Query types in the meeting domain: assessing the role of argumentative structure in answering questions on meeting discussion records

Vincenzo Pallotta[1], Violeta Seretan[2], Marita Ailomaa[3]

We define a new task of question answering on meeting records and assess its difficulty in terms of types of information and retrieval techniques required. The importance of this task is revealed by the increasingly growing interest in the design of sophisticated interfaces for accessing meeting records such as meeting browsers.
We ground our work on the empirical analysis of elicited user queries. We assess what is the type of information sought by the users and perform a user query classification along several semantic dimensions of the meeting content. We found that queries about argumentative processes and outcomes represent the majority among the elicited queries (about 60%). We also assess the difficulty in answering the queries and focus on the requirements of a prospective QA system to successfully deal with them. Our results suggest that standard Information Retrieval and Question Answering alone can only account for less than 20% of the queries and need to be complemented with additional type of information and inference.

1. Introduction

The digital information era and the recent development of information retrieval technologies (ranging from simple keyword search to sophisticated natural language Question Answering) revolutionized the access to information in every sector of human activity. Quick access to relevant information stored in different formats – text, image, audio, video - is nowadays crucial in our time-driven society. In a business context, for instance, the vast amounts of documents stored in corporate memories need to be exploited in order to ensure work efficiency and enable productivity.

Meeting records constitute a particularly important and rich such source of information. Meetings are a frequent and sustained activity, in which multi-party dialogues take place that are goal-oriented and where participants perform a series of actions usually aimed at reaching a common goal: they exchange information, raise issues, express opinions, make proposals, propose solutions, provide arguments (pro or con), negotiate alternatives, make decisions. As an outcome of the meeting, agreements on the future action items are reached, tasks are assigned, conflicts are solved etc. Meeting outcomes have a direct impact on the efficiency of organization and team performance, and the stored and indexed meeting records serve as reference for further processing such as information retrieval, understanding of the decision process, meeting summarization, preparation of future meetings, and planning/control of meetings (Post et al. 2004). They can also be used in future meetings in order to facilitate the decision making process by accessing relevant information from previous meetings (Cremers et al., 2005), or in order to make the discussion more focused (Conklin, 2006). The means to exploit meeting information involve defining a metadata schema for representing the content, then structuring and analyzing the content according to the metadata schema, and finally, accessing the analyzed content via browsing, querying or filtering (Tucker and Whittaker, 2004). Table 1 summarizes this discussion.

This description we proposed applies to debate meetings, while a possible typology of meetings would include other types of meeting in which the deliberative aspect is less essential. This paper focuses exclusively on debate meetings, precisely because of the richness of information which is contained therein concerning the decision making process. Debate meetings are further classified, in

[1] University of Fribourg
[2] University of Geneva
[3] Swiss Federal Institute of Technology - Lausanne

NIST (Cugini, 1997), in staff meetings, information exchange and decision-making meetings, and information gathering and decision-making meetings.

Table 1. Debate Meeting Conceptualization

support	audio video transcription mixed (multimedia)		
content	factual	setting, events, timeline, actions, dynamics	
	thematic	topics, details	
	deliberative	objective	reach a common goal (make decisions, solve conflicts)
		process	make proposals raise issue propose solutions propose alternatives express opinions provide arguments negotiate solutions exchange information
		outcome	decisions made conflicts solved tasks assigned
modeling	semantic model structure analysis (understanding)		
exploitation type	browse query filter		
application	information retrieval (of factual, thematic, *deliberative* content) explaining the rational of outcome intelligent meeting summarization (minutes) facilitation of deliberation in current meeting future meetings preparation/planning/control		

This paper adopts an abstract perspective on meetings along the semantic and pragmatic dimensions, as it focuses on the retrieval of tacit information from meetings,. If we consider that a meeting content can be organized on three levels (cf. Table 1): *factual* level (what happens: events, timeline, actions, dynamics), *thematic* level (what is said: topics discussed and details), and *argumentative* level (how common goals are reached), our endeavor relates to the third and most abstract level. The information on the first two levels is explicit information that can be retrieved directly by searching the meeting records with appropriate indexing techniques (i.e., TF-IDF[4]) depending on the support type (audio, video, text transcription). The third level, on the contrary, contains more abstract, tacit information pertaining to how the explicit information contributes to the rationale of the meeting, and this information is not present as such in meeting records: whether or not the meeting goal was reached; what issues were debated; what proposals were made; what alternatives were discussed; what arguments were brought; what decisions were made, what task were assigned etc.

Our interest in the third level (i.e., on the pragmatic dimension of meetings: what intentions and goals are behind a speaker's utterance) is motivated by the strong user interest in the deliberative

[4] TF-IDF stands for Term Frequency – Inverted Document Frequency (Salton and Buckley, 1989)

information and by the inapropriateness of existing information retrieval techniques in accounting for it (since implicit). Our ultimate goal is to find means to facilitate the access to the argumentative information from debate meetings (given at least the meeting transcriptions).

The scenario is the following: A user needs information about a past meeting, either in quality of a participant who wants to recollect a discussion (memories of co-participants are in fact often inconsistent (Banerjee et al., 2005)), or as a non-participant who missed, for instance, that meeting. Instead of consulting the entire meeting-related information which is usually heterogeneous and scaterred (AV recorings, notes, minutes, e-mails, handouts, etc.), the user asks natural language questions to a query engine in which the meeting data is enriched with argumentative information and which shows the user the relevant passages of the meeting records where the answer to the question can be found.

The paper assesses the users' interest in retrieving argumentative information from meetings and the kind of information and techniques that are needed in answering users' queries, in general (be them argumentative or not). Section 2 reviews the existing work on exploiting meeting information and in conceptualizing meetings, then several meeting data models are described in detail in Section 3. The user query analysis and the main findings are presented in Section 4.

2. Exploiting Argumentation in Meeting Information Retrieval

Although meetings are sometimes judged as time-consuming and unrewarding in terms of results achieved especially if badly organized, they constitute as substantial and important source of information that benefit a corporate organization and performance (Corrall, 1998; Romano and Nunamaker, 2001). Novel multimedia techniques have been dedicated to both meeting recording and access for finding certain elements of interest, e.g., (Cremers et al. 2005; Tucker and Whittaker, 2004).

(Tucker and Whittaker, 2004) identified 4 types of meeting browsers: audio browsers, video browsers, artifacts browsers (exploiting meeting minutes or other meeting-related documents), and browsers working with derived data, such as discourse and temporal structure information.

Depending on the browser type, different levels of meeting content become accessible for information retrieval: audio and video browsers both offer factual and thematic information, while artifact browsers might also touch on deliberative information, as long as it is present, for instance, in the meeting minutes. In contrast, the derived-data browsers aim to account for the argumentative information not explicitly presented in the meeting content, but can be inferrable from it. If minutes are likely to contain only the most salient deliberative facts, the derived-data browsers are much more useful in that they offer access to the full meeting record, and thus to relevant details about the deliberative information sought.

As shown by (Rosemberg and Silince, 1999), tracking argumentative information from meeting discussions is of central importance for building project memories, since, in addition to the "strictly factual, technical information", these memories also need to store relevant information about decision making processes. Seemingly, in a business context, the information derived from meetings is useful for next business processes, as it can explain phenomena and past decisions and can support future actions by mining and assessment (Pallotta et al., 2004a).

Argumentative structure of meeting discussions, possibly visualized in form of argumentation diagrams or maps, can be helpful in meeting browsing. To our knowledge, there are currently three meeting browser that have adopted argumentative structure: ARCHIVUS (Lisowska et al., 2004), ViCoDe (Marchand-Maillet et al, 2005) and the Twente-AMI JFerret browser (Rienks and Verbree, 2006). From these, ARCHIVUS only allows the formulation of natural language queries for searching the meeting data. However, the processing of the query is still keyword-based information retrieval.

The users' interest in argumentation dimension of meetings has been proved by a series of recent studies that attempted to elicit potential user questions (Lisowska 2003; Benerjee at al. 2005; Cremers et al., 2005).

Lisowska's query elicitation study (2003), that (like our current work) is part of the IM2 project[5], was intended to collect user requirements for a meeting retrieval system. The study was performed in a

[5] IM2 stands for Interactive Multimodal Information Management. IM2 is a Swiss National Center of Competence in Research (NCCR) aimed at the advancement of research, and the development of prototypes, in the field of man-machine interaction (http://www.im2.ch).

simulated environment in which users were asked to imagine themselves in a particular role from a series of 4 scenarios. The participants were both IM2-members and non IM2-members and produced about 300 retrospective queries on recorded meetings. Although this study has been criticized (Post et al., 2004; Cremers et al., 2005; Banerjee et al., 2005) for being biased because artificial, obtrusive, and not conforming to strong HCI methodologies for survey research, it was one of the first that shed light on potential queries and classified them in two broad classed (classification that seems to correspond to the argumentative vs. non-argumentative distinction) (p. 994):

- "elements related to the interaction among participants: acceptance/rejection, agreement/disagreement; proposal, argumentation (for and against); assertions, statements; decisions; discussions, debates; reactions; questions; solutions")

- "concepts from the meeting domains: dates, times; documents; meeting index: current, previous, sets; participants; presentations, talks; projects; tasks, responsibilities; topics").

Unfortunately, the authors do not provide precise information on the relative proportions of queries for the classification provided, but simply state that overall more queries belong to the second class, and that queries requiring understanding of the dialogue structure still comprised a sizeable proportion of the queries.

(Banerjee et al., 2005)'s survey concerned instead real, not simulated interviews of busy professionals about actual situations -either meetings to which they previously participated, or meetings they missed. About half of the information sought by interviewees concerned, in both cases, the argumentative dimension of meetings. For non-missed meetings, 15 out of the 26 instances (i.e., 57.69%) concerned argumentative aspects:

- what the decision was regarding a topic (7);
- what task someone was assigned (4);
- who made a particular decision (2);
- what was the participants' reaction to a particular topic (1);
- what the future plan is (1).

The other instances relate to the thematic dimension, i.e., specifics of the discussion on a topic (11). As for missed meetings, the argumentative instances were equally represented (18/36):

- decisions on a topic (7);
- what task was assigned to interviewee (4);
- whether a particular decision was made (3);
- what decisions were made (2);
- reasons for a decision (1);
- reactions to a topic (1).

The thematic questions concern topics discussed, announcements made, and background of participants. The authors showed that interviewees had a level of understanding of missed meetings from the meeting documents (artefacts) provided to them. More importantly, they showed that the recovery of information from meeting recordings is significantly faster when discourse annotation is available, such as the distinction between discussion, presentation, and briefing. It is worth noting that this result comes from an experiment that was performed with subjects other than the interviewees.

Another non-obtrusive user requirement experiment was performed by (Cremers et al., 2005) in a "semi-natural setting" related to the design of a meeting browser. The top 5 search interests highlighted by the 60 survey participants were: *decisions made*, participants/speakers, topics, agenda items, and *arguments for decision*. Two of these, here shown in italics, are clearly argumentative. In fact, the authors acknowledge the necessity to include some "functional" (argumentative) categories as innovative search options. Interestingly, from the user interface evaluation presented in their paper one can indirectly infer how salient the argumentative information is perceived by users: the icons that the

authors intended for emotions, i.e., for a emotion-based search facility, were actually interpreted by users as referring to persons' opinions: *What is person X's opinion?*: positive, negative, neutral.

3. Argumentative Meeting Models

We review in this section some of the current approaches in modelling meeting discussions that take argumentative structure into consideration.

In the IM2 approach, after the transcription of the meeting audio recordings, a first level of annotation covers the description of various metadata about static meeting information such as *date*, *location*, *participants*, *shared documents*, etc. (Marchand-Maillet, 2003).

The second level of annotation is the a discursive annotation which first segments the dialogues into turns and assigns to each utterance a set of dialogue acts highlighting its communicative function (e.g., in our case, the MALTUS tagset (Clark and Popescu-Belis, 2004), a modification of ICSI-MRDS meeting annotation guidelines (Shriberg et al., 2004)).

The third level of annotation is based on the Shallow Dialogue Model, proposed in (Armstrong et al., 2003). This model provides a simple logical structure of dialogues based on the following categories:

- a *dialog* is a non empty set of thematic episodes; a new episode is identified by a topic shift.
- a *thematic episode* is a non empty set of turns; a new turn is introduced by a speaker change.
- a *turn* is a non empty sequence of utterances. Each turn is annotated with one or more dialogue acts.

The argumentation information resides in the dialogue acts that highlight the communicative function of an utterance. This model, however, cannot accommodate multiple communicative functions per utterance, non-utterance communicative actions (such as agreement or disagreement by applause, facial expressions, silence), social behavior of participants (such as greetings, digressions, and jokes), and, more importantly, it does not keep trace of participants' opinion formation and their role in the deliberation process.

The IBIS (Issue Based Information System) model, proposed by (Kunz and Rittel, 1970) and adopted in a series of computer-supported collaborative argumentation systems (e.g., Zeno (Gordon and Karakapilidis, 1999), HERMES (Karakapilidis and Papadias, 2001), Compendium (Selvin et al., 2001)) provides an abstract description of the discussion rationale. It outlines the main lines of a meeting discussion in terms of issues discussed, alternatives proposed in order to solve the issues, positions (acceptance or rejection) taken by participants, conflicts arisen and possibly solved, and decisions made. IBIS considers meeting discussions as *Collaborative Decision Making processes* and focuses on the communicative actions performed by agents in order to accomplish a joint goal, and on how these actions contribute to the established objective. The main dimensions of the IBIS model are:

- an overall task issue
- a set of alternative proposals
- a set of arguments (pros and cons) for each proposal
- a collection of choice criteria (perspectives, preferences) settled upon the agents
- a decision (evaluation) function that combines the criteria to judge the alternatives.

In (Smith et al., 2000), the meeting is viewed as a threaded discussion forum to which each participant contributes by posting a message. Each posted message has an argumentative force, according to which the message is attached to a given discussion thread. The meeting thus becomes a complexly structured, multi-threaded document. The messages, i.e., the dialogue acts having argumentative force, are called *argumentative acts* and are considered as backward-looking acts with forward-looking expectations. This model abstracts away from the dynamics of the discussion (which in turn is modeled with IBIS structures).

Other approaches to the argumentative structuring of conversations can be grouped along the following perspectives:

- communication: the role of each utterance in performing a communicative action, such as request, inform, suggest, confirm (Searle, 1969; Austin, 1962);
- planning: humans do not perform actions at random, but plan them in order to achieve a given goal (Cohen and Perrault, 1979; Litman and Allen, 1990; Carberry and Lynn, 1999);
- structuring: the way participants take part in the discussion and argue their standpoints, according to discourse theories such as Rhetorical Structure Theory (Mann, 1988; Mann and Thompson, 1988).

The latter perspective highlights argumentative phenomena, such as agreement and disagreement, conflicts, complications, contradictions, negotiations, inquiries, persuasions, decision making (Walton and Krabbe, 1995). Within this perspective, several applied studies proposed different models of argumentative structure, for different application domains. The most relevant work in this area is reviewed below.

One of the first attempts in modelling the argumentative structure of meeting discussion has been proposed by (Pallotta et al., 2005; Pallotta et. al., 2004b, Pallotta, 2003). In these works, a coding scheme for argumentative structure has been proposed, the Meeting Discussion Scheme (MDS). In MDS, a small number of argumentative categories are used to classify turns in meeting dialogues. Turns are the basic discourse units and assumed to be already identified. Adjacent turns from the same speaker are grouped into *argumentative segments* when a certain degree of cohesion exists. By cohesion it is meant that a set of adjacent turns produced by a speaker in a discussion perform an atomic argumentative action (e.g., a suggestion, a justification). In addition to argumentative classes, the MDS schema contains two relations, namely *replies-to* and *elaborates*. These relations establish backward links and are used for building rhetorical dependencies between argumentative segments. They differ in their usage, because "replies-to" requires different speakers, whereas "elaborates" does not; it establishes a subordinate rhetorical dependency between the two segments, where the first is the nucleus one and the second the satellite (in RST terms). The MDS schema is not based on a specific argumentation theory, but takes inspiration from the well-known IBIS model (Kunz and Rittel, 1970).

Seemingly, (Rienks and Verbree, 2006; Rienks et al., 2005) propose the argumentative annotation of recorded meetings using the Twente Argumentative Schema (TAS), the only coding schema besides MDS that has been intentionally designed to model argumentation in meeting discussions as argumentative trees. TAS differs from MDS in a crucial point, namely, by modelling argumentation by means of rhetorical relations between simple turns. Turns are classified as ordinary dialogue acts (i.e. Statements, Weak-statement, Open-issue, A/B-issue, Yes/No-issue) according to widely adopted schema such as the ICSI-MRDS (Shriberg et al., 2004) or the MALTUS schema (Clark and Popescu-Belis, 2004). TAS also imposes a stricter structure of the conversational flow as stated in this sentence taken from the paper: "TAS was constructed in a way that it preserves the conversational flow. By applying a left-to-right, depth first search, walk through on the resulting trees, the reader is able to read the resulting trees as if reading transcripts. This was realized by assuring that in principle every next contribution of a participant becomes a child of the previous contribution, unless the current contribution relates more strongly to the parent of the previous contribution". In our opinion, it is not always the case that the conversational flow agrees with the argumentative structure – for instance, whenever two proposals are considered at the same time but the speaker does not necessarily refer to the most recent. In TAS work, an interesting empirical study has been conducted on the usability of argumentative structure in solving a given task (i.e. answering questions on a meeting using a Meeting Browser). As a result, it has been shown that the cognitive effort required in order to understand the argumentative structured (increasing response time) is paying off on the long term by improving the quality of the answers and changing of perception of questions from difficult without argumentative information to easy using argumentative information. Finally, a learning algorithm is presented which learns classifiers for unit labels. Its performance is justified by the fact that categories are easily marked and by the fact that Statements are highly probable and Questions are easily detectable. However, they do not compare their result with other works on the detection of similar types of dialogue acts (Popescu et al., 2005; Ang et al., 2005).

Stent, proposed in (Stent 2000b) a coding scheme based on DAMSL (Core and Allen, 1997) and Rhetorical Structure Theory (Mann and Thompson, 1988), which was used to annotate argumentation acts in some dialogues from the Monroe corpus (Stent, 2000a). She proposed an annotation schema for

dialog based on the specific data and an ad-hoc decision-trees procedure. She also showed that the RST relations *elaboration* and *joint* are too broad classes, and added relations for structural patterns in dialog (make-plan, describe-situation, list, joke).

(Clark and Schaefer, 1989) proposed the Contribution Graphs (CGs) schema, an annotation schema for simple dialogues that was successively extended to cope with multi-party discussions by (Novick et al., 1996) and used to annotate a corpus of meeting discussions (the Watergate Tape transcripts). The coding schema is based on two broad categories, *presentation* and *acceptance*, which are used to tag one or more dialogue turns. Presentations and acceptances can be linked together to form a *contribution*. A contribution can also play the role of presentation or acceptance, this making then the model recursive. The contribution relation in CGs is intended to model the fact that the information that has been presented in the presentation node has actually been understood by the addressee and thus "accepted". The acceptance turn can also include a subsequent presentation, which might be accepted by another acceptance turn, thus forming another contribution node. The CGs schema captures the structural connections between turns in a way that is similar to MDS's "replies to" relation. However, it fails in capturing the argumentative force of each component of the relation.

With the Adjacency Pair model (Schegloff & Sacks, 1973) and similar projects (Galley et al., 2004; Hillard et al., 2003), the importance of tracking agreement and disagreement in discussions has been again recognized. These methods have the great advantage of being automatic, but they help only partly in reconstructing the argumentative information needed in order to answer real user queries.

Within the Information State Update Dialogue Model, a new argumentative perspective has been introduced by Larsson's notion of Issue-Under-Negotiation (IUN) (Larsson, 2002). This notion naturally extends the Ginzburg's notion of Question-Under-Discussion (Ginzburg and Fernandez, 2005) to the new class of negotiation dialogues,where argumentative threads are seen as pertaining to particular Issues, which in turn are modelled as questions on the IUN stack. *Introduce moves* introduce new issues, *proposals* introduce possible alternative answers thereto, *acceptances* or *rejections* remove those alternatives. The model is very similar in spirit to the IBIS model and is also grounded on a fully-fledged dialogue theory. It has been adopted by (Niekrasz et al., 2005) for the real-time reconstruction of an argumentative structure by overhearing discussions in design meetings, and the resulting analyses are to be used for answering questions about the meeting discussion.

Finally, the annotation schema for dialog proposed by (Traum and Hinkelman, 1992) includes four levels of conversational acts, with argumentation acts on the top. Other schemas for argumentative modelling of discussion have been developed mainly for formal argumentation, socio-linguistic and ethnographic studies. A review of these works is outside the scope of the paper.

The reliability of the surveyed coding schemes has been only partially evaluated by their authors. It is still an open problem whether the argumentative annotation is a feasible task and the automatic argumentative parsing is possible. An alternative solution that seems appropriate is the production of argumentative information while the meeting takes place. In (Whittaker et al. 2005) it was suggested that meeting-capture systems should allow using personal notes of participant as indexes for the recorded meeting. In our opinion, a smart system for taking personal notes, possibly based on argumentation diagramming tools such as Compendium, would certainly fit the on-line argumentative annotation task. Moreover, the collaborative use of these note taking tools would help in establishing a consensus in identifying the argumentative categories. As a useful side effect, the enhanced meeting recording system would easily produce high-quality annotated data, which are otherwise difficult and costly to produce. These annotated data could then be used for training an automatic algorithm with machine learning techniques.

4. Assessing the difficulty of answering queries on meetings

Meetings contain heterogeneous information. As shown in Table 1, at least three main dimensions for the meeting content can be identified: factual, thematic, and argumentative. Accordingly, the user questions concern various factual, thematic, and/or argumentative elements of interest, as proved by the query elicitation studies outlined in section 2. Searching relevant information through the recorded meeting dialogues poses important problems when using standard IR indexing techniques (Baeza-Yates and Ribeiro-Nieto, 2000), because users ask different types of queries for which a single retrieval strategy, as used in the traditional IR systems, might be insufficient. Depending on the type of query, the required recall and precision may vary since in some cases the query needs to be answered

with precise information (e.g. who, where, when, why), but other queries might need, for instance, a whole passage from the meeting recording, where the element of interest is contained (Show me the passage where they talked about me). Another problem for meeting querying is that it still unclear what level of granularity the document base should have (from turns and episodes to main topics segments or entire meetings).

Spoken-language information retrieval (Vinciarelli, 2004) and automatic dialogue act extraction techniques (Stolke et al., 2000; Clark and Popescu-Belis, 2004; Ang et al., 2005) have been applied to meeting recordings and provided good results under the assumption that the user is interested in retrieving either topic-based or dialog act-based information. But this assumption is invalidated by users query elicitation research that shows that such information is only sought in a small fraction of the users' queries. A particular problem for these approaches is that the topic looked for is usually not a query itself (*Was topic T mentionned?)*, but just a parameter in more structured questions (*What was decided about T?*). Moreover, the relevant participants' contributions (dialog acts) need to be retrieved in combination, not in isolation (*The reactions to the proposal made by X*).

The study we present in this section is aimed at assessing how difficult it is to answer the questions users typically ask about a meeting. Our goal is to provide insights into: How many queries can be answered using standard IR techniques on meeting artefacts only (e.g., minutes, written agenda, invitations)? How many can be answered with IR on meeting recordings? When IR does not apply or is insufficient, what kind of additional information and inference is needed (e.g., factual information about the participants and the meeting dynamics, external information about the meeting's context such as the relation to a project; semantic interpretation of question terms and references, computation of durations, aggregation of results, etc.)?

4.1 Why a new query classification experiment

The existing query elicitation experiments highlighted a series of question types that users typically would like to ask about past meetings. It has also been possible to roughly separate the information sought into two broad classes: argumentative information (about the argumentative process and the outcome of debate meetings), and non-argumentative information (factual, i.e., about the meeting as a physical event, and thematic, i.e., about what has been said in terms of topics).

However, assessing the level of difficulty of a query based on these broad categories of queries couldn't provide insightful results, because these would be general, thus less interpretable. Again, the complex queries requiring mixed information would escape observation because assigned to a more general class. We therefore considered it necessary to perform a detailed analysis of each query instance, as this provides not only detailed, but also traceable information.

4.2 Data

- IM2-set

  The only available set of user queries on meetings at the time this work was done was the IM2-set (Lisowska, 2003). It comprises 270 questions shortly described in section 2. Questions are annotated with a label showing whether or not it was produced by an IM2-member.

- Manager Survey Set (M.S.-set)

  As IM2 queries were not derived from a natural business setting, we conducted another, our own survey of user requirements on meeting querying among a number of 3 managers from a company. We collected 35 questions (showed in Annex 1).

- BET observations

  This set, developed in the framework of the IM2-AMI's meeting browser evaluation test (Wellner et al., 2005), does not contain queries but a series of statements about an IM2 recorded meeting. The statements were made by IM2 members and retained as characteristic of the meeting. The set comprises 294 natural language statements. We use it as a "validation" set for the IM2 queries: a query is considered as "realistic" or "empirically grounded" if there is a BET observation that can answer the query.

4.3 Annotation Experiment

The queries from the IM2-set and the M.S.-set were analyzed by 2 different teams of 2 judges. Each team discussed each query, and classified it along two main dimensions (we refer to these as to the first and second classifications):

- query type: the type of meeting content to which the query pertains (cf. our model, Table 1)
- query difficulty: the type of knowledge required to provide the answer.

There were 4 categories for the first classification and 10 categories for the second; the two sets of categories are described below. A query can be assigned with one category from the first set perceived as the most salient, and with multiple categories from the second set. With only one exception (for the category *Role of IR*), the values for the categories are all Boolean:

Set 1 – query type:

1. *factual*: the query pertains to the factual meeting content
2. *thematic*: the query pertains to the thematic meeting content
3. *process*: the query pertains to the argumentative meeting content, more precisely to the argumentative process
4. *outcome*: the query pertains to the argumentative meeting content, more precisely to the outcome of the argumentative process

Set 2 - query difficulty:

1. *Role of IR*: states the role of standard[6] Information Retrieval (in combination with Topic Extraction techniques[7]) in answering the question.
   0 = irrelevant (N/A)
   Example: *What decisions have been made?*
   1 = successful (these techniques are sufficient alone or in combination with additional inference);
   Example: *What topics were discussed?*
   -1 = insufficient (these techniques are not sufficient alone but require additional inference and information (e.g., argumentative, cross-meeting, external corporate/project knowledge).
   Example: *Who rejected the proposal made by X on issue Y?*
2. *Artefacts*: answering the query requires information such as Agenda, minutes of previous meetings, e-mails, invitations and other documents related and available before the meeting.
   Example: *Who was invited to the meeting?*
3. *Recordings*: answering the query requires the meeting recordings (audio, visual, transcription). This is almost always true, except for queries where Artefacts or Metadata are sufficient, such as *What was the agenda?, Who was invited at the meeting?*).
4. *Metadata*: answering the query requires context knowledge kept is static metadata (e.g., speakers, place, time).
   Example: *Who were the participants at the meeting?*
5. *Dialogue Acts & Adjacency Pairs*: answering the query requires techniques based on dialogue acts and adjacency pairs.
   Example: *What was "John" response to my comment on the last meeting?*
6. *Argumentation*: answering the query requires metadata (annotation) about the argumentative structure of the meeting content.
   Example: *Did everybody agree on the decisions, or were there differences of opinion?*

[6] By standard IR we mean techniques based on bag-of-word search and TF-IDF indexing.

[7] Topic extraction techniques are based on topic shift detection (Galley et al., 2003) and keyword extraction (van der Plas et al., 2004).

7. *Semantics*: answering the query requires semantic interpretation of terms in the query and reference resolution, including deictics (e.g., for *how long, usually, systematically, criticisms; this, about me, I*). The terms requiring semantic interpretation are underlined.
   Example: *What decisions got made easily?*
8. *Inference*: answering the query requires inference (deriving information that is implicit), calculation, and aggregation (e.g. for "command" queries asking for lists of things - participants, issues, proposals).
   Example: *What would be required from me?*
9. *Multiple meetings*: answering the query requires cross-meetings analysis.
   Example: *Who usually attends the project meetings?*
10. *External*: answering the query requires external knowledge, not present in the meetings (e.g., information about the corporation or the projects related to the meeting).
   Example: *Did somebody talk about me or about my work?*

4.4 Annotation Example

We provide an example of annotation for the classification of the query:

Were any of X's suggestions included in the final specifications?

Type of meeting content: outcome
Query difficulty:
Role of IR = Insufficient.
IR can be used to retrieve passages on "final specification"; other terms must instead be filtered out. IR would fail to identify the suggestions of X, because it is very unlikely that "suggestions" will be lexicalized in recordings, e.g., "I suggest/propose to do A for the final specification".
Recordings = 1.
Answering requires the access to meeting recordings.
Metadata = 1
Information about the author, X, of the contribution representing suggestions is required.
Multiple meetings = 0
No information about other meetings is required.
Dialogue Acts & Adjacency Pairs = 0
Dialogue Acts/Adjacency Pairs techniques do not help.
Argumentation = 1
The contributions of X that are "suggestions" must be marked as such in the argumentative structure of the meeting.
Semantics = 1
For answering this query it is necessary to know what is meant by "specification", in order to look up external related documents where references and/or the content of X's suggestion can be found. "Final" also needs to be interpreted, probably as the last version of the specifications.
Inference = 1
In answering, it is required to compute the list of suggestions made by X and then check if at least one item of this list appears in the external document "final specification".
External = 1.
To answer, external knowledge is required about the set of documents to be looked up in order to check out whether X's suggestions are included.

It is apparent from the above analysis the type of query just exemplified needs a complex aggregate of information beyond the lexical one found in transcriptions. Moreover, simply searching for terms contained in the query without a careful interpretation would certainly lead to the retrieval of false positives. We will show with this study that complex queries like this are not exceptions, but they represent a large portion of the elicited user queries.

4.5 Some Recurrent Problems in Annotating

A frequent problem that arose in the classification of queries was the fact that no clear distinction could be made between "discussion" seen as a series of argumentative processes and "discussion" as mention of the topic of the conversation without any confrontation/debate:

Argumentative query:
What decisions got made only after prolonged or intense discussion?
What were the most discussed issues?)

Thematic query:
What topics were discussed?
I seem to recall Topic A discussed before? Was it?

A similar problem concerned the term "issue", which can be either thematic or argumentative.

In general, some of the queries (from the IM2-set) were ambiguous to annotators with respect to users' elements of interest:

Is there something that was not mentioned at the beginning and that appeared later?
Who discussed new ideas?
I seem to recall Topic A discussed before? Was it?
Who was actively participating to the meeting?
Who is originator of the project idea?

While they have been assigned the most plausible interpretation by the human judges, for an automatic system they would be hard to treat (e.g., *something mentioned at the beginning* might easily be identified as an agenda item by a human, but what about a system?)

Some of the queries were perceived by the annotators as very hard to answer:

Who is the most dominant person in meetings?
Are there any software issues that I should know about or modifications that are needed?
What did participant X actually intend when saying P during the discussion of issue I?
Can I have a short summary of the last meeting.
Are there any dates mentioned that I should put in my agenda (deadlines, appointments)?

Yet other queries were estimated as impossible to answer:

Does an employee have initiative - do they do what they're told and when, or do they come up with their own work/schedules?
I'd like to send a memo to people involved in Topic A.
Were there any documents produced recently in relation to the project? Give me their authors, titles, dates, and types (publication-where?, deliverable, report, etc.)

Another problematic query is the one found in the M.S.-set:

To whom are they (action items) assigned & by when?

In this example, one needs to look at passages where decisions were made and detect who was the addressee (Jovanovic et al., 2006) of the task assignment. A possible situation of task assignment is the following: a person B is assigned the task X:

A: *You will do X by the end of the month*
B: *OK*

Which month is referred in the turn can be inferred from the metadata by looking at the date of recording.

4.6 Classification Results

4.6.1 Query type

Results from the first classification task for both query sets are reported in Table 2. In both sets, the information most sought was argumentative: about 55% of the IM2-set queries are argumentative (process or outcome). This contradicts Lisowska's initial estimation (Lisowska et al., 2004:994) that the non-argumentative queries prevail, but confirms the figures obtained in (Banerjee et al., 2005), i.e., 57.69%. In our real (as opposed to simulated) survey, we obtained even higher percentages for argumentative queries (60% and 68%, depending on the annotation). The argumentative queries are followed by factual and thematic in both query sets, with a slight advantage for factual queries.

Table 2. Query classification along the first dimension (meeting content type)

Category	IM2-set (size: 270)		Manager Survey set (size: 35)	
	Annotation 1 (%)	Annotation 2 (%)	Annotation 1 (%)	Annotation 2 (%)
Factual	24.81	45.56	20	20
Thematic	18.52		20	11.43
Process	30.00	32.59	22.86	28.57
Outcome	26.67	21.85	37.14	40
Process, Outcome	56.67	54.44	60	68.57

The inter-annotator agreement for this first classification is reported in Table 3. The proportion of queries on which annotators agree are significantly high (about 80% on the fist set and 85% on the second). We only report here the results for the argumentative categories, Process and Outcome. A high percentage of the agreed queries were marked as being argumentative: 54.93% in IM2-set and 63.33% in the M.S.-set.

Table 3. Inter-annotator agreement for the first classification task

Category	IM2-set (size: 270)			Manager Survey set (size: 35)		
	number	Ratio (%)	k-score (%)	number	Ratio (%)	k-score (%)
Process	229	84.81	82.92	31	88.57	87.81
Outcome	245	90.74	89.58	32	91.43	90.86
Process, Outcome	213	78.89	76.25	30	85.71	84.76
agreed argumentative (label: AA)	117	117/213 = **54.93%**		19	19/30 = **63.33%**	

Table 4 shows the distribution of IM2 queries according to query origin (IM2-member vs. non IM2-member); the label AA refers to agreed argumentative queries.

Table 4. IM2 set: Distribution according to query origin (IM2-member vs. non IM2-member)

Query	G1: IM2-members	G2: non IM2-members
All	163/270 = 60.37%	107/270 = 39.63%
AA	81/117 = 49.69%	36/117 = 30.77%
AA ratio in groups	81/163 = 69.23%	36/107 = 33.64%

As for the IM2 queries that are realistic according to BET observations, their proportion is 41.85% (they match a BET observation, i.e., that can be answered from a BET statement). From these, a further 28.89% are argumentative, while the ratio of BET statements that contain argumentative information is 66.92%.

4.6.2 Query difficulty

The next table shows the number of queries in each set that require argumentative information in order to be answered; the figures are reported both for the queries overall and by type. As expected, no argumentation information is necessary for answering factual queries, but some thematic queries do need it (24% in the IM2-set and 42% in the M.S.-set). Almost all the argumentative queries require

argumentation information; some not. Overall, the majority of queries in both sets need argumentation information in order to be answered (56.30% from IM2 queries, and 65.71% from M.S. queries).

Table 5. Queries requiring argumentative information

Category	IM2-set, Annotation 1			Manager Survey set, Annotation 1		
	total	Require argumentation	Ratio (%)	total	Require argumentation	Ratio (%)
Factual	67	0	0	7	0	0
Thematic	50	12	24	7	3	42.86
Process	81	73	90.12	8	7	87.5
Outcome	72	67	93.06	13	13	100
All	270	152	56.30	35	23	65.71

More results for the second classification task (on query difficulty) are shown in the Annex, which lists what kind of information is required most often. Table 7 below synthesizes the role of IR in answering users' queries: IR might help for 14.44% of the IM2 queries, and for 20% of the M.S. queries. In 50% and 25.71% of the cases, respectively, it simply cannot be applied. Finally, IR is not enough in 35.56% of the queries from the IM2-set, and in 54.29% of the M.S. queries.

Table 7. Role of IR and Topic Extraction in answering the queries

Role	IM2-set		Manager Survey set	
	All	AA	All	AA
Irrelevant (N/A)	135/270 = 50%	55/117 = 47.01%	9/35 = 25.71%	3/19 = 15.79%
Sufficient	39/270 = 14.44%	1/117 = 0.85%	7/35 = 20%	1/19 = 5.26%
Insufficient	96/270 = 35.56%	61/117 = **52.14%**	19/35 = 54.29%	15/19 = **78.95%**

If we report to only the agreed argumentative queries (column AA), then IR success rate drops to almost 0. Only 1 argumentative query of each set could be answered with IR:

What were the decisions to be made (open questions) regarding the topic t1?
When is the NEXT MEETING planned? (e.g. to follow up on action items)

The first query can be answered by searching topic t1 in the Agenda of the meeting (part of the Artefact set). The second is answerable by searching the terms "next meeting" (and their lexical variants) in the recordings.

For the majority of argumentative queries IR is insufficient, and in the rest of the cases it is irrelevant (N/A). For those queries that cannot be answered with IR, we list below the most frequent combination of information types that is needed.

Table 8. Types of information required when IR fails

Set	IR alone fails	Information required in order to answer the queries (unless specified otherwise, recordings are also required)	Freq.	(%)
IM2	IR insufficient 96 cases	Argumentation, Semantics, Inference	15	15.62
		Argumentation, Semantics	11	11.46
		Metadata, Argumentation, Semantics, Inference	9	9.38
		Cross-Meetings, Argumentation, Semantics, Inference	8	8.33
		Metadata, Argumentation, Semantics	7	7.29
		Argumentation	5	5.21
		Argumentation, Inference	4	4.17
	IR irrelevant 135 cases	Metadata, Argumentation, Semantics, Inference	14	10.37
		Artefacts, Argumentation, Semantics, Inference	9	6.67
		Metadata, Semantics, Inference	8	5.93
		Argumentation, Semantics, Inference	8	5.93
		Metadata, Cross-Meetings Semantics, Inference (no recordings)	7	5.19
		Metadata (no recordings)	5	3.7

M.S.	IR insufficient 19 cases	Argumentation, Semantics, Inference	6	31.58
		Argumentation, Inference	4	21.05
		Argumentation, Semantics	2	10.53
		Argumentation, Semantics, Inference, External	2	10.53
		Artefacts, Argumentation, Semantics	1	5.26
	IR irrelevant 9 cases	Artefacts, Metadata, Semantics, Inference (no recordings)	2	22.22
		Artefacts, Metadata (no recordings)	2	22.22
		Artefacts, Metadata, Semantics (no recordings)	1	11.11
		Artefacts, Semantics (no recordings)	1	11.11

4.7 Conclusions on the Assessment Experiment

The analysis of the annotations obtained for the 305 queries (35 from a natural set, M.S., and 270 from the IM2-set) revealed that:

- the information most sought by users from meetings is argumentative (i.e., pertains to the argumentative process and it outcome). It constitutes more than half of the total queries, while factual and thematic information are similar in proportions;
- there was no significant difference in this respect between the IM2-set and the M.S-set;
- the decision as to whether a query is argumentative or not is easy to draw (as suggested by the high inter-annotator agreement);
- standard IR and topic extraction techniques are insufficient in answering most of the queries. Less than 20% of the whole query set can be answered with IR, and almost no argumentative question.
- Argumentative information is needed in answering the majority of the queries;
- The information types that are needed most are: Argumentation, Semantics, Inference, Metadata.

5. Conclusion

This paper provides insights into the importance of argumentative information in accessing meeting records. We reviewed a number of existing coding schemes for the annotation of meeting data which consider the argumentative structure as one of the dimensions in annotation. While in its infancy, argumentative structuring of meeting discussion is gaining increasing interest in Multimedia Information Retrieval.

We quantified the proportion of queries involving argumentative dimension of the meeting content by performing an in-depth analysis of queries collected in two different elicitation surveys.

The analysis of the annotations obtained for the 305 queries (270 from the IM2-set, 35 from M.S.-set) was aimed at providing insights into different matters, such as:

Information sought: What type of information is sought by users from meetings? Do results differ in the natural set (M.S.-set) from the unnatural set (IM2-set)? How many queries pertain to the argumentative process and it outcome? What is the distribution of argumentative queries in the IM2-set among biased authors (IM2-members) and non-biased? How many of the IM2 queries were realistic, according to BET observations?

Information and techniques needed in order to answer the queries: How many queries require IR (plus topic extraction techniques)? For how many queries IR is sufficient? For what percentage of the argumentative queries is IR sufficient? When IR fails, what additional information and inference is needed? What information is needed more frequently for answering queries?

This work represents an initial step towards a full understanding of queries on the meeting domain. In the future, ee expect to cover aspects like the automatic classification of answer types and, more importantly, automatic extraction of argumentative features and their relations with other components of the query (e.g. topic, named entities, events).

References

Ang J., Liu Y., Shriberg E. 2005. Automatic Dialog Act Segmentation and Classification in Multiparty Meetings. In *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA. 2005.

Armstrong S., Clark A., Coray C., Georgescul M., Pallotta V., Popescu-Belis A., Portabella D., Rajman M., Starlander M., Natural language queries on natural language data: a database of meeting dialogues. In *proceedings of NLDB'2003 (8th International Conference on Applications of Natural Language to Information Systems),* Burg/Cottbus, Germany. 2003.

Austin, J. L. 1962. How to Do Things with Words. Oxford. 1962.

Baeza-Yates, R. and Ribeiro-Nieto, B. 2000. Modern Infor-mation Retrieval. Addison Wesley. 2000.

Banerjee S., Rose C., Rudnicky A. 2005. The Necessity of a Meeting Recording and Playback System, and the Benefit of Topic--Level Annotations to Meeting Browsing. In *Proceedings of the Tenth International Conference on Human-Computer Interaction*, 12-16 September 2005, Rome, Italy.

Carberry S., and Lambert L. 1999. A Process Model for Recognizing Communicative Acts and Modeling Negotiation Subdialogues. *Computational Linguistics* 25(1): 1-53 (1999).

Cohen, P. R. and Perrault C.R. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3).

Clark A. and Popescu-Belis A. 2004. Multi-level Dialogue Act Tags. *Proceedings of the SIGDIAL'04 (5th SIGdial Workshop on Discourse and Dialogue)*, Cambridge, MA, USA, p.163-170. 2004.

Clark, H.H. and Schaefer, E.F. 1989. Contributing to Discourse. *Cognitive Science* 13, 1989, p. 259-294.

Conklin J. 2006. Dialogue Mapping : Building Shared Understanding of Wicked Problems. John Wiley & Sons, 2006.

Core, M. G., and Allen, J. F. 1997. Coding dialogs with the DAMSL annotation scheme. In *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*. 1997.

Corrall, S. 1998. Knowledge management. Are we in the knowledge management business? *ARIADNE: the Web version*, 18. 1998.

Cremers, A. H. M., Hilhorst, B, and Vermeeren, A.P.O.S. 2005. "What was discussed by whom, how, when and where?" personalized browsing of annotated multimedia meeting recordings. In *Proceedings of HCI 2005: 11th international conference on human computer interaction* (pp. 1-10). Lawrence Erlbaum Associates.

Cugini, J. et al. 1997. Methodology for evaluation of collaborative systems. *Technical Report Revision 3.0, The Evaluation Working Group of the DARPA Intelligent Collaboration and Visualization Program*.

Galley M., McKeown K., Fosler-Lussier E., Jing H. 2003. Discourse Segmentation of Multi-Party Conversation. In *Proceedings of ACL 2003 conference*: 562-569.

Galley, M., McKeown K., Hirschberg J., and Shriberg E. 2004. Identifying agreement and disagreement in conversational speech: Use of bayesian networks to model pragmatic dependencies. In *Proceedings of 42nd Meeting of ACL,* Barcelona, July 21-26 2004.

Ginzburg, J. and Fernandez, R. 2005. Scaling up to multilogue: some benchmarks and principles. In *Proceedings of the 43rd meeting of the Association for Computational Linguistics*, Ann Arbor, Michigan, 2005.

Gordon, T. and N. Karacapilidis. 1999. The Zeno argumentation framework. German *Artificial Intelligence*, 3(20):20–29.

Hillard D., Ostendorf M., Shriberg E. 2003. Detection Of Agreement vs. Disagreement In Meetings: Training With Unlabeled Data. In *Proceedings of HLT-NAACL 2003 Conference*, Edmonton, Canada.

Janin, A., J. Ang, S. Bhagat, R. Dhillon, J. Edwards, J. Macias-Guarasa, N. Morgan, B. Peskin, E. Shriberg, A. Stolcke, C. Wooters, B. Wrede. 2004. The ICSI Meeting Project: Resources and Research. In *Proceedings of the NIST ICASSP 2004*

Meeting Recognition Workshop, Montreal, Canada. May 2004.

Jovanovic N., op den Akker R., Nijholt A. 2006. Addressee Identification in Face-to-face Meetings. In *Proceedings of the 11th Conference EACL 2006*. Trento, Italy, April 3-7 2006.

Karacapilidis, N. and D. Papadias. 2001. Computer supported argumentation and collaborative ecision making: The HERMES system. *Information Systems*, 24(4):259–277.

Kunz, W. and H. W. J. Rittel. 1970. Issues as elements of information systems. *Technical Report 131*, Universität Stuttgart, Institüt für Grundlagen der Planung, Germany.

Larsson S., 2002. Issue-based Dialogue Management. *Ph.D. Dissertation*, Göteborg University. Also published as *Gothenburg Monographs in Linguistics* n. 21. 2002.

Lisowska, A. 2003. Multimodal Interface Design for the Multimodal Meeting Domain: Preliminary Indications from a Query Analysis Study. Geneva, *Technical Report n. 30* of University of Geneva. 2003.

Lisowska A., Popescu-Belis A. & Armstrong S. 2004. User Query Analysis for the Specification and Evaluation of a Dialogue Processing and Retrieval System. In *Proceedings of the LREC 2004 (Fourth International Conference on Language Resources and Evaluation)*, Lisbon, Portugal, vol.III, p.993-996. 2004.

Lisowska A., Rajman M., & Bui T.H. 2004. ARCHIVUS: A System for Accesssing the Content of Recorded Multimodal Meetings. In *Proceedings of the Joint AMI/PASCAL/IM2/M4 Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, Martigny, Switzerland, 2004.

Litman, D. J. and Allen, J.F. 1990. Discourse planning and commonsense plans. In *Philip Cohen, Jerry Morgan, and Martha Pollack, (eds), Intentions in Communication*. MIT Press, Cambridge, Massachusetts, pages 366–388.

Mann, W. C. 1988. Dialogue games: Conventions of human interaction. *Argumentation*, 7:289–327.

Mann, W.C and S.A Thompson. 1988. Rhetorical Structure Theory: Towards a Functional Theory of Text Organization. *Text*, 8(3):243–281.

Marchand-Maillet, S. 2003. Meeting record modelling for enhanced browsing. *Technical Report 03.01*, Computer Vision and Multimedia Laboratory, Computing Centre, University of Geneva, Switzerland. 2003.

Marchand-Maillet, S., and Bruno, E. 2005. Collection Guiding: A new framework for handling large multimedia collections. In *Proceeding of the First Workshop on Audio-visual Content And Information Visualization In Digital Libraries*, AVIVDiLib05, Cortona, Italy, May, 2005.

Niekrasz, J. and Gruenstein, A. 2006. NOMOS: A Semantic Web Software Framework for Annotation of Multimodal Corpora. In *Proceedings of LREC 2006 conference*, Genova, Italy, 2006.

Niekrasz J., Purver M., Dowding, J., Peters S. 2005. Ontology-based discourse understanding for a persistent meeting assistant. In *Proceedings of the 2005 AAAI Spring Symposium on Persistent Assistants*, Stanford CA, USA, American Association of Artificial Intelligence. 2005.

Novick, D.G., Walton, L. and Ward, K. 1996. Contribution Graphs in Multiparty Discourse. In *Proceedings of the International Symposium on Spoken Dialogue (ISSD-96)*, Philadelphia, PA, October, 1996, 53-56.

Pallotta, V. 2003. A computational dialectics approach to meeting dialogues tracking and understanding. Linguistic and new professions, In *Materiali Linguistici*, 1095.41:511–532. 2003.

Pallotta V., Ghorbel H., Ballim A., Lisowska A., Marchand-Maillet S., 2004a. Towards meeting information systems: Meeting knowledge management. In *Proceedings of the International Conference on Enterprise Information Systems (ICEIS)*, 464–469, Porto, Portugal. 2004.

Pallotta V., Ghorbel, H., Coray G. and Ruch P., 2004b. An argumentative annotation schema for meeting discussions. In *Proceedings of the LREC 2004 international conference*, 26-28 May 2004, Lisbon, Portugal. 2004.

Pallotta V., Nieksraz J. and Purver M. 2005. Collaborative and Argumentative Models of Meeting Discussions. In *Proceeding of*

CMNA-05 international workshop on Computational Models of Natural Arguments (part of IJCAI 2005). July, 30th 2005, Edimburgh, UK.

van der Plaas L., Pallotta V., Rajman M. and Ghorbel H. 2004. Automatic keyword extraction from spoken text: A comparison between two lexical resources: the EDR and WordNet. In *Proceedings of the LREC 2004 international conference*, 26-28 May 2004, Lisbon, Portugal.

Popescu-Belis A., Clark A., Georgescul M., Zufferey S. & Lalanne D. 2004. Shallow Dialogue Processing Using Machine Learning Algorithms (or not). In *Proceedings of the MLMI 2004 Workshop on Machine Learning for Multimodal Interaction*, LNCS 3361, Springer-Verlag, Berlin, p. 277-290.

Post, W.M., Cremers, A.H.M., and Blanson Henkemans, O.A. 2004. A Research Environment for Meeting Behavior. In *Proceedings of the 3rd Workshop on Social Intelligence Design*, pp. 159-165, University of Twente, Enschede.

Purver M, Ehlen P., Niekrasz J. 2006. Detecting Action Items in Multi-Party Meetings: Annotation and Initial Experiments. In *Proceedings of the 3rd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI 2006)*, Washington DC, USA.

Rienks R., Heylen D., van der Veijden E. 2005. Argument Diagramming of Meeting Conversations, in Multimodal Multiparty Meeting Processing, In *Proceedings of the 7th International Conference on Multimodal Interfaces (ICMI),* October 07 2005, A. Vinciarelli and J-M. Odobez (eds), Trento, Italy, pp. 85-92.

Rienks R., Verbree D., About the Usefulness and Learnability of Argument-Diagrams from Real Discussions. In *Proceedings of MLMI 2006*. 1-3 May 2006, Washington DC, USA.

Romano N. C. Jr., Nunamaker J.J.F. 2001. Meeting Analysis: Findings from Research and Practice. In *Proceedings of 34th Annual Hawaii International Conference on System Sciences (HICSS-34)*, January 3-6, 2001, Maui, Hawaii - Track 1. IEEE Computer Society.

Rosemberg, D. and Silince J.A.A. 1999. Common ground in computer-supported collaborative argumentation. In *Proceedings of the Workshop on Computer Supported Collaborative Argumentation for Learning Communities (CSCL99)*.

Salton, G. and Buckley C. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management* 24(5): 513–523. 1988.

Schegloff E.A., Sacks H. 1973. Opening up closings. *Semiotica*, 7(4):511–532. 1973.

Selvin, A., Buckingham Shum S., Sierhuis M., Conklin J., Zimmermann B., Palus C., Drath W., Horth D., Domingue J., Motta E., G. Li. 2001. Compendium: Making meetings into knowledge events. In *Proceedings of Knowledge Technologies Conference*, March 4-7, Austin, TX.

Searle J. 1969. Speech Acts: An essay in the philosophy of language. Cambridge University Press.

Shriberg E., Dhillon R., Bhagat S., Ang J., and Carvey H. 2004. The ICSI Meeting Recorder Dialog Act (MRDA) Corpus. In *Proceedings of HLT-NAACL SIGDIAL Workshop*, April-May 2004, Boston.

Smith, M., Cadizz J.J., and Burkhalter B. 2000. Conversation trees and threaded chats. *Technical Report* MSR-TR-2000-43, Microsoft Research, Redmond, WA.

Stent, A. 2000a. The Monroe Corpus, *Technical Report 728 and Technical Note 99-2*, Computer Science Dept., University of Rochester, March 2000.

Stent A., 2000b. Rhetorical structure in dialog. In *Proceedings of the 2nd International Natural Language Generation Conference (INLG'2000)*, June 2000.

Stolke, A., Ries K., Coccaro N., Shriberg E., Bates R., Jurasfky D., Taylor P., Martin R., Van Ess-Dykema C., and Meeter M. 2000. Dialogue act modelling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–373, 2000.

Traum D. R. and Hinkelman E. A. 1992. Conversation acts in task-oriented spoken dialogue. Special Issue on Non-literal language of *Computational Intelligence*, 8(3):575-599, 1992.

Tucker, S. and Whittaker, S. 2004. Accessing multimodal meeting data: systems, problems and possibilities. In *Proceedings of the Fourth Joint AMI/PASCAL/IM2/M4 workshop on multimodal interaction and related machine learning algorithms*, June 21-2, Martigny. 2004.

Vinciarelli, A. 2004. Noisy text categorization. In *Proceedings of International Conference on Pattern Recognition (ICPR)*. 2004.

Walton, D. N. and Krabbe, E. C. W. 1995. Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. State University of New York Press, Albany, NY. 1995.

Wellner, P., Flynn, M., Guillemot, M. 2004. Browsing Recorded Meetings with Ferret, In *Proceedings of MLMI'04 workshop*, Springer-Verlag, 2004.

Wellner, P., Flynn, M., Tucker, S., Whittaker, S. 2005. A Meeting Browser Evaluation Test, In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, Portand, Oregon, USA, 2nd-7th April. 2005.

Whittaker S., Laban R., Tucker S. 2005. Analysing Meeting Records: An Ethnographic Study and Technological Implications. In *Proceedings of 2nd Joint Workshop on Multimodal Interaction and Machine Learning Algorithms (MLMI 2005)*, Edinburgh, UK, 11-13th, July. 2005.

Annex1. Manager Survey set. Query classification.

Query	F	T	P	O	Artef.	IR	Rec.	Metadata	Mult.	DA	Arg.	Sem	Inf	Ext
Who chaired the meeting?	1				1			1						
How did the meeting chair present the goals of the meeting?			1			-1	1				1	1		
Who was present at the meeting?	1				1			1						
What decisions where made?				1		-1	1				1	1	1	
Show me each recorded decision.				1		-1	1				1		1	
What action points where decided?				1		-1	1				1	1	1	
Where any actions attributed to me?				1		-1	1				1	1	1	1
Will there be a follow up meeting?				1		-1	1				1	1		
For action "A", show me the passage before the action was recorded so I know what it pertains to.			1			-1	1				1	1	1	
Did "John" (someone) give a status on "P?"		1				1	1	1						
Did anyone raise any issues?			1			-1	1				1			
For the issue ...I..., what actions where proposed?			1			-1	1				1		1	
Where any future risks identified?		1				1	1					1		
For the risk ...R... where any mitigating actions proposed?			1			-1	1				1	1	1	
Show me the passage where "John" talked about R.		1				1	1							
what were the key DECISIONS made?				1			1				1	1	1	
Everybody agreed these, or were they differences of opinion?			1				1	?			1	1	1	
What are the NEXT STEPS / ACTIONS?				1		-1	1				1	1	1	
To whom are they assigned & by when?				1		-1	1	?			1		1	
Am I supposed to DO anything after this meeting? (e.g. actions for me?)				1			1	?			1	1	1	1
When is the NEXT MEETING planned? (e.g. to follow up on action items)				1		1	1				1			
Who were the INVITEES? The PARTICIPANTS?	1				1			1				1	1	
What was the planned AGENDA?	1				1							1		
The effective list of TOPICS DISCUSSED?		1				1	1				1	1	1	
Where can I find the complete MINUTES? (to read & use full text search)	1				1	1	1	1				1		
Detailed agenda and key topics of the discussion		1			1	1	1					1	1	
Objective of the session		1			1	-1	1				1	1		
Name of the facilitator/leader	1				1			1				1		
Names of the participants and their roles	1				1			1				1	1	
Subject/Project status key points		1				-1	1				1	1	1	1
List of issues			1			-1	1				1		1	
List of agreed solutions and actions				1		-1	1				1	1	1	
Participants mood meter (scale Low/1-10/High) and reasons if low score			1			-1	1					1	1	
What would be required from me				1		-1	1	?			1	1	1	1
Next steps				1		-1	1				1		1	

Annex2.

Set	Information required most in order to answer the queries (unless specified otherwise, recordings are also required)	Prevailing query category	Freq.	Role of IR			(%)
				N/A	Insuff	Suff.	
IM2	Metadata, Argumentation, Semantics, Inference How was X's contribution received by the others at the meeting?	process	25	14	11	0	9.26
	Argumentation, Semantics, Inference What decisions got made easily?	outcome	24	8	15	1	8.89
	Metadata, Semantics, Inference Who talked most during the meeting?	factual	13	11	0	2	4.81
	Artefacts, Argumentation, Semantics, Inference What problems still remain pending?	outcome	12	9	2	1	4.44
	Metadata, Argumentation, Semantics When did X contradict himself about the issue X?	process	12	5	7	0	4.44
	Argumentation, Semantics Show me the final decision made about issue X.	outcome	12	1	11	0	4.44
	Cross-Meetings, Argumentation, Semantics, Inference What important decisions have been made during the last 6 months?	process	11	3	8	0	4.07
	Argumentation What decisions have been made?	outcome	11	5	6	0	4.07
	Semantics, Inference Did any participant leave the meeting early?	factual	9	6	1	2	3.33
	Recordings only What did P say about X?	thematic	9	0	0	9	3.33
	Metadata, Cross-Meetings, Semantics, Inference Was X present in all the meetings?	factual	7	7	0	0	2.59
M.S.	Argumentation, Semantics, Inference what were the key DECISIONS made?	outcome	8	1	6	1	2.96
	Argumentation, Inference For the issue ...I..., what actions where proposed?	outcome	4	0	4	0	1.48
	Argumentation, Semantics, Inference, Metadata(?) What would be required from me	outcome	2	1	1	0	0.74
	Artefacts, Metadata, Semantics, Inference (no recordings) Names of the participants and their roles	factual	2	2	0	0	0.74
	Argumentation Did anyone raise any issues?	process/outcome	2	0	1	1	0.74
	Metadata Who was present at the meeting?	factual	2	2	0	0	0.74
	Argumentation, Semantics Will there be a follow up meeting?	process/outcome	2	0	2	0	0.74
	Argumentation, Semantics, Inference, External Where any actions attributed to me?	thematic/outcome	2	0	2	0	0.74