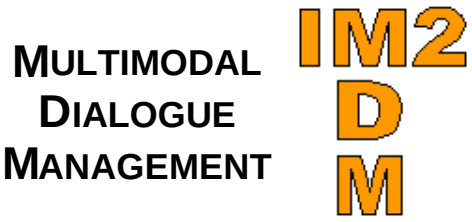
http://www.im2.ch



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

ISSCO/TIM/ETI, Université de Genève

Multimodal Interface Design for the Multimodal Meeting Domain: Preliminary Indications from a Query Analysis Study

Agnes Lisowska

IM2.MDM Report 11 – November 2003

Multimodal Interface Design for the Multimodal Meeting Domain: Preliminary Indications from a Query Analysis Study

Plan of the report

Abstract

The design of a multimodal interface for retrieving the content of multimodal meeting recordings from a database is a complicated process that requires careful consideration of the needs of real users of such an interface in real use situations. This report describes the results of a study in which potential queries to such an interface were elicited from real users, with the aim of developing a basic set of user requirements for interface design. Results and indications extrapolated from an analysis of the queries are discussed, and areas where further research is required are highlighted. Finally, the report discusses future directions to pursue in order to continue and refine the design process, and how the data set that was collected can be used for the design and evaluation of components other than the interface itself.

Chapter 1 Introduction

The IM2 project (www.im2.ch), funded by the Swiss National Science Foundation, proposes to create a database that stores recordings of meetings in an annotated and multimodal format. Users are expected to be able to use this database to access those recordings and find out information about the content of the meetings. However, the types of interactions that can occur between the users and the type of database described are expected to be richer and more complex than those occurring while accessing conventional databases. Not only is the information that is stored in the foreseen database multimodal and multimedia, but it is expected that the user's interaction with the system itself can be as well. The presence of multimodality at both the interface and database level poses new challenges for multimodal human-computer interaction.

When designing and developing any type of interactive system it is important to understand what that system will be used for, and the nature of the interaction between it and the user. This knowledge plays a key role in how the interface should be designed and implemented, and also potentially influences the design and organization of other system components as well. Because the goals of the IM2 project are research rather than a commercially oriented, current user and use case specifications are based on the intuitions of project participants rather those of real users. The study presented in this report, and the consequent analysis of the results, is intended as a first step in the process of defining design specifications based on real user requirements.

The remainder of this report is comprised of 6 chapters. Chapter 2 lays out the assumptions made when designing the study. Chapter 3 explains the importance and necessity of developing user requirements models and use cases, and their role in interface design. Chapter 4 gives details about the study that was carried out, while chapter 5 delves into the specific results and their implications on design. Chapter 6 explores the problem areas in the design of the study and the results, and how they can be addressed in future work. Finally, Chapter 7 summarises the results and implications of the study and explores directions for the future.

Chapter 2 How we view the problem

Because there are many possible ways to conceive of a system such as the one proposed for the IM2 project, we feel that it is important to first set out the specific assumptions that the work here takes into consideration, and provide a simple analogy to help the reader understand more clearly how we envision the functionality of the interface (and to a larger extent the system) from the user's perspective, and some of the a priori problems that we believe will be encountered during its design and implementation.

2.1 Assumptions

2.1.1 About the domain

The domain being considered is that of accessing and retrieving information from and about recorded multimodal meetings. It is assumed that a database exists, henceforth referred to as the meeting database, which contains recordings of meetings, where the data in those recordings is stored in multiple media formats (text, images, documents, videos etc) and annotated to take into consideration both the content of the recordings (topics, decisions, discussions, statements etc.) and the modalities contained within (speech, gesture etc.).

2.1.2 About the multimodal meeting content retrieval (MMCR) system

The multimodal meeting content retrieval system is taken to mean the meeting database connected to a user interface and a processing and retrieval component, all of which are tailored to suit the information finding needs defined for the domain. The work in this report focuses predominantly on the interface and the processing and retrieval components described in more detail below.

The Interface

There are many ways in which one could imagine interacting with a database containing information about meetings – for example being able to browse the information in a database visually or access specific fields and values in the database using traditional methods such as SQL - and thus as many possible different types of interfaces. This work will concentrate on directed information seeking through a multimodal interface. This means that the interface should allow the user to ask specific questions about the data in the database using any combination of modalities from the set that is available to them.

There are four important underlying assumptions being taken into consideration:

1. The user is not familiar with the structure of the database (and should not need to be familiar with it in order to be able to find the information they need)
2. The user has a specific goal in mind when they are interacting with the system – there is some specific information that they are seeking
3. The user should be able to use any means that they are comfortable with to achieve this goal, whether it be through natural language, more traditional point and click methods or a novel combination of modalities (although for the purposes of this report we are assuming that natural language will be a principal modality)
4. The system does not necessarily provide the user with a precise answer to a question, but rather provides the data or excerpts of data that contain the information that answers the user's questions.

What the most appropriate modalities to include in the interface are, given the potential tasks and needs of the users, is one of the long-term goals for which the work described in this report lays the foundations.

Information processing and retrieval
When we refer to processing we are referring to the mechanism that analyses the input from the user (independent of what modality that input is in) and translates the information contained in that input into a defined internal search language. The retrieval mechanism then executes the search. One of the research issues that will have to be addressed is the scope and nature of the search space over which the search is run. The multimodal nature of the data in the database implies that the search will need to be executed on annotations of the data in addition to the textual content of the data itself as many of the multimodal features will not be present in the textual content. Moreover, depending on the nature and extent of the annotations and the requirements imposed by the domain, it may be discovered that running the search over raw textual data may not be necessary at all.

2.1.3 About the users

Throughout the last sections we have referred to users. For any kind of interface design it is important to define who potential users of that interface will be, and what they will be using it for, at the very least in broad terms. (The need for a much more detailed analysis of users is elaborated in chapter 3) The MMCR system can conceivably be used for different purposes by different people. For example for testing purposes by researchers who are developing software to create the data and/or the annotations that are stored in the database. However, the work described in this report considers as the users of the system only those who use it in order to find out something about the content of the meetings themselves. An example of such a user would be an employee of a company that records and stores their meetings in the MMCR system. This employee has missed a meeting that they were supposed to attend and, as they are in a hurry do not have time to view the whole meeting, they use the system to find out about the things that are most important to them, such as what decisions were made.

2.2 The archivist analogy

Interaction with an MMCR system can be thought of as interacting with an old fashioned human archivist. The archivist works in an archive room stacked with various types of documents that are stored according to a particular system with which the archivist is familiar, and they can find documents quickly and efficiently. When someone wants needs a specific piece of information they go down to the archive room and tell the archivist what information they are looking for. The archivist then goes into the archives and retrieves the documents that contain that information. Furthermore, people

may want to know about information in varying degrees of detail, depending on what they intend to use that information for and who they are. The archivist usually knows that from the way in which a request is made (or who the person is) the level of detail in which the answer should be provided.

There are two important things to note which have an impact on how the MMCR system is viewed:

1. The archivist does not directly answer a question, rather their role is to provide the documents which contain it
2. If the archivist is unsure about what is being requested, they can ask for a clarification or more specific information

The envisioned MMCR system replaces the human archivist (the interface, processing and retrieval) and the archives (the database). This implies that as many as possible of the tasks and functionalities performed by the archivist should be transferred to the system, and those that cannot be transferred directly should be compensated for in a manner that minimally inconveniences the user.

An immediate concern in such a transfer is that many of the actions performed by the archivist involve things that human beings do naturally and often unconsciously such as

- understanding what is being asked of them (interpreting language)
- understanding how different modalities are used together to express information (for example knowing when a gesture is directly related to something that is being said – someone saying "I work in that building" while pointing to a building across the street)
- knowing what an appropriate response to a given question is (If someone asks "What was the agenda" people know that it is the topics in the agenda that are of interest and not that it was a piece of paper)

Additionally, human beings

- have knowledge about the real world (for example that a meeting has participants and topics etc)
- make assumptions about the abilities and desires of their dialogue partner (if a person knows that their dialogue partner is not proficient in the language being spoken, that person is unlikely to use uncommon words or complicated sentence structure unless it is absolutely necessary)

All of these points are actually hard problems for computers (and open research issues in their respective areas), and yet they seem to need to be taken into consideration when designing the interface. Consequently, one of the first steps in the interface design process will be to determine whether these elements actually are necessary, and the extent of the role that they play in human-computer interaction in the chosen domain. Additionally, means by which to overcome these technological gaps in as smooth and transparent a manner as possible should be investigated as we believe that their presence does not mean that the design and development of a robust and satisfactory system must wait for them to be filled first.

Human beings, when communicating with other people, make accommodations to account for differences between themselves and their dialogue partner [4]. For example, if a person is speaking to a child, they know that different vocabulary needs to be used to explain things than if the same person were explaining the same things to an adult. The same is true for explaining things to people from different areas of expertise or backgrounds. Additionally, if a person is communicating with someone who is not also trying to make accommodations and modify their behaviour to suit the situation, they often quickly become frustrated and give up, or reach a conversational impasse. We believe that it is not unreasonable to assume that humans will be willing to make similar accommodations for a computer system. But, just as between people, the accommodation must be mutual. Determining the scope and breadth of the mutual accommodation is another research challenge for interface design in this domain.

Chapter 3 User requirements modelling and use cases

3.1 User requirements modelling

It is important to note that in the framework of this report *user requirements modelling* is not equivalent to the development of *user models* as suggested in literature on the subject [6, 7]. In the latter case, user models are often used to help fine tune a system to the needs of a particular user, and in the case of natural language interfaces, to generate an appropriate dialogue model. What we are attempting here is more general, and follows more closely user requirements modelling as described in [5], further elaborated below.

User requirements modelling is an important part of any software design process, and is particularly crucial to interface design. There are two sides to the requirements model, both of which need to be given equal consideration in a well designed system – the functionalities that required in the underlying software in order for the user to be able to accomplish their tasks, and the functionalities that are required in the interface so that accomplishing this task is made as easy and natural as possible. Thus, user requirements modelling includes determining who potential users of the system are, what their roles are, the tasks that they will use the system to accomplish, what the important elements of those tasks are, and how they want to be able to use the system.

Having a well defined model of the user and what they require of and expect from the MMCR system can help in the design of mechanisms for overcoming some of the problems caused by technological gaps, as user models can help to tailor and tune the system to be as accommodative as possible, allowing for smoother human-computer communication even in the presence of technical gaps.

One of the key elements of user requirements modelling is to develop the models based on real potential users of the system. Given the domain in question, it seems natural to assume that those involved in the project itself qualify as real users, since they too go to meetings and might want to know about their content. While this is true, it is important to bear in mind that developers' views of a particular application and its uses are not always representative of actual needs. This is a common problem in software design, and is equally applicable to interface design. Users not involved in the development process view the system top down. In other words, they are exposed to an interface, a surface representation that ideally allows them to understand and work with what lies beneath. It is through interaction with this interface that users slowly learn the capabilities of the system and in so doing, the limitations that the system imposes on the satisfaction of their own particular needs. Those involved in the design, however, tend to view a system bottom up. They are aware of what the system can and cannot do at the most basic level because they have built it (or intend to), and are familiar with the technological limitations that constrain the development.

Careful user requirements modelling facilitates the bridging of these two points of view by drawing out real-world user requirements that are constrained by technical knowledge. In turn, these requirements can then be fed back into the design process to ensure that the software and system architecture is developed in such a way as to accommodate for them without compromising the technical capabilities of the system. Such a cycle is commonly called the iterative design process in human computer interaction literature [5]. In order to create a system that is both innovative and satisfies the needs of potential users, the iterative design process should be taken into account. The study outlined in this report is a first step in this process.

3.2 Use Cases

The development of use cases is in fact a sub-task of user requirements modelling. Use cases are intended to determine the types of tasks that real users might want to use the system for. [5] They can range from the very broad, for example reviewing the progress of a project over time, to the very narrow, such as finding out what a particular person said about a particular topic during a specific meeting. Use cases are important because they help to constrain the set of tasks which system design needs to take into account, and help make analysis of user requirements easier by providing a more coherent and refined set of results across participants.

Furthermore, use cases are not only useful in the design process itself, but also in the evaluation of a system at various stages of development. One must bear in mind though that while the use cases for the design process should ideally come from the users themselves because the aim is to design a system that is useful in real situations, those used for evaluation purposes do not have the same constraint. Developers may want to test a very specific aspect of the functionality of a system, and in order to do this while controlling for other variables, the user may have to perform a task that is much more specific than would be found under typical use circumstances. For example, to test the ease of use of a particular retrieval mechanism a task such as searching for a particular image in the database and then timing how long it takes the user to find that image may be specified. However, care must be taken in the development of evaluative use cases as they should remain rooted in the realm of real use cases. Otherwise, the evaluator risks obtaining results that do not reflect their domain or the usefulness of the application to the domain.

To date there have been no use cases determined for this domain that come from users not involved in the IM2 project, and those that have been specified are discussed briefly in section 4.3. While these use cases may in fact correspond to real use cases, it would be a useful exercise for the future to determine a set of at least general use cases for the multimodal meeting domain that are agreed upon and taken into consideration by all IM2 partners.

Chapter 4 The study

4.1 Previous work and motivations for the study

In October 2002, sets of possible queries to a multimodal meeting database were elicited in an effort co-ordinated by ISSCO, at the University of Geneva. The result was a set of 176 queries conceived by various IM2 partners, an initial categorisation and analysis of the query content[1], and the definition of three potential use cases. These use cases were: someone who has missed a meeting, someone who was present at a meeting and wants to review it, and high-level project management and employee tracking.

There are two difficulties posed by the use of this data set for the development of user requirements models. The first is that the participants were all involved in the IM2 project. As previously discussed, this means that their results cannot be assumed to be representative of the general user population. The second problem is that while some of the queries were constrained by the very general use cases mentioned, it was not the case that all were, which makes coherent and controlled analysis difficult.

As has been argued earlier, proper interface design must take into consideration carefully developed user requirements models. The study described in this report builds on and extends the original effort, to controlling for various variables and providing a more representative, structured and analysable data set.

4.2 Goals of the study

The study discussed in the following sections had several overall aims:

1. To determine the types of things that users want to know about – what elements of meetings are important to users
2. To determine how users ask about those things – what language is used
3. To determine what modalities users might want to have available in the database – what needs to be recorded, annotated and made available for retrieval
4. To control for and assess the impact of biasing by including both participants who are involved and uninvolved in the IM2 project
5. To obtain a larger, more balanced and representative data set

The goals described are intentionally rather broad. As this is only the first step in what we intend to be a much larger and ultimately iterative process, the initial aim was to obtain as wide a range of results as possible for the set of use cases selected. The hope was that these broad results would give

[1] These queries and the analyses mentioned can be found at http://issco-www.unige.ch/projects/im2/mdm/

preliminary indications about trends in the data and possibly uncover new avenues to explore in future experiments and analyses.

4.3 Methodology

The studies were conducted by means of a questionnaire, the final version of which can be found in Appendix A. The questionnaire first briefly describes the IM2 project itself and its general goals, then goes on to present four possible use cases in which a potential user might be interacting with the system. The four use cases were derived based on the three use cases defined in work carried out during the October 2002 effort and were chosen to attempt to maximise the variety of queries that could be made. They are outlined briefly below:

1. A manager is tracking employee performance
2. A manager is tracking project progress
3. An employee has missed a meeting about a project that they are involved in
4. A new employee wants to learn about a project that they will be starting work on soon

The final part of the questionnaire asks users to provide some background information about themselves, in order to be able to determine if any anomalies in the data can be explained by the background of the participant(s) from whom the data came. The questionnaire was sent and received by e-mail, to participants both involved and uninvolved with the IM2 project.

Initially, a pilot study was run to test the sufficiency of the questionnaire. The results of the pilot study were positive and the contents of the questionnaire went unchanged when the full scale study was run. The only difference between the pilot study and the full scale study was that in the pilot study the participant was assigned a particular use case and was not aware of the other three potential cases, whereas in the full scale study, the participant was free to choose among the four use cases and was thus aware of the other possible contexts of use of the system. Given that the difference between the two studies not thought to be of the nature to influence results, and in order to increase the number of queries gathered to have a more statistically viable data set, it was decided to pool the results of both the pilot and the full scale study into one set. Thus, all references to the study from this point on refer to the combined set of queries from the pilot and full scale studies. Figure 1 shows an overview of the results obtained.

	Study Set				
	Full scale		Pilot		Combined
	IM2	Non-IM2	IM2	Non-IM2	
# of responses	11	9	3	5	28
# of queries	109	76	43	69	297

Figure 1: Overview of results obtained

Note: approximately 4% of all queries gathered can be considered unanswerable by the envisioned system.

4.4 Inter-case differences

An interesting aspect to investigate when selecting design requirements is whether user behaviour changes if the goal that they are working towards changes. In the case of an MMCR system such as the one under consideration, the general task itself (looking for information about or within a meeting) does not change, but the way in which that information is used by the user might, and it is that difference that is the subject of the inter-case analysis presented here.

The first thing to note is that among the users who were given a choice of which use case they were working with, the case where the participant was pretending to be someone who has missed a meeting (use case 3) was the most popular, 2-3 times as frequently selected as the other cases. This is interesting for two reasons. The first is that future elicitation studies for queries for the managerial use cases may require people in managerial positions as participants since it may be too difficult for non-managers to effectively put themselves in that role. The other is that the missed meeting use case may be more representative of what people imagine an MMCR system being used for, since even

among the non-managerial use cases, the missed meeting use case was chosen twice as often as the new employee use case.

In terms of differences between the types of queries that were asked for each particular use case, there do not appear to any marked differences beyond what might be expected. For example, it is expected that in the project management use case, there would be more questions regarding the project than in the person management use case, and this expectation was verified. The only case that did stand out was in fact the person management use case, where the questions tended to be more abstract and subjective. For example '*Who has the most positive/negative attitude with respect to the other participants?*' However, it should be kept in mind that this apparent lack of disparity between the use cases may be due to insufficient data from which to uncover real differences. Further analysis will be necessary as more experiments are run and the data sets for each use case increase.

4.5 Biasing

One of the primary considerations when performing this study in comparison to the October 2002 effort was to try to control for and investigate biases that may surface when users are also directly involved with the IM2 project and therefore aware of its possibilities and limitations, and their own vested interests in it. To this end, the current study involved a total of 28 responses, 14 of whom were involved in the IM2 project and 14 completely uninvolved with the project and whose only knowledge of it stemmed from the brief description provided in the questionnaire.

Only one evident, but broad, case of biasing was discovered. Participants who were involved in the IM2 project tended to pose more queries about agreement/disagreement, acceptance/rejection, decisions, criteria, presentations, proposals, suggestions, non-results and open questions while those uninvolved in IM2 tended to pose more queries about statements, dates and tasks. Both groups however were equally interested in discussions, arguments for or against something, documents that accompany the meeting, information about a project, and the various topics that were covered in the meeting. Importantly, there was no indication that either group tended to pose more questions about very general classes of elements (more detail is available in section 5.6.1), rather, the differences lay in the number of queries for a particular element in each group.

The type of biasing described above is important because it impacts the design and development of annotation sets for the data in the database and on query processing requirements for the system. It will have to be investigated more carefully in future studies and as more information is gathered, to determine whether the elements involved in the biasing are in fact not really of interest to general users, or whether the non-interactive format of the questionnaire played a role in the biasing by not giving users uninvolved in the IM2 project enough of an indication as to the types of elements that they could in fact ask about (the IM2 participants would likely have known that they could ask about them because of their familiarity with the intended system).

Chapter 5 Implications of the study

An analysis of the queries brought to light several areas and issues that need to be taken into consideration during the interface design process. This chapter gives an overview of what those issues are, how they influence the design, and which of them demand further investigation.

5.1 Inter-relatedness

One of the first things that became apparent during the analysis of the collected queries is that the aspects that were being analysed, and by extension the elements necessary for interface design, are highly connected and inter-dependent, and thus must be considered together when planning the interface and, on a more general level, the overall design of the system itself and the integration of the individual components. The diagram in figure 2 highlights some of the main areas and the relationships between them.

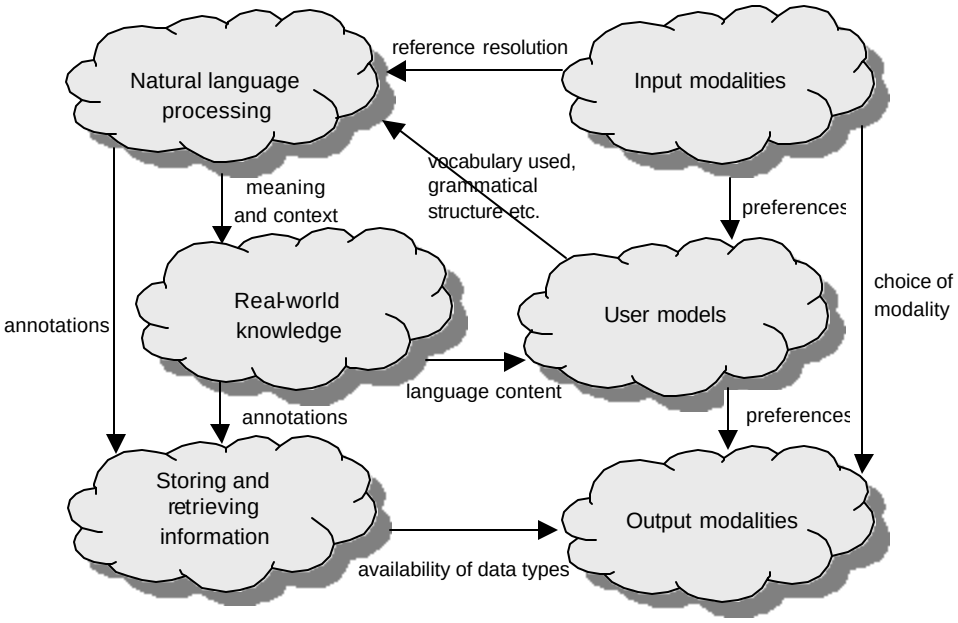


Figure 2: Areas of inter-dependency between components involved in the interface and user-system interaction

5.2 Why this particular type of analysis

The queries gathered during this study predominantly gave indications about the language processing and database and data annotation aspects of system and interface design (The reasons for this are explained in detail in chapter 6). Consequently, the analysis described in this chapter was primarily undertaken with the following objectives in mind:

- To assess in a broad manner linguistic phenomena that would need to be accounted for when handling natural language input in the interface (independently of whether the input is spoken or written)
- The types of processing that would need to be done on the data or annotations in the meeting database in order to satisfy user needs for retrieving information from the system
- To uncover phenomena particular to the type of interaction envisioned that would have to be handled either at the interface or language processing levels.

5.3 Implications for natural language processing

We have already mentioned that in the scope of this study we are working under the assumption that natural language will be a primary input modality for the interface, but we have not specified which form, spoken or written, it can take as we believe that both forms are likely to be used in the final interface. A consideration that evolves from the presence of both forms in the interface is that there are likely to be significant differences between queries in written and spoken language [3], and while there may be overlap in developing tools to handle both types of language, it cannot be assumed that the two can be exhaustively handled in the same way. Consequently, one area that will require further research is the extent of the differences between spoken and written queries, and how those differences should be handled within the MMCR system. However, as the questionnaire elicited only written queries, the analyses that follow will focus on written language.

5.3.1 Relative terms

The use of relative terms in queries poses a problem for natural language processing because relative terms are often used to qualify (and at times quantify) important nuances that help determine possible answers to a query. In order for the system to provide appropriate information, it must have a means for understanding what those nuances are and applying

them in the search process. There are two general categories of relative terms that were found in the queries, quantifiable and unquantifiable.

Quantifiable

Quantifiable relative terms are those where the precise information that is required to resolve the relativity can be determined based on information about one or more known points to which the sought after information is relative. This category in itself can be broken down into two additional categories, definitely resolvable relative terms and scalably resolvable relative terms

1. **Definitely resolvable relative terms**

   Definitely resolvable relative terms are those where something is relative to a specific and known point in time and relativeness can be determined based on that fixed point. For example, if the meeting that a person is viewing is from November 11th, 2003 and they ask to see the next meeting, then the fixed point is the date of the current meeting (November 11th, 2003), and the relativeness of the term *next* can be reduced to a date that is the closest to that fixed point but comes after it on a timeline.

   Examples of definitely resolvable relative terms found in the queries are: *next, last, following, new, further, future, first, final, past, beginning, so far, one before last, other, early, still, before, previous, external*

2. **Scalably resolvable relative terms**

   Scalably resolvable relative terms are those where the resolution is not fixed to a particular point but can be assigned a specific upper and lower bound. Often the resolution is in fact be a range of points (a scale) rather than a single point. For example, in the query '*Did someone briefly present the project?'*, the relative term *briefly* can be defines as 2-5 minutes, or 0-30 minutes, and might depend on the context (if the meeting was very long, a couple of minutes may not be sufficient time) or the person (some people consider 2 minutes brief, while for others it may be 15).

   Examples of scalably resolvable relative terms found in the queries are: *main, usually, most, recently, initially, often, majority, long, prolonged, short, briefly*

Unquantifiable

Unquantifiable relative terms are particularly hard because not only are they heavily dependent on the context in which they are being used or on the person who uses them, but they often cannot be defined, neither numerically nor linguistically, in any consistent manner. For example, in the query '*What important decisions have been taken during the last meeting?'* what is considered *important* may vary from person to person and even for the same person from situation to situation. It is impossible to foresee all possible contexts of importance for all possible users, and thus becomes impossible to determine computationally what piece of information is being referred to. At best, one could develop a kind of rating scale based on frequency of events, but even that would likely only be successful for a portion of the relative terms. Thus, one challenge will be determining how to handle queries that involve these terms, either by developing novel ways to resolve the relative term itself, or by designing an interface mechanism that allows the user to help the system specify what they consider the relative term to mean.

Examples of unquantifiable relative terms found in the queries are: *the, this, that, there, important, unusual, intense, controversial, actively, they, main, major*

5.3.2 Abstract concepts

Abstract concepts such as *tension*, *dominance, constructive criticism* and *initiative* pose problems for natural language processing because they often cannot easily be defined, and their definition can be slightly (but sometimes in an important way) different for different people. If one takes for example the query *Who made constructive criticisms about the*

proposal?' how is the system to know what constitutes a *constructive criticism*? What constitutes a constructive criticism can vary for different people. One person might consider that an important part of constructive criticism is the suggestion by the critic of a possible solution to the 'problem', whereas another person might believe that the core of a constructive criticism is rather the pointing out of legitimate problematic areas and the explanation of why they are problematic that is the core of what a constructive criticism is. Such differences influence both what would need to be annotated in the data, and what might be returned as a possible response to the example query.

Some examples of abstract concepts found in the queries are: *paying attention, involvement, negative, positive, constructive criticism, degenerate, dominant, sleeping, initiative, leadership responsibilities, tension, valuable.*

5.3.3 Indirect questions

The way in which a query is phrased can pose its own problems when determining what type of information needs to be found and presented to the user. If the question is in fact a command (e.g. '*Show me the passage where X made a proposal about issue Y'.*) or a direct question (e.g. '*Who was at the meeting?'*) then the type of answer that is required is clear. Fortunately, commands and direct questions account for approximately 85% of all queries. The other 15% however, fall into the category of indirect questions, where, depending on the context or the circumstances, the desired answer may be more than the literal answer to the query. Take for example the query '*Have I been assigned a new task*?' From a strictly grammatical point of view, the answer is either a yes or a no. However, humans know from real life experience that it is very likely that the person asking the question actually wants something more than a yes or no answer – that they also want to know what that new task is. While one might assume that the person could request the additional information as a follow-up question if the answer to the indirect question is positive, and that this would not significantly harm the communicative process, it makes the interaction between the user and the system less natural.

Of course, the MMCR system does not have the same type of knowledge or experience as a person to determine when the appropriate answer is an implied sought after answer, but mechanisms could be developed which recognize cases where there is a need for more information and provide the additional information immediately. In this way, the user can choose to disregard the additional information if they did not want it in the first place, but if that information is what they actually intended by posing the query, then they do not have to go through any extra steps to get it.

Dalhbäck et al. [4] question whether users of natural language interfaces will pose indirect questions given the fact that a user is interacting with a machine rather than another human being, and believe that only empirical studies can provide an answer. Following this thought, it will need to be investigated whether users continue to use indirect questions when interacting with an actual system, as the questionnaire format and the fact that the users subconsciously knew that the questions would be read by a human, may have influenced the use of indirect questions in the study.

5.3.4 Reference resolution

Reference resolution has long been an interesting problem for computational linguists, and while it typically applies to longer texts or pieces of discourse and is thus more likely to be thought of as a problem for data annotation rather than the interface, the queries gathered indicate that at least to a limited extent, reference resolution will also have to be accounted for in the interface in order to be able to understand what information is being sought after in the query.

In most cases of reference resolution at the interface level, the resolution involves queries that refer to information contained in a previous query, and thus the resolution relies on correctly identifying which previous query is being referred to, and more specifically, what part of the information in it. In some cases, this type of resolution, given the appropriate algorithm, can be

done quite easily. Take for example the query *'Did Dave talk at all during the meeting?'* followed immediately by '*Did he make any suggestions about the design?'*. It is fairly easy to resolve that *he* refers to *Dave* in this case. More complicated cases however are those where there are intermediary questions that contain other people who might be possible referents, or cases where the intermediary dialogue is so long that the intended referent is too far back in the dialogue to make the connection. There must be some mechanism in place in the system that can make resolutions independently of user input whenever possible, and in cases where it cannot make the resolution, the system should handle the problem in a fluid manner that does not interfere with the natural flow of communication between the user and the system.

A special case of reference resolution in human-computer interaction is the resolution of the use of "I" in interaction with the system. In the query '*Have I been assigned a new task?'* the system needs to resolve who "I" is referring to in order to be able to process and answer the query. Having the system ask the user to identify themselves would likely interrupt the natural flow of a dialogue, and it is unreasonable to ask the user to talk about themselves in the third person when interacting with the system. Fortunately, given the framework of the IM2 project, we can work on the assumption that if the MMCR system were a commercial product, it would have some kind of password, or verification system for users integrated into it. Based on that password or verification, the system would be able to 'know' who the user is, and thus resolve "I" transparently.

5.4 Real-world knowledge

Human beings use a significant amount of real-world knowledge to solve even the simplest problems. For example there are several things that a person knows when answering the query *'What was the agenda?'*. First of all, they know that an agenda is related to a specific meeting, and serves the purpose of structuring the meeting. Additionally, they know that an agenda has content, that the content has a specific meaning, and that it is that content that is usually of interest to someone, and not for example the form of the agenda (a specific type of document, or the fact that it was spoken). All of these factors contribute to the person being able to answer the question in a sensible way.

The encoding of real-world knowledge for computational purposes, often referred to as ontology creation, is a notoriously difficult problem and there are two primary reasons for this. The first is that human beings gain knowledge not only through being taught, but through personal experience as well. Computers, on the other hand, cannot really gain knowledge from personal experience, at least not to the same extent as human beings do. As a result, any knowledge that a computer has is knowledge that has in some way been specified by software developers. The Cyc effort (www.cyc.com), begun by artificial intelligence pioneer Doug Lenat, has been attempting for the last 17 years, and continues to this day, to encode general world knowledge. Clearly, this is a non-trivial task. Consequently, the notion of trying to encode all world-knowledge for an MMCR application is unreasonable.

However, recent trends in ontology creation have been leaning towards the development of domain-specific ontologies - ontologies that contain quite detailed information about a specific area, and very little about everything else. We believe that this approach can be applied to the MMCR system to facilitate language understanding and information retrieval. If this approach is adopted, then the first step is determining the subset of knowledge that is appropriate for the multimodal meeting domain. The queries elicited in this study form the basis for a seed ontology by providing two important elements necessary for ontology development:

1. The concepts that users are interested in when interacting with the MMCR system (document types, discussions, decisions, people, roles etc) and more generally, the vocabulary they use to name those concepts. For example, there were 14 different possible types of 'document' found in the query set: *lists, reports, diagrams, flow charts, pictures, graphs, slides, presentations, posters, recordings, videos, summaries, agendas, meeting minutes*.

2. The relevant relationships between those concepts (decisions are made by people, an agenda can be a thing etc). Figure 3 shows a possible skeleton ontology for the entry 'project'.

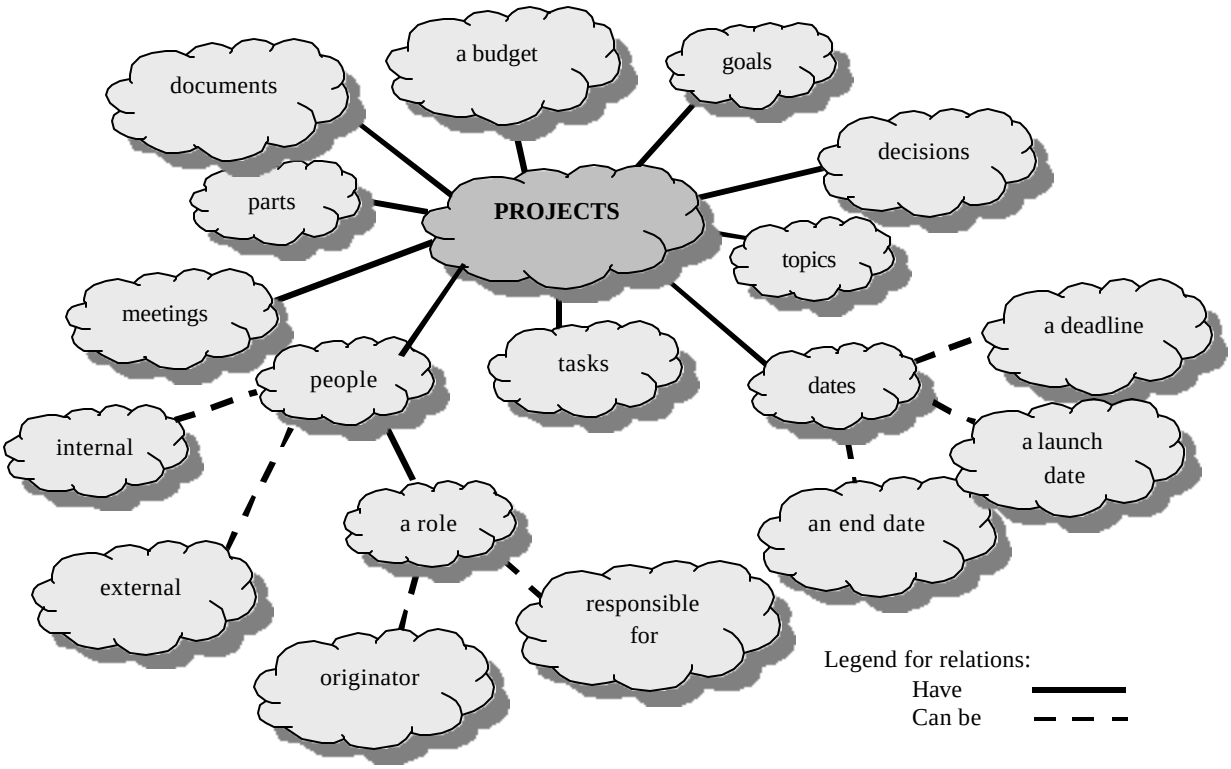


Figure 3: Possible skeleton ontology for the entry 'project'

Over time, as new queries are collected and the user models are refined, the ontology can be expanded and/or modified to better reflect the domain.

5.5 Beyond simple search and retrieve

Analysis of the queries also revealed that finding answers to them may involve more than just a simple search and retrieve on even annotated data. In approximately 40% of the queries, it was necessary to perform some form of background processing in order to find an answer. This section gives an overview of the types of background processing that were most commonly needed to answer the queries.

5.5.1 Spans over meetings

Around 15% of the queries (and half of all queries explicitly involving meetings) require looking at, or gathering data over, several meetings in order to find an answer. Examples of such queries are '*Was X present in all the meetings?', 'How often have there been meetings on the project?', ' Were any issues discussed/resolved relating to discussions of previous meetings?'*. The implication is that related meetings need to be grouped in a meaningful way and that when answering a query, whether the answer is contained in a single meeting or requires looking at several meetings needs to be taken under consideration.

5.5.2 Mathematical calculations

16% of all queries require the system to perform simple mathematical calculations in order to find an answer. Furthermore, some queries such as '*When is the next meeting?*' require calculating differences between dates and times (these types of calculations are necessary for about 14% of all queries), while others such as '*Who talked most during the meeting?*' require simple mathematical calculations such as differences, averages etc. The query set provides specific indications as to what types of calculations users expect the system to be able to perform, and so algorithms to perform those calculations should be incorporated into the system.

5.5.3 References to events in the past

In order to answer some queries connections or relations between events in the present and those that happened in the past need to be identified. Examples of this are follow-up discussions, answers to questions that had been previously posed etc. Queries such as '*What was discussed that was followed up from a previous meeting?'* or *'What issues have been rediscussed in the current meeting'* pose a double difficulty for the system because not only must it first be able to recognize that terms such as *rediscuss* and *follow up* make implicit references to things that have already happened, but it must then also find the correct event in the past based on information about a more recent or current event. The extent to which the presence of specific types of annotations in the database and appropriate entries in an ontology can help with this problem will have to be investigated.

5.5.4 References to non-events

References to non-events pose particular problems because they involve deducing events that did not occur (non-events) from those that did and/or should have occurred. The difficulty lies in the fact that the information on which these types of deductions are based may not always be available in the database. For example, answering the query *What issues were left undiscussed during the meeting?'* requires knowing what issues should have been discussed (for example from a list of issues based on an agenda) and those that were in fact discussed (for example from a list of issues from the topic segmentation of the meeting), and then looking at the differences between the two lists (also taking into consideration other factors such as vocabulary differences between entries in the lists where the semantic meaning is actually the same). Similarly, the query *'Which people missed meetings which they were supposed to attend?'* requires knowing who actually did attend certain meetings and who should have attended them, and again comparing the lists. In order to be able to answer these types of queries, the system must be able to recognize such queries and either have algorithms in place to retrieve the information and make the necessary deductions, or inform the user that it is incapable of answering this type of query.

5.5.5 References to non-dialogue events

Certain queries involve recognizing events that will likely not be referred to in a dialogue and thus need to be identified directly from the visual or auditory recording of a meeting. For example, retrieving answers to queries such as '*Who left the meeting before the end?'* and '*Who was sleeping during the meeting?'* requires directly annotating events like leaving and sleeping. While these types of events aren't numerous (they were found in approximately 3% of all queries), they should still be taken into consideration given the multimodal nature of the domain. While it is unreasonable to assume that all such events can be foreseen and annotated in the data, the existing query set and those gathered in future studies will be helpful in determining the level of detail and the types of events that will be most useful for the system to recognise and annotate.

5.6 Implications for data storage and annotation

An important aspect of processing the queries and finding answers to them is establishing that the connections between the modalities (language, pointing, etc.) used to express the query and the way in which the information being queried is stored is appropriate for the task. This relationship is in fact circular in the sense that how a query is processed depends on how the information is stored in the database, and how the information is stored in the database should depend on the overall query processing capabilities of the system.

What is actually stored in the meeting database should depend on what users want to use that information for, and how they want to view it. Determining this involves carefully considering what kinds of media files (.jpeg, .doc, .wav, .ppt, .pdf etc) users request and, on a deeper level, what types of information contained in those media files they ask for – in other words, how the media files need to be processed and annotated in order for information retrieval to be both efficient and sufficient. The latter point can be further divided into the physical and thematic dimensions of the meeting. The physical dimension involves annotation of events such as people's movements, when they speak etc. The thematic dimension pertains to the annotation of the thematic content of the meetings – what is being talked about (topics) and how (discussion, decision…). As the results of the study gave little indication about the physical dimension, it is the implications of the thematic dimension that will be

discussed in the remainder of this section. Three primary areas were highlighted by the results of the study – the types of annotations necessary, the degree of annotation and relationships between meeting elements.

5.6.1 Types of annotation

The types of things being asked about can be divided into two general groups:

(a) Things that involve and/or describe interaction among participants in the meeting

> Agreement/disagreement, acceptance/rejection, controversy, decisions, discussion, arguments for/against, opinion, emotion/reaction, criteria/requirements, presentation, proposals, questions, statements/assertions, response, suggestion

(b) Things that involve elements from the meeting domain itself

> Date/time, deadline/timeline, documents, project, task/responsibility, topic, summary, open/unresolved questions/problems, solutions, results, problems, non-results, meetings, people

There are two important things to note about this very high level classification. The first is that three times as many queries were found to fall under category (b) than category (a), and the second is that depending on the context of the query, the two groups are likely not disjoint.

5.6.2 Degree of annotation

Annotations on any piece of data can range form the very specific (who said something, the exact times at which something was said etc.) to the more general (some piece of text is a discussion, an argumentation for some point, what was being talking about). The degree of annotation is thus the level of detail at which annotations need to be made in order to satisfy user driven retrieval needs. At this stage, it is still too early to determine which degree of annotation, and more generally which types of annotations within the different degrees, will prove to be the most useful. However, it is important to bear these distinctions in mind in the early stages of development, particularly if the annotation process is to be automated. For example, while not many queries directly involved low-level annotations such as the annotation of questions, it may be found that these low level annotations are in fact necessary in order to derive or identify candidate cases of higher-level annotations such discussions or decisions, which were much more common in the query set. One of the key tasks in this area will be to find the right balance of necessary and derivable annotations that will make the automation process most efficient (both in terms of time taken to do the annotation and the needs imposed by system requirements).

5.6.3 Relationships

We user the term *relationship* to refers to any link between two pieces of information that would have to be annotated or otherwise derived in order to be able to answer a question or find a piece of requested information. 134 separate relationships were identified in the query set, composed of 36 different elements. Moreover, a single query often involves more than one relationship, and the combination of relationships can be used to help reduce the search space when finding an answer, by providing a more constrained set of search parameters. The list below shows, in order of frequency, the 13 most common relationships found in the query set.

1. meeting-person	2. topic-meeting	3. topic-discussion
4. person-statement	5. meeting-discussion	6. person-talk
7. meeting-decision	8. topic-said	9. person-person
10. topic-decision	11. topic-person	12. person-suggestion
13. person-document		

5.7 Additional phenomena

5.7.1 Multi-part queries

Multi-part queries are queries that contain the conjunction 'and' or the disjunction 'or' and thus in essence require two searches, but for possibly closely related sets of data. Examples of such queries are '*What has each person promised to do and when did they promise it?'* and '*What decisions were made only after intense or prolonged discussion?'.* We hypothesize that these types of queries are more likely to be found in a cases where written language is a primary input modality, but this hypothesis will have to be verified in the course of testing with an interactive system.

5.7.2 Chain queries

Chain queries are a sequence of queries where each query that is posed is dependent on the answer to the previous query. For example, '*Was my proposal put forward?'* followed by '*How did people react to my proposal?'* just after the answer to the former query. Although not many instances of this type of query were found in the current data set, we expect chain queries to appear more frequently in actual interaction with the MMCR system since they seem more suited to an interactive, dialogue-driven environment. In fact, several study participants commented on the fact that had they been interacting with a real system the queries that they would pose would most likely depend at least to some extent on the answers that they received to previous queries. Actual interaction with an MMCR system would be needed to verify this hypothesis.

5.7.3 Number of terms and search criteria

The number of terms used in the queries, and thus the number of possible search criteria available to the system, is also an important factor. In this work, search criteria are considered to be terms that can be found in the ontology or annotations based on which data can be extracted. In general, almost 75% of queries contain between 2 and 3 search criteria, and 1 or 4 criteria are found in just over 10% of queries each, while less than 2% of queries contain more than 5 search criteria.

Chapter 6 Shortcomings of the study

The query analysis process revealed certain shortcomings in the study that was carried out, although we do not believe that they impact the integrity of the results extrapolated from the data set. This section outlines these shortcomings and proposes ways to account for them in future studies.

Lack of multimodal information
Multimodal information (both in terms of multimodal information stored in the database, and how to query it multimodally from the interface) was largely missing from the study data set, leading us to conclude that this type of information is difficult to elicit given the static and uni-modal format of the questionnaire. In order to gather indications about multimodality in an MMCR system interaction between participants and an actual system is required.

Insufficient data for adequate statistical analysis
It has been pointed out to the author that 300 queries is an insufficient amount for carrying out a sound statistical analysis on the results. Thus, the trends that have been discovered cannot be taken as representative until the data set is enlarged and a statistically sound analysis can be performed. Consequently, as has been pointed out earlier, we take the trends discovered in the current data set to be indicators rather than truths.

Bias of instruction set
There is concern that the explanation of the MMCR system that is contained in the questionnaire may have biased users as to the types of things that would be available to them. The only way that is seen to rectify this problem is to allow users to interact with an actual system and learn its capabilities through personal experience with it. The difficulty with this approach is that it is a long and difficult process that may never actually see the user exploit the full capabilities of the system. However, this in

itself would be telling as it could point out areas where the system would need to be fine-tuned to allow the user to learn about its capabilities in an interactive manner.

Breadth of query set
On average, each user gave about 8 queries. Taking into account the relatively low number of chain queries and the nature of the questionnaire, there are doubts as to whether the variety of the types of questions and topics covered by the queries in this study is fully representative of the core queries for real use. In order to verify this, we would need to look at extended interactions between a single user and the system, either over a single session, or perhaps more realistically, over several sessions.

Chapter 7 Conclusions and future work

As the results described in the preceding sections indicate the study has achieved four of the five goals outlined in section 4.2. It provides the IM2 research community with a large query set that is balanced, coherent and consistent, controls for bias and is easily analysable from several different perspectives, while at the same time allowing for the control of various variables. Moreover, it forms a solid foundation on which to begin the design and development of user requirements models, a base ontology and a language processing module for the multimodal meeting domain, as well as indications about the types of annotations that are necessary on the data and the type of data that needs to be available in the interface in order for the system to be useful to a user. The only goal of the study that was not met was the elicitation of multimodal information both in terms of what is stored in the database and the modalities that are used in a users' interaction with the system. The way in which we propose to address this issue in future studies is discussed in the following section.

Wizard of Oz Studies
The shortcomings of the study, described in chapter 6, and the nature of several other potential research issues mentioned throughout this report, point to the need for interactive studies to better understand how a real user would interact with the MMCR system. Consequently, we propose to design, develop and run a series of interactive studies using the Wizard of Oz technique. The Wizard of Oz technique is a method that is commonly used to test systems that are under development as it does not rely on all components of a system to be fully functional at the point of testing [3, 4, 8]. In this technique, a user interacts with a system whose lack of full functionality is compensated for by a human 'wizard' who simulates any missing functionality from a behind the scenes position in such a way that the user is never aware of the fact that any functionality is actually missing from the system. This method allows a full range of testing at a stage where it is early enough to make crucial modifications in the design if they are necessary.

In addition to further expanding the user requirements models being developed for the MMCR system, we hope that in particular Wizard of Oz studies will allow careful and detailed examination of the multimodal aspect of interaction with an MMCR system. Specifically, we need to determine whether:

a) users would request more multimedia information or information from different modalities if they were faced with a real system,
b) what modalities (and in what combinations) people prefer to use to obtain specific types in information from the database to do specific tasks,
c) how they prefer to receive that information.

These involve investigating input modalities, specifically the relationships between modalities – the nature and common context of the modalities and how to recognise the relationships, and the output modalities, where the general preferences for output modalities would have to be determined based on the tasks and users. In the case of some queries, the choice of output modality is explicitly stated in the query itself. If a person asks '*List all the topics that were not resolved'*, the only choice the system has is to show text (barring text-readers being incorporated into the system). In others queries such as '*Who attended the meeting?'* the choice of output modality is less clear. The system could show an image or a text with list of names, or both. Furthermore, the theory of matched modality from human-human dialogues, which postulates that people prefer to receive information in the same medium in which they give it [2], would need to be investigated to determine whether it holds for human-computer dialogues as well as this would have a direct influence on the output modalities selected by the system.

Additionally, given the static nature of the questionnaire, we cannot be sure that the vocabulary and sentence structure used in the queries are representative of the those that would be used in interaction with a real system, as the complexity of the language used may in fact change in an interactive situation [3, 4]. Gathering new query sets, or observing user interaction, in the course of Wizard of Oz studies will help to further tune the language processing components for the multimodal meeting domain by providing a clearer indication about the range of vocabulary and sentence structure in actual use situations.

A final consideration in the execution of Wizard of Oz studies will be which MMCR system to use. The question is whether to strictly follow software engineering and interface design guidelines and create a system based strictly on requirements extrapolated from the query set in this study, or whether to use MMCR systems that already exist within the IM2 project [1] (but which have been developed based on designer intuitions about what the requirements are), and augment or alter them to accommodate for any additional requirements that results from the query analysis.

Use of the queries for evaluation purposes

If the system as a whole, or any of its subcomponents, are to be tested in view of their use in an actual use environment, it is important to have a test set that is 'real' in the sense that it comes from real potential users trying to carry out real potential tasks. Queries such as those gathered in this study can not only be used to test the system as a whole - both the information retrieval and the interface components – but can also be used to test individual underlying components at intermediary stages in development. And while we concede that this test set is not ideally suited for the evaluation task because it may be incomplete and possibly underdetermined, it provides a baseline with which to work until a more valid set can be elicited.

Furthermore, the query set can be used both for refining the system as it is being developed, and for testing the breadth of the system once the design and development at a particular stage is complete. It is generally not considered good practice to use as an evaluation set for a system the same set on which the design of the system was based. However, given that the technology that is currently available is not yet at the stage of being able to immediately accommodate all of the required design features, we believe that in the meantime the existing query sets can be used as a measure of progress for the various iterations of the system. For example, if version 1 of the system can cover 10% of the queries, then version 2 should be able to cover more than 10%. Additionally, analysis of the query sets can also determine the most common or important features for the domain and/or the system and the lack of those features in a version could impact the overall evaluation of that particular version.

Thank you to Susan Armstrong and Andrei Popescu-Belis for their insights and comments on this work.

References

[1] Armstrong et al., “Natural Language Queries on Natural Language Data: a Database of Meeting Dialogues”, Proceedings NLDB’2003 (8th International Conference on Applications of Natural Language to Information Systems, Burg/Cottbus, Germany.

[2] Bilici, V., Krahmer, E., te Riele S., and Veldhuis R., ‘’Preferred Modalities in Dialogue Systems’’. http://citeseer.nj.nec.com/424290.html

[3] Coutaz, J and Salber, D., “Requirements for Multimodal Wizard of Oz Platforms.**”** http://citeseer.nj.nec.com/522152.html

[4] Dahlbäck, N, Jönsson, A and Ahrenberg L., “Wizard of Oz Studies – Why and How”, in W.D. Gray, W.E. Helfley and Murray, D. (eds). Proceedings of the 1993 Workshop on Intelligent User Interfaces (pp. 193/200) Orlando, FL. New York, ACM Press.

[5] Dix et al., “Human Computer Interaction” Second Edition, Prentice Hall, 1998.

[6] Johansson, P., “User Modeling in Dialogue Systems” from http://citeseer.nj.nec.com/johansson02user.html

[7] Kobsa, A., “User Modeling: Recent Work, Prospects and Hazards“, Adaptive User Interfaces: Principles and Practice. Eds. M Schneider-Hufschmidt, T. Kuhme and u. Malinowski. North-Holland, Amsterdam, 1993.

[8] Salber, D and Coutaz, J., “Applying the Wizard of Oz technique to the study of Multimodal Systems “, 3rd International Conference EWHCI’93, East/West Human Computer Interaction, Moscow. L. Bass, J. Gornostaev, C. Unger Eds. Springer Verlag Publ. Lecture notes in Computer Science, Vol. 73. pp. 219-230.

Appendix A - A copy of the questionnaire used for the elicitation study

Questionnaire: Eliciting input for a multimodal database query system

Introduction

In the context of the IM2 (Interactive Multimodal Information Management, www.im2.ch) project we are developing a system that stores electronic recordings of meetings between human beings, in a variety of formats - video, audio, text transcripts etc. In addition to the content of the meeting itself, any files that are associated with the meeting, such as presentation slides and distributed papers are also stored in electronic form. The fact that the information is stored in a variety of annotated formats means that the user can ask questions about the actual content of the meetings, in addition to requesting to see or hear parts of the meeting.

A user is able to access the information in the system by simply posing a question to the system (much as they might if they were trying to get the same information from a colleague) about the information in those meetings, or requesting to see some or all of a particular meeting. The user can ask the question through any combination of typing on a keyboard, speaking and using a pointing device (such as a mouse or laser pen). What we are currently interested in is what aspects of a meeting people would want to know about and how they would pose their questions.

Scenarios

Below you will find four scenarios in which someone might want to use the system that has been described. Please pick one, and then list as many questions as you can think of that you would want to ask the system to get the answers you need, given the context described in the scenario. Note: If you are willing to do this for more than one scenario, that would be very helpful, but we ask that the questions be listed separately for each scenario.

1. Imagine that you are managing a project, but are too busy to attend all of the meetings related to it. This isn't a problem, because you know that all meetings in your institute have been recorded and stored in the IM2 system. You want to find out how particular members of the group are contributing overall to the project.

2. Imagine that you are managing a project, but are too busy to attend all of the meetings related to it. This isn't a problem, because you know that all meetings in your institute have been recorded and stored in the IM2 system. You want to find out how the flow of ideas for the project has been progressing, what directions the project is taking and what decisionsare being made.

3. Imagine that you have missed a meeting about a project that you are working on. You want to catch up on what happened in that meeting (what was discussed, what was decided etc), and you know that all meetings in your company are recorded and stored in the IM2 system. Also remember that since all of the meetings are stored, you can also ask questions about previous meetings that you attended, if you feel that that can be helpful.

4. Imagine that you have just been hired at a company to work on a project. The project actually started six months ago, so you have some catching up to do. Fortunately, all of the meetings regarding this project have been recorded and are stored in the IM2 system which you will use to help you catch up.

About you

It would be helpful if you could answer the following questions, but you are not obligated to do so.

1. Are you involved in any way with the IM2 (Interactive Multimodal Information Management) project? If yes, please specify how.

2. What sort of computer experience do you have? Please erase all those options that do not apply to you, leaving only those that best describe your experience.

Basic personal computer use (word processing, internet browsing etc.)
Extensive personal computing use
Computer programming experience
A computer science (or related) degree
Experience with building databases
Experience with natural language processing

3. What is your professional position? Again, please erase those that do not apply to you.

Researcher
Manager/director
Other (please specify)

4. What is your area of expertise?

5. Please specify your native language:

6. Would you be willing to volunteer to participate in a hands-on experiment in this area in the future?

Thank you for taking the time to help us with our research. We greatly appreciate it.

Appendix B: Unedited Lists of Queries from the elicitation study

This document contains the unedited lists of queries gathered during the query elicitation study run by Agnes Lisowska as part of the IM2.MDM IP and contains a total of 297 queries. The queries are primarily divided by use case – the brief description of the use case that was given in the questionnaire is also provided. Each use case section is further divided into queries from participants involved in the IM2 project and those who were not.

1. Use Case 1

Use case: Imagine that you are managing a project, but are too busy to attend all of the meetings related to it. This isn't a problem, because you know that all meetings in your institute have been recorded and stored in the IM2 system. You want to find out how particular members of the group are contributing overall to the project.

IM2 participants

- 1-Y-2 -1 I would like a list of all participants to at least a meeting, with the proportion of meetings attended by each participant. Could you sort the list starting with those that attended most of the meetings?
- 1-Y-2 -2 Which people missed meetings which they were supposed to attend? (use the call for meeting)
- 1-Y-2 -3 Were there any talks longer than 10 minutes at some meetings? Who gave these talks?
- 1-Y-2 -4 Who were the moderators of each meeting?
- 1-Y-2 -5 Was someone accused of not doing his/her job? By whom? What was the reply? Was there any consensus in favour or against the accusation?
- 1-Y-2 -6 Who reported major results? For instance, who reported publications related to the project?
- 1-Y-2 -7 Who contributed to the project reports? Who promised to contribute but failed to do so?
- 1-Y-3-1 Did X talked at all during the meeting?
- 1-Y-3-2 If so, did he make any suggestions on the design?
- 1-Y-3-3 Did Y accept/reject his suggestions?
- 1-Y-3-4 Did Y often interrupt the other team members?
- 1-Y-3-5 Did Y make his own ideas prevail?
- 1-Y-3-6 Were any of X's suggestions included in the final specifications
- 1-Y-4-1 Who was actively participating to the meeting?
- 1-Y-4-2 Which role was played by X during the discussion about issue Y?
- 1-Y-4-3 Who headed the meeting?
- 1-Y-4-4 Show me the arguments between X and Y.
- 1-Y-4-5 Why the proposals made by X were systematically rejected?
- 1-Y-4-6 Who talked most during the meeting?
- 1-Y-4-7 Why the proposal made by X was not accepted?
- 1-Y-4-8 Show me the passage where X made a proposal about issue Y.
- 1-Y-4-9 Show me the conflicts of opinion between X and Y.
- 1-Y-4-10 Who made the proposal on issue X which was finally accepted by all the participants?
- 1-Y-4-11 How was the conflict of opinion between X and Y solved?
- 1-Y-4-12 Which was the level of attention/involvement of participant X during the meeting/discussion of issue X?
- 1-Y-4-13 Who has the most positive/negative attitude with respect to other participants?
- 1-Y-4-14 Who made the highest number of negative/positive comments?
- 1-Y-4-15 Who made constructive criticisms about the proposal X on issue Y?
- 1-Y-4-16 When did X contradict himself about the issue X?
- 1-Y-4-17 Who left the meeting before the end?
- 1-Y-4-18 Why X left the meeting before the end?
- 1-Y-4-19 Were there any coalitions in accepting/rejecting the proposal X on issue Y?

- 1-Y-4-20 Give me all the contributions of participant X in favour of alternative A regarding the issue I
- 1-Y-4-21 What participant X actually intended by saying P during the discussion of issue I?

Non-IM2 participants

- 1-N-1-1 What has each person promised to do? And when did they promise it?
- 1-N-1-2 Have they delivered? On time?
- 1-N-1-3 How long does it take each person to do each task?
- 1-N-1-4 Does an employee have initiative - do they do what they're told and when, or do they come up with their own work/schedules?
- 1-N-1-5 How valuable is a task to the project?
- 1-N-1-6 Is an employee taking on leadership responsibilities?
- 1.1.1-N What was actually discussed at the meeting? (which may not be quite the same as the advertised topic for discussion)
- 1.1.2-N Who presented what?
- 1.1.3-N What did X present?
- 1.1.4-N What did X precisely say about topic Y?
- 1.1.5-N Did X come up with new information/insights/plans for topic Y?
- 1.1.6-N How was X's contribution received by the others at the meeting?
- 1.1.7-N Given a particular comment or criticism, how did X react to this?
- 1.1.8-N What was decided?
- 1.1.8a-N Was it decided to do A or B?
- 1.1.8b-N Did X volunteer to do A or B?
- 1.1.8c-N What did X promise to do next?

2. Use Case 2

Use Case: Imagine that you are managing a project, but are too busy to attend all of the meetings related to it. This isn't a problem, because you know that all meetings in your institute have been recorded and stored in the IM2 system. You want to find out how the flow of ideas for the project has been progressing, what directions the project is taking and what decisions are being made.

IM2 participants

- 2-Y-1-1 What was agenda item 3 about?
- 2-Y-1-2 Was there again a discussion about how to organize the multimodal data in a database?
- 2-Y-1-3 What requirements for the data model have been mentioned?
- 2-Y-1-4 What where the arguments against using database product X?
- 2-Y-1-5 Was any new argument for product X brought up, i.e., an argument that was not mentioned in a former meeting?
- 2-Y-1-6 Who questioned this argument?
- 2-Y-1-7 Was there a consensus about using a commercial product at all?
- 2-Y-2-1 What were the topics of the last meetings?
- 2-Y-2-2 What were the most discussed issues?
- 2-Y-2-3 Were there any formal decisions made? Who proposed them? Who was against?
- 2-Y-2-4 Were there any talks longer than 10 minutes at some meetings? Who gave these talks?
- 2-Y-2-5 Were there any major clashes? Between which participants?
- 2-Y-2-6 Were there any documents produced recently in relation to the project? Give me their authors, titles, dates, and types (publication-where?, deliverable, report, etc.)
- 2-Y-2-7 Has any of the people initially involved in the project stopped working on the project or stopped coming to meetings? Are there any newcomers to the project?

- 2-Y-3-1 What issues from the agenda where discussed during the meeting?
- 2-Y-3-2 What decisions have been taken?
- 2-Y-3-3 Who rejected the proposal made by X on issue Y?
- 2-Y-3-4 How was the proposal made by X discussed during the meeting?
- 2-Y-3-5 When the discussion on issue X degenerated?
- 2-Y-3-6 Why the issue X from the agenda was not discussed?
- 2-Y-3-7 Which proposal were accepted without any discussion?
- 2-Y-3-8 Show me a diagram highlighting the pro and con with respect to the proposal made by X on issue Y.

Non-IM2 participants

- 2.1.1-N Who was at the meeting?
- 2.1.1a-N Who chaired it?
- 2.1.2-N What topics were discussed?
- 2.1.3-N How long was spent on each topic?
- 2.1.4-N Did any topic give rise to unusual tension?
- 2.1.5-N How much did each participant talk?
- 2.1.6-N Which participant talked on which topic?
- 2.1.7-N Were there long interactions between any two or maybe three participants?
- 2.1.8-N How long did the whole meeting last?
- 2.1.9-N What decisions got made?
- 2.1.10-N What decisions got made easily?
- 2.1.11-N What decisions got made only after prolonged or intense discussion?
- 2.1.12-N Who was involved in the discussion in these cases?
- 2.1.13-N Did any topic lead to a request for further information to be developed after the meeting? (e.g., can someone get these figures? can someone do as a list of pros and cons ...)
- 2.1.14-N Did any participant leave the meeting early?

3. Use Case 3

Use case: Imagine that you have missed a meeting about a project that you are working on. You want to catch up on what happened in that meeting (what was discussed, what was decided etc), and you know that all meetings in your company are recorded and stored in the IM2 system. Also remember that since all of the meetings are stored, you can also ask questions about previous meetings that you attended, if you feel that that can be helpful.

IM2 participants

- 3-Y-2-2 What were the major items on the agenda for this meeting?
- 3-Y-2-3 What was the follow up to question X raised in last week's meeting?
- 3-Y-2-4 Has there been any decision taken on that point?
- 3-Y-2-5 Did person X present the results of his work concerning matter Y?
- 3-Y-2-6 For when is the next meeting planed?
- 3-Y-2-7 What were the major decisions taken during the meeting?
- 3-Y-2-8 What problems, still remain pending?
- 3-Y-2-9 Who is taking care of problem X?
- 3-Y-2-10 Where there any handouts distributed on the meeting I missed? From whom can get them?
- 3-Y-2-11 X raised question Y on last meeting about problem Z, did you come up with a solution?
- 3-Y-4-1 What important decision have been taken during the last meeting?
- 3-Y-4-2 Have any decision been taken regarding subject X?
- 3-Y-4-3 Did anyone mention that I was not there?
- 3-Y-4-4 Were there tasks allocated to me?
- 3-Y-4-5 Can I have a short summary of the last meeting.
- 3-Y-4-6 Are there any dates mentioned that I should put in my agenda (deadlines,

appointments)?

- 3-Y-4-7 What were the points on the agenda?
- 3-Y-4-8 Who were there?
- 3-Y-4-9 Who was missing?
- 3-Y-7-1 What issues left undiscussed in the previous meeting have been discussed in this meeting?
- 3-Y-7-2 Why X changed his mind on issue Y in the current meeting?
- 3-Y-7-3 What issue have been rediscussed in the current meeting?
- 3-Y-7-4 Was X present in all the meetings?
- 3-Y-7-5 Was X actively participating to the discussions of all the meetings?
- 3-Y-7-6 What is the average number of decision taken across the meetings?
- 3-Y-7-7 Why the issue left undiscussed in the previous meeting was not included in the new agenda?
- 3-Y-7-8 For which open question there was no solution adopted? Why? What are the open questions for a next meeting?
- 3-Y-7-9 Which criteria were chosen to take the decision D1?
- 3.1.1-Y What was the agenda of the meeting?
- 3.1.2-Y Who were the participants?
- 3.1.3-Y From which institution are coming the X and Y participants?
- 3.1.4-Y Who were the participants representing the I1 institution?
- 3.1.5-Y What were the decisions to be taken (open questions) regarding the topic t1?
- 3.1.6-Y How long have they been discussing about topic t1?
- 3.1.7-Y Give me, please, that part of the meeting when people were discussing about topic t1.
- 3.1.8-Y For which open question there was no solution adopted? Why?
- 3.1.9-Y What are the open questions for a next meeting?
- 3.1.10-Y What decisions were adopted?
- 3.1.11-Y Who suggested the adopted solution s1?
- 3.1.12-Y Which participant suggested more of the adopted solutions?
- 3.1.13-Y Which members agreed / disagreed on taken the decision d1?
- 3.1.14-Y Which criteria were chosen to take the decision d1?
- 3.1.15-Y Which criteria invoked the members who disagreed the taken decision d1?
- 3.1.16-Y What were the documents presented during the meeting?
- 3.1.17-Y Who presented the document d1?
- 3.1.18-Y I would like to see the first slide from the presentation given by person X.
- 3.1.19-Y What was X's information request concerning the document d1?
- 3.1.20-Y What was the reaction / response to the request r1?
- 3.1.21-Y Who was sleeping during the meeting? at what moment?
- 3.1.22-Y What was the topic in discussion when the majority of people were "sleeping" / laughing / acclaiming/applauding?

Non-IM2 participants

- 3-N-3-1 Who attended the meeting
- 3-N-3-2 What were the main topics discussed (e.g. A, B, C)
- 3-N-3-3 Any decisions made regarding the topic C
- 3-N-3-4 Was there an action plan determined for topic C and what was it
- 3-N-3-5 Can I hear discussion re: Topic B
- 3-N-3-6 If there was a key person that I trust (Mr X), what did he say re: such and such
- 3-N-3-7 Was there any data presented re:Topic A? Can I see it?
- 3-N-3-8 When is the next meeting.
- 3-N-3-9 I seem to recall Topic A discussed before? Was it?
- 3-N-3-10 What was discussed and what were the decisions made then?
- 3-N-3-11 Any financial obligations mentioned re Topic A? What were they
- 3-N-3-12 I'd like to send a memo to people involved in Topic A
- 3-N-3-13 Did they serve Tim Horton's, Starbuck's, and did they have the jelly filled donuts?

- 3-N-3-14 Can I see the minutes of the meeting regarding Topic B
- 3-N-1-1 Who was invited to the meeting?
- 3-N-1-2 Who attended the meeting?
- 3-N-1-3 What was the agenda?
- 3-N-1-4 What was discussed that was followed-up from previous meetings?
- 3-N-1-5 What directives/requests were given to committees / project members?
- 3-N-1-6 What is the timeline for those requests, and how do they relate to other timelines/deadlines in this project?
- 3-N-1-7 What items were set to be discussed in the next meeting?
- 3-N-1-8 What comments were made, what were the transcribed comments, and what was the duration of each comment?
- 3-N-1-9 What was said when each slide was presented? (annotation of slides with transcription of presentation, and questions)
- 3-N-5-1 How long did the meeting last?
- 3-N-5-2 Who were the participants?
- 3-N-5-3 What topics were discussed?
- 3-N-5-4 What decision was achieved on topic #3?
- 3-N-5-5 List all positive and negative sides of the decision #3.
- 3-N-5-6 What actions are required to implement the decision on topic #3?
- 3-N-5-7 Were all topics successfully resolved?
- 3-N-5-8 List topics that were not resolved.
- 3-N-5-9 Why was topic #5 not resolved?
- 3-N-5-10 Play recording of discussion on topic #5.
- 3-N-5-11 Did the manager inquire on what kept me from attending the meeting?
- 3-N-6-1 What was "John" response to my comment on last meeting.
- 3-N-6-2 Who has a new suggestion how to solve this problem
- 3-N-6-3 Who discussed new ideas
- 3.2.1 -N How long did the meeting last?
- 3.2.2-N What was the main topic discussed?
- 3.2.3-N Did somebody bring about a subject that was not in the "ordre du jour"?
- 3.2.4-N What did X say? And what did Y respond to X about that?
- 3.2.5-N Was Y angry at X?
- 3.2.6-N Did somebody talk about me or about my work?
- 3.2.7-N Has something been said about the one-before-last meeting?
- 3.2.8-N Did somebody answer the question I asked last week?
- 3.2.9-N Did somebody leave the room during the meeting?
- 3.2.10-N What will be dealt with in the next meeting?
- 3.2.11-N Have I been assigned a new task?
- 3.2.12-N Has somebody ever said when the project was due to end?
- 3.3.1-N What was discussed?
- 3.3.2-N Did they discuss X? (where X is some particular topic that is of interest to me)
- 3.3.3-N What was decided? (again depending on the content of the meeting and my interests, the questions can be more specific:
- 3.3.4-N Did they decide to do X?
- 3.3.5-N When are we going to do X?
- 3.3.6-N How are we going to do X?
- 3.3.7-N Why did they decide to do (or not to do) X?
- 3.3.8-N Who is going to do X?
- 3.3.9-N Did they mention me?
- 3.3.10-N Was my proposal put forward?
- 3.3.11-N How did people react to my proposal?
- 3.3.12-N What did P say about me?
- 3.3.13-N What did P say about X?
- 3.3.14-N Who suggested X?

4. Use Case 4

Use case: Imagine that you have just been hired at a company to work on a project. The project actually started six months ago, so you have some catching up to do. Fortunately, all of the meetings regarding this project have been recorded and are stored in the IM2 system which you will use to help you catch up.

IM2 participants

- 4-Y-1-1 Did someone briefly present the project during some meeting at the start of the project?
- 4-Y-1-2 Who is involved in the project?
- 4-Y-1-3 Is there an email list of every body involved in the project?
- 4-Y-1-4 How often has there been meetings on the project?
- 4-Y-1-5 Who usually attends the project meetings?
- 4-Y-1-6 Who is responsible for the project?
- 4-Y-1-7 Are there any external partners to the project
- 4-Y-1-8 What were the major decision taken so far on the project?
- 4-Y-1-9 What budget do we have for the project?
- 4-Y-1-10 Are there any existing reports on issue x?
- 4-Y-1-11 What are the following deadlines?
- 4-Y-3-1 Who are the participants of the project who attended all (or most of) the meetings?
- 4-Y-3-2 Who were the external people who participated to the project meetings?
- 4-Y-3-3 Find the first meeting where the issue X was addressed.
- 4-Y-3-4 Show me the final decision taken about issue X.
- 4-Y-3-5 Show me all the discussions about issue X.
- 4-Y-3-6 Enumerate all the issue discussed during the first 6 months of the project.
- 4-Y-3-7 Show me the issues (discussed during the first 6 months of the project) for which no final decision have been taken.
- 4-Y-3-8 Show me all the issues discussed so far on topic X.
- 4-Y-3-9 Show me the presentations (slide shows, monologues) occurred during the first 6 months project meetings.
- 4.2.1-Y What important decision have been taken during the last 6 months?
- 4.2.2-Y Who is the most dominant person in meetings?
- 4.2.3-Y Who disagrees a lot?
- 4.2.4-Y Where do X and Y talk about Z?
- 4.2.5-Y Where do X and Y disagree?
- 4.2.6-Y Give me a summary of all past meetings about subject Z.
- 4.2.7-Y List the subjects of all past meetings.
- 4.2.8-Y List the parts of meetings where people are talking about Z.
- 4.2.9-Y What is the name of the man with the white shirt?
- 4.3.1-Y For every meeting, name and position of the participants
- 4.3.2-Y What were the topics discussed (detailed)
- 4.3.3-Y Summary of the meeting
- 4.3.4-Y Main decisions taken
- 4.3.5-Y Who asked what to whom
- 4.3.6-Y What were the conclusions
- 4.3.7-Y Other participants involved in the topic
- 4.3.8-Y Suggestions
- 4.3.9-Y Unsolved problems
- 4.3.10-Y Further work attributed to the various participants
- 4.3.11-Y Topics to be discussed in future meetings
- 4.3.12-Y What documents were consulted during the meeting (by topic: name of person who provided document, type of document: report, biblio, etc.). Document should also be available in the database.

Non-IM2 participants

- 4-N-2-1 Who is originator of project idea
- 4-N-2-2 What is a goal of this project
- 4-N-2-3 Who will implement this project in customer location
- 4-N-2-4 What a responsibilities of each group member
- 4-N-2-5 Who was working on my part of project before me
- 4-N-2-6 What a result of his job. If any show slides or play video.
- 4-N-2-7 Who previously commented or discussed my part of project
- 4-N-2-8 Why this project was assigned to me.
- 4-N-4-1 What is the deadline for this project?
- 4-N-4-2 Have any prototypes been built and will there be a pilot run?
- 4-N-4-3 Are there any software issues that I should know about or modifications that are needed?
- 4-N-4-4 What is the expected launch date for this project?
- 4.1.1-Y Who suggested to take a cup of coffee ?
- 4.1.2-Y Who initiated the IM2 project first?
- 4.1.3-Y When did that meeting happened and where?
- 4.1.4-Y Did someone disagreed?
- 4.1.5-Y What were some of the objections (if exist)?
- 4.1.6-Y What arguments did X use in the matter Y?
- 4.1.7-Y Who attend to this meeting?
- 4.1.8-Y What was the reaction of X when Y said Z? any kind of reaction (verbal, non-verbal) this question is too much deep I think or is it among your plans of realisations ?
- 4.1.9-Y Can you show me the poster presented by A.L. the 02/15/2004 ?
- 4.1.10-Y Who are the other participants and what are there qualifications?
- 4.1.11-Y What are the mains ideas that prevail through all these meetings?
- 4.1.12-Y What do each one of the people have to do?
- 4.1.13-Y Are there some ?? about some topics ?
- 4.1.14-Y Is there something that was not mention at the beginning and that appear later?
- 4.1.15-Y Does a controversy exist? if yes, where?
- 4.1.16-Y How do some deal with the matter of 'conversation', eg (if you work on conversation)?
- 4.1.17-Y What are their opinions about X?