk = 2890 times, etc. There are 868 combinations among the 1650 that occur only once.

A.2.1. Vertical bars and colons

There are 4027 utterances that are labelled using one or two vertical bars (only 8 have two bars). In our view, these should be separated into two (resp. three) utterances, each with its own DA label.

The same applies for reported speech (‘:’), of which there are 521 occurrences. These should be also separated into two utterances (but this is not feasible yet since the colons are not inserted in the transcriptions of the utterances in the present state).

For our tests and counts, we separate dialog act labels that have a ‘|’ or a ‘:’ in two parts, and obtain a list of **68766** labels that no longer contain separators. There are now only **963** different types of labels.

A.2.2. Disruptions

There are 49 labels with two ‘%’, most of them with reported speech (‘:’). There are 241 labels with an ‘x’ (non speech), the most frequent being ‘s.x’ 70 times.

Regarding disruption marks with a ‘%’, there are **8872** labels that contain such marks. The cases when the disruption marks are alone are: %- = 1288 times; %-- = 751 times; % = 127 times (remember also the 16 combinations of %, %- and %-- described in section A.1.2). Note also that ‘x’ appears 241 times, of which 58 times alone, and ‘z’ appears 1331 times, always alone (and remember the ‘x.%’, twice).

To have a better view of the variety of occurring labels *without* the disruption marks, we perform the following operations:

- we remove from our list tokens that consist only in disruption marks (%-, %--, %, x, z, %.%--, %.%-);
- we delete from the other labels the additional disruption marks (., %-, .%--, .x);
- we finally check that the predicted number of removed labels (3573) is indeed removed.

The result is 65193 occurrences of labels without separators or disruption marks.

After some corrections (related to the list in A.1.5), we are left with **65188** occurrences, with **685** different types of labels.

A.2.3. Number of tags in the labels

In the remaining 65188 labels, there are 97088 occurrences of tags. Remember that there are 11 general tags [1st tier] and 39 specific tags [2nd tier] since ‘fw’ is never used. The numbers of occurrences are:

| | | | | | | | |
|---------|--------|---------|---------|---------|---------|---------|---------|
| 42886 s | 9152 b | 5145 fh | 4551 rt | 3748 bk | 3437 qy | 3287 aa | 2043 cs |
| 1943 fg | 1925 d | 1837 e | 1725 df | 1419 bu | 1289 qw | 1287 ba | 959 g |
| 783 na | 581 co | 579 h | 578 no | 571 tc | 550 ar | 539 j | 533 2 |
| 486 m | 449 t | 403 nd | 377 r | 366 f | 340 cc | 332 fe | 301 ng |
| 236 qrr | 233 am | 224 bd | 214 t3 | 201 qh | 194 t1 | 173 qr | 164 aap |
| 159 br | 158 fa | 156 qo | 136 bs | 135 arp | 118 bsc | 73 bh | 52 ft |
| 48 bc | 11 by | | | | | | |

The total here is 97088.

If we are interested only in the list of general tags (whether they are or not followed by specific tags), of which there are 11, the figures are the following. The numbers sum up here to 65188, since each label has one and only one general tag.

| | | |
|---------|---------|---------|
| 42886 s | 9152 b | 5145 fh |
| 3437 qy | 1941 fg | 1289 qw |
| 573 h | 236 qrr | 201 qh |
| 173 qr | 156 qo | |

Remember there are 685 different types of labels. If we look only at the combinations of specific tags following the general tag, we are still left with 466 different combinations of specific tags.

Going back to the 685 occurring combinations of tags, here are their frequencies:

| Nb. of tags in label | Number of types | Number of tokens |
|----------------------|-----------------|------------------|
| 1 | 11 | 39694 |
| 2 | 135 | 20884 |

| | | |
|--------|-----|-------|
| 3 | 361 | 3075 |
| 4 | 131 | 1283 |
| 5 | 42 | 245 |
| 6 | 5 | 7 |
| 7 | 0 | 0 |
| Total: | 685 | 65188 |

One of the most obvious conclusions from these figures is that a program aimed at automatically annotating the DA labels (using the ICSI-MR set and the present data) should probably not be allowed to hypothesize more than two or three tags per label, in order to reduce the search space without losing too much accuracy, as the following table shows.

| | | | | | | |
|--|------|------|------|-------|------|---|
| Maximal number of tags hypothesized for a label | 1 | 2 | 3 | 4 | 5 | 6 |
| Maximal accuracy reachable on the present data set | 0.61 | 0.93 | 0.98 | 0.949 | .996 | 1 |

In fact, the maximal accuracy could even be higher if partial matches between labels also count as correct answers. For instance, one could count a partial match when the system proposes 's^ba' for an utterance that is annotated s^ba^rt (the score of this match could be 67%).

Remember also that we do not consider in the figures above the disruption marks and the separators.

A.2.4. Theoretical combinations of ICSI-MR tags into labels

In this section, we compute the number of possible ICSI-MR labels (combinations of tags). We consider only at the 11 general tags and the 39 specific tags (since for some reason 'fw' is not used), without the disruption marks and the separators. We thus compute the size of the search space for a program trying to assign DA labels automatically – that is, the number of possible combinations of tags, as follows. Note that the order of the specific tags (the former 2nd tier) is not relevant, and they could be for instance ordered alphabetically by convention.

- for the general tag, there are 11 possibilities (s, b, h, fg, fh, qo, qr, qrr, qw, qy, qh)
- for the specific tag:
 - if no specific tag is used, 1 possibility
 - if exactly one specific tag is used, 39 possibilities
 - if exactly two specific tags are used, $C(2,39)=741$ possibilities
 - if exactly three specific tags are used, $C(3,39)=9139$ possibilities
 - if exactly four specific tags are used, $C(4,39)=82251$ possibilities
 - if exactly five specific tags are used, $C(5,39)=575757$ possibilities

Therefore, if combinations of tags are no further constrained by the annotation scheme, as it is currently the case in the ICSI specification, then the theoretical number of combinations of ICSI-MR tags into labels is:

| | | | | | | |
|-----------------------------------|----------|-----------|-----------|-----------|-----------|-----------|
| Maximal number of tags in a label | 1 (gen.) | 2 (1g+1s) | 3 (1g+2s) | 4 (1g+3s) | 5 (1g+4s) | 6 (1g+5s) |
| Number of possible combinations | 11 | 440 | 8'591 | 109'120 | 1'013'881 | 7'347'208 |

Again, the conclusion is that allowing only two tags per label in an automated annotation system (one general tag out of 11 and one specific tag out of 39) reduces considerably the search space, to only 440 possible tags, while the maximal accuracy is 93% as shows above.

A.3. Conversion to MALTUS labels

In this section, we describe very briefly the conversion of the 65188 ICSI-MR labels to the MALTUS tagset. A similar study should be done also using the class maps provided by ICSI in the last release.

We use the table given in section 4.3.1 to write a ‘sed’ conversion script, followed by a ‘sed’ cleaning script that reorders the specific tags alphabetically within each label, and discards duplicates. This is especially important since in the conversion process, the same MALTUS tag may appear twice in a label.

There are 51 resulting types of tags, shown below, ordered by number of occurrences:

| | | |
|-------------|------------|--------------|
| 28365 S | 31 S^RI^RN | 2 S^AT^RP |
| 9137 B | 30 S^DO^RN | 2 S^AT^PO |
| 7646 H | 27 S^PO^RU | 2 Q^RP |
| 4242 S^AT | 21 S^DO^RI | 2 Q^DO^RN |
| 3945 S^RP | 15 S^RI^RU | 2 Q^AT^RN |
| 3309 Q | 15 B^RI | 1 S^RP^RU |
| 2713 S^DO | 13 Q^RN | 1 S^DO^RI^RP |
| 1928 Q^AT | 11 S^DO^RU | 1 S^DO^PO^RU |
| 1297 S^RN | 11 H^RI | 1 S^AT^RU |
| 746 S^RU | 5 S^PO^RI | 1 Q^RI^RU |
| 543 S^RI | 5 S^AT^RN | 1 Q^AT^RU |
| 401 S^PO | 4 S^AT^DO | 1 Q^AT^RI^RP |
| 246 S^RI^RP | 3 S^RN^RU | 1 Q^AT^PO |
| 188 S^AT^RI | 3 S^PO^RN | 1 H^RP |
| 141 Q^DO | 3 Q^RU | 1 H^AT |
| 46 Q^AT^RI | 3 Q^AT^DO | |
| 38 Q^RI | 2 S^PO^RP | |
| 33 S^DO^RP | 2 S^DO^PO | |

Another possible display of the same data is to sort the tags alphabetically:

| | | |
|--------------|--------------|-------------|
| 9137 B | 13 Q^RN | 11 S^DO^RU |
| 15 B^RI | 2 Q^RP | 401 S^PO |
| 7646 H | 3 Q^RU | 5 S^PO^RI |
| 1 H^AT | 28365 S | 3 S^PO^RN |
| 11 H^RI | 4242 S^AT | 2 S^PO^RP |
| 1 H^RP | 4 S^AT^DO | 27 S^PO^RU |
| 3309 Q | 2 S^AT^PO | 543 S^RI |
| 1928 Q^AT | 188 S^AT^RI | 31 S^RI^RN |
| 3 Q^AT^DO | 5 S^AT^RN | 246 S^RI^RP |
| 1 Q^AT^PO | 2 S^AT^RP | 15 S^RI^RU |
| 46 Q^AT^RI | 1 S^AT^RU | 1297 S^RN |
| 1 Q^AT^RI^RP | 2713 S^DO | 3 S^RN^RU |
| 2 Q^AT^RN | 2 S^DO^PO | 3945 S^RP |
| 1 Q^AT^RU | 1 S^DO^PO^RU | 1 S^RP^RU |
| 141 Q^DO | 21 S^DO^RI | 746 S^RU |
| 2 Q^DO^RN | 1 S^DO^RI^RP | |
| 38 Q^RI | 30 S^DO^RN | |
| 1 Q^RI^RU | 33 S^DO^RP | |

It appears that some of the 51 tags are quite infrequent, especially for instance the combinations of four tags. Also, although RP, RN, RU (responses) are mutually exclusive, they sometimes co-occur in the above tables (1 S^RP^RU, 3 S^RN^RU). It should be checked whether this comes from incoherencies in the ICSI-MR annotation or from a problem in the conversion process (we favour the first hypothesis...).

In section 4.1, we estimated that the number of possible MALTUS labels is 770 (compare this to the numbers of ICSI-MR combinations in section A.2.4). With the present conversion script, the MALTUS labels are made of the following tags:

- one of the four general tags S, Q, B, H
- one or more specific tags: AT, RI, DO, PO and/or either of RN, RP, RU

So the number of MALTUS tags here is $4 \times 2 \times 2 \times 2 \times 4 \times 2 = 256$. The difference with 770 (section 4.1) comes from three points: we do not consider here U (unintelligible) and D (disrupted) tags, and RI is not divided in RIP and RIC.

For MALTUS too, the idea to use at most three tags per label in an automatic annotation program might reduce the search space without decreasing the accuracy too much. Another idea is to use only the labels that appear above, that is, only 51 labels, or even better, the labels that occur above more than ten times (for instance). It is necessary however either to check the data to see if the other labels are exceptions, or to ignore the other labels simply by noting that they account only for a total of 50 occurrences, that is 0.077% of the total number of labels (occurrences). There are only 27 labels that appear more than ten times, which is a considerable reduction of the search space that should be tested empirically to make sure that no “important” tags have been left out.

The 27 MALTUS labels that occur more than ten times are, arranged alphabetically:

```

9137 B
      15 B^RI
7646 H
      11 H^RI
3309 Q
      1928 Q^AT
           46 Q^AT^RI
      141 Q^DO
      38 Q^RI
      13 Q^RN
28365 S
      4242 S^AT
           188 S^AT^RI
      2713 S^DO
           21 S^DO^RI          30 S^DO^RN          33 S^DO^RP   11 S^DO^RU
      401 S^PO
           27 S^PO^RU
      543 S^RI
           31 S^RI^RN          246 S^RI^RP          15 S^RI^RU
      1297 S^RN
      3945 S^RP
      746 S^RU

```

Further analysis should tell us whether this list should be enriched with useful labels that are absent from it. Also, a comparison of MALTUS to the five ICSI-MR class maps, as well as a similar analysis for each class map, would bring useful insights, in the near future.

Appendix B – Notes on the 11-02-2004 release of the ICSI-MRDA corpus

The following observations originate in an attempt to derive some statistics from the dialog acts that occur in the ICSI-DA corpus release of February 11, 2004, which is substantially the final release.

B.1 Statistics from the analysis of ICSI-MRDA

In this release, there are 75 meetings and 112,032 utterances (with and without '|')⁵. There are 2'054 different labels (types). Of the 112.032 occurrences of labels, 533 are empty and 2,442 are 'z'. The labels that appear more than 1,000 times each are:

| | | | |
|-----------|------------|-----------|-----------|
| 25684 s | 14466 b | 6160 fh | 5673 s^bk |
| 4626 s^aa | 4190 s.%-- | 2927 s^df | 2628 s^e |
| 2442 z | 2212 s.%- | 2153 %- | 1810 s^ba |
| 1710 s^cs | 1420 fg | 1318 s^rt | 1194 %-- |
| 1173 fh s | | | |

On the contrary, there are also 998 labels which appear only once!

Labels with ‘%’ (disruptions): there are 66 labels with two ‘%’. All are coherent since they all represent reported speech, being formed of two labels separated by ‘:’. There were 16 occurrences which weren't fully coherent: %--.% and %.%--, respectively %-.% and %.%-. All have been corrected, so there are now 8 times %.%- and 8 times %.%-- (unintelligible + aborted/interrupted). There are no ‘x.%’ in the final release.

Labels with ‘x’ (non speech noise): 326 labels contain ‘x’, the most frequent ones being ‘s.x’ (92 times), ‘x’ (90), ‘b.x’ (41), ‘fh|s.x’ (13), etc.

Labels with ‘.’: 11205 occurrences.

Composed labels (with '|'): one has two bars, and 7,192 have two bars (meaning the utterance was split in two, according to the two functions). We now remove the bars, splitting the labels into two, and talk from now on about elementary labels (no '|'). **There are now 119,226 labels** (112,032 + 7,192 + 2).

Labels with ‘:’ (reported speech): 1,512 labels. None has two colons or more. We split these two, and find **120,738 occurrences of labels**, out of which 533 are empty (non-labelled utterances)⁶. Removing empty ones leads to **120,205 labels**.

There are now only **964 different labels**, of which those that appear more than 1,000 times are:

| | | | |
|-----------|------------|-----------|-----------|
| 30619 s | 14472 b | 8348 fh | 6770 s^bk |
| 5638 s^aa | 4746 s.%-- | 3123 s^df | 3107 fg |
| 2805 s^e | 2493 s.%- | 2442 z | 2357 %- |
| 2044 s^ba | 1962 s^cs | 1533 s^rt | 1327 s^na |

⁵ These are extracted by the ‘extraire_da’ script into ‘da_list_total’ (see IM2\data\datrans-new\manips_040211).

⁶ File : da_list_nobars_nocolons, renamed da_list_n2.

1293 %--

1031 qy^rt

1000 qw

Disruption marks (marked ‘%’): 12’059 labels contain a disruption mark, among which: 4746 s.%--, 2493 s.%-, 2357 %-, 1293 %--, 511 b.%, 463 %, etc. ‘x’ appears 326 times (109 times ‘s.x’, 90 times ‘x’, etc.), and ‘z’ appears 2,442 times, always alone (unmarked utterance).

We now remove occurrences of ‘%-’, ‘%--’, ‘%’, ‘x’, ‘z’ alone and their combinations among themselves. We are left with **113,560 labels**⁷. There are 776 occurring labels, i.e. combinations of tags. Recall that in SWBD, there were only 220 combinations of tags in about 205,000 utterances. The most frequent ones are:

| | | | |
|-----------|------------|-----------|-----------|
| 38035 s | 15026 b | 8381 fh | 6837 s^bk |
| 5703 s^aa | 3523 s^df | 3114 fg | 3005 s^e |
| 2284 s^cs | 2101 s^ba | 1593 s^rt | 1378 s^na |
| 1248 qw | 1064 qy^rt | | |

We now focus on the relations between the complexity of the labels and their frequencies. This table shows the number of possible labels (combinations of tags), theoretical vs. actual, and their occurrences in the corpus. See the previous annex for an explanation on how this is computed.

| Nb. of tags in label | Nb. of theoretical combinations of labels | Nb. of occurring combinations | Nb. of tokens |
|----------------------|---|-------------------------------|---------------|
| 1 | 11 | 11 | 68,213 |
| 2 | 429 | 129 | 37,889 |
| 3 | 8,151 | 402 | 5,054 |
| 4 | 100,529 | 176 | 2,064 |
| 5 | 904,761 | 49 | 326 |
| 6 | 6,333,327 | 9 | 14 |
| 7 | ... | 0 | 0 |
| Total: | 7,347,208 | 776 | 113,560 |

Therefore, it is quite easy to compute the maximal accuracy of DA tagging that could be reached on the ICSI-MR data, using a limited number of tags per label (and thus reducing greatly the search space).

| Maximal_nb_of_tags | 1 | 2 | 3 | 4 | 5 | 6 |
|---|-------|-------|-------|-------|-------|---|
| Maximal_theoretical_accuracy_on_ICSI-MR | 0.601 | 0.934 | 0.979 | 0.997 | 0.999 | 1 |

We provide now some statistics about tags (labels are made of one or more tags). There are **169,125 occurrences of tags** once the ‘^’ are removed. Only 50 types occur, since we removed disruption tags (several anomalous tags were found and corrected: a, baa, e_, grt, rt(32b)). Their frequencies are as follows:

| | | | | | |
|---------|---------|---------|---------|---------|---------|
| 77226 s | 15180 b | 8381 fh | 7307 bk | 6823 rt | 6088 aa |
| 5580 qy | 4128 df | 3642 e | 3129 d | 3114 fg | 3051 cs |
| 2464 ba | 2311 bu | 2064 qw | 1753 na | 1607 g | 1034 no |
| 957 ar | 891 j | 865 2 | 830 co | 793 h | 737 f |

⁷ In the process we found an occurrence of ‘fh.s’ in Bmr006, corrected in the final release.

| | | | | | |
|--------|---------|---------|---------|---------|--------|
| 717 m | 701 nd | 680 tc | 611 r | 598 t | 520 fe |
| 511 ng | 444 bd | 440 cc | 407 qh | 399 qrr | 379 am |
| 368 t3 | 292 t1 | 286 fa | 262 aap | 245 br | 225 qr |
| 191 qo | 180 arp | 171 bsc | 164 bs | 160 bh | 139 ft |
| 57 bc | 12 by | 6 fw | | | |

Note that 'fw' never occurred in the previous release of 50 meetings.

Looking now at the **occurrences of the eleven general tags** (tier 1) only, regardless whether they are followed by another tag (total is 113,560, the same as the number of labels above, since each label has exactly one general tag):

| | | | | | |
|---------|---------|---------|---------|---------|---------|
| 77226 s | 15180 b | 8381 fh | 5580 qy | 3114 fg | 2064 qw |
| 793 h | 407 qh | 399 qrr | 225 qr | 191 qo | |

There are 45,347 occurrences of combinations of secondary tags, with 466 different combinations.

B.2 Conversion of ICSI-MRDA to MALTUS

Use icsi2maltus.sed, we get 113,560 resulting MALTUS labels (in 'da_maltus' file). There are only 72 different labels when the order of the tags that compose them is not normalized. When we normalize it, there are only **50 different types**. The result sorted by frequency is:

| | | | | |
|------------|------------|--------------|--------------|------------|
| 51304 S | 15180 B | 12288 H | 8280 S^AT | 7612 S^RP |
| 5320 Q | 3935 S^DO | 3137 Q^AT | 2219 S^RN | 1298 S^RU |
| 791 S^PO | 765 S^RI | 436 S^RI^RP | 273 S^AT^RI | 239 Q^DO |
| 69 Q^AT^RI | 61 S^PO^RU | 60 Q^RI | 46 S^RI^RN | 41 S^DO^RP |
| 38 S^DO^RN | 32 S^DO^RI | 19 Q^RN | 18 S^RI^RU | 16 S^DO^RU |
| 13 S^PO^RI | 8 S^RN^RU | 6 S^PO^RN | 6 S^DO^PO | 6 S^AT^RN |
| 5 S^AT^DO | 5 Q^AT^DO | 4 S^PO^RP | 4 S^AT^RU | 4 Q^RP |
| 3 S^AT^RP | 3 Q^RU | 2 S^AT^PO | 2 Q^DO^RN | 2 Q^AT^RN |
| 1 S^RP^RU | 1 S^RP^AT | 1 S^DO^RI^RP | 1 S^DO^PO^RU | 1 Q^RI^RU |
| 1 Q^RI^RP | 1 Q^PO | 1 Q^AT^RU | 1 Q^AT^RI^RP | 1 Q^AT^PO |

If we look only at the **26 MALTUS labels occurring more than 10 times** (which amounts to neglecting only 0.061% of the occurrences), and if we re-order them according to their label, then we get:

| | |
|---------------|---------------|
| S (51304) | S^PO^RU (61) |
| S^AT (8280) | S^RI (765) |
| S^AT^RI (273) | S^RI^RN (46) |
| S^DO (3935) | S^RI^RP (436) |
| S^DO^RI (32) | S^RI^RU (18) |
| S^DO^RN (38) | S^RN (2219) |
| S^DO^RP (41) | S^RP (7612) |
| S^DO^RU (16) | S^RU (1298) |
| S^PO (791) | |
| S^PO^RI (13) | |

B (15180)
 H (12288)
 H^RI (11)

 Q (5320)
 Q^AT (3137)

Q^AT^RI (69)
 Q^DO (239)
 Q^RI (60)
 Q^RN (19)

A quick comparison between this list and the one in Appendix A, generated from fewer data, shows that the proportions are pretty much the same. The maximal theoretical accuracy using this mono-dimensional tagset of 26 labels is 99.939%.

Following the above analysis and other checks, several changes have been done, documented in the 'doc/preliminary-release-updates.txt' file.

B.3. Update for the final release: 17 March 2004

The minor changes observed above have been done. The release updates from the 'secondary-release-updates.txt' are the following:

Minor DADB file fixes:

 -Blank lines were removed from the beginning of Bed013.dadb, Bed014.dadb, Bro022.dadb, Bro028.dadb, and Btr002.dadb.

-A missing original DA label was inserted the following line in the Bns002.dadb file:

5609.1,5609.57,A,5609.1+5609.57+o_k,,Bns002-cB,s^bk,,,okay .,o_k .

-Five incoherent DA labels were fixed:

| Incoherent Tag | Fixed Tag | Meeting | Channel | Start Time |
|----------------|--------------|---------|---------|------------|
| s^e.%_:s.%-- | s^e.%-:s.%-- | Bed015 | 5 | 3762.39 |
| s^a s^na | s^aa s^na | Bmr016 | 4 | 722.9 |
| s^baa | b | Bmr019 | 5 | 2988.85 |
| qy^rt(32b) | qy^rt | Bmr028 | 0 | 2845.78 |
| s^bu qy^d^grt | s^bu qy^d^rt | Btr002 | 6 | 2169.47 |

References

- Allen James F. and Mark G. Core 1997, DAMSL: Dialog Act Markup in Several Layers (Draft 2.1), Multiparty Discourse Group, Discourse Research Initiative.
- Allwood Jens, Joakim Nivre and Elisabeth Ahlsén 1993, On the Semantics and Pragmatics of Linguistic Feedback, *Journal of Semantics*, **9**, 1, p. 1-26.
- Brown Penelope and Stephen C. Levinson 1987, *Politeness: Some Universals in Language Use*, Cambridge University Press, Cambridge, UK.
- Carletta Jean, Amy Isard, Stephen Isard, Jacqueline C. Kowtko, Gwyneth Doherty-Sneddon and Anne H. Anderson 1997, The Reliability of a Dialogue Structure Coding Scheme, *Computational Linguistics*, **23**, 1, p. 13-32.
- Dhillon Rajdip, Sonali Bhagat, Hannah Carvey and Elizabeth Shriberg 2004, Meeting Recorder Project: Dialog Act Labeling Guide, Technical Report ICSI, TR-04-002.
- Jurafsky Daniel 2003, Pragmatics and Computational Linguistics, *Handbook of Pragmatics*, Blackwell, Oxford, UK.
- Jurafsky Daniel, Elizabeth Shriberg and Debra Biasca 1997, Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation (Coders Manual, Draft 13), Technical Report University of Colorado, Institute of Cognitive Science, 97-02.
- Jurafsky Daniel, Elizabeth Shriberg, Barbara. Fox and Tracy Curl 1998, Lexical, prosodic, and syntactic cues for dialog acts, Proceedings ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers, Montréal, Canada, p. 114-120.
- Levinson Stephen C. 1979, Pragmatics and social deixis, Proceedings Proceedings of the Fifth Annual Meeting of the Berkeley Linguistic Society. Berkeley Linguistics Society, Berkeley, CA, USA, p. 206-223.
- Levinson Stephen C. 1983, *Pragmatics*, Cambridge University Press, Cambridge, MA, USA.
- Levinson Stephen C. 1992, Activity types and language, *Talk at Work: Interaction in Institutional Settings*, Cambridge University Press, Cambridge, MA, USA, p. 66-100.
- Lycan William G. 2000, *Philosophy of Language: a Contemporary Introduction*, Routledge, London, UK.
- Moeschler Jacques 1989, *Modélisation du dialogue : représentation de l'inférence argumentative*, Hermès, Paris.
- Moeschler Jacques 2002, Speech act theory and the analysis of conversations: sequencing and interpretation in pragmatic theory, *Essays in speech act theory*, John Benjamins, Amsterdam, p. 239-261.
- Schegloff Emanuel A. 1988, Presequences and indirection: Applying speech act theory to ordinary conversation, *Journal of Pragmatics*, **12**, 1, p. 55-62.
- Searle 1976 via Levinson 1983
- Shriberg E., R. Dhillon, S. Bhagat, J. Ang and H. Carvey 2004, The ICSI Meeting Recorder Dialog Act (MRDA) Corpus, Proceedings 5th SIGdial Workshop on Discourse and Dialogue, Cambridge, MA, USA, p. 97-100.
- Sinclair J. and M. Coulthard 1975, *Towards an Analysis of Discourse*, Oxford University Press, Oxford, UK.

- Stolcke Andreas, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Marie Meteer and Carol Van Ess-Dykema 2000, Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech, *Computational Linguistics*, **26** 3, p. 339-371.
- Traum David R. 2000, 20 Questions for Dialogue Act Taxonomies, *Journal of Semantics*, **17**, 1, p. 7--30.