An Exploratory Study of Topic Importance in Requirements Elicitation Interviews

Corentin Burnay[1,2,3], Ivan J. Jureta[1,2,3], and Stéphane Faulkner[2,3]

[1] Fonds de la Recherche Scientifique – FNRS, Brussels
[2] Department of Business Administration, University of Namur
[3] PReCISE Research Center, University of Namur
{corentin.burnay,ivan.jureta,stephane.faulkner}@unamur.be

Abstract. Interviewing stakeholders is a common way to elicit information about requirements of the system-to-be and the conditions in its operating environment. One difficulty in preparing and doing interviews is how to avoid missing the information that may be important to understand the requirements and environment conditions. Some information may remain implicit throughout the interview, if the interviewed stakeholder does not consider it important, and the business analyst fails to mention it, or a topic it relates to. We propose the so-called Elicitation Topic Map (ETM), which is intended to help business analysts prepare elicitation interviews. ETM is a diagram that shows topics that can be discussed during requirements elicitation interviews, and shows how likely it is that stakeholders tend to discuss each of the topics spontaneously (as opposed to being explicitly asked questions on that topic by the business analyst). ETM was produced through a combination of theoretical and empirical research.

Keywords: Elicitation, Interviews, Context, RE Topics, Exploratory study.

1 Introduction

Research Context - Requirements Elicitation via Interviews: Requirements Engineering (RE) focuses on the elicitation, modelling, and analysis of requirements and environment of a system-to-be, in order to produce its specification. Requirements elicitation [1–4], only *elicitation* hereafter, refers to activities done in RE, in order to obtain information from stakeholders of the system-to-be; the aim is to use this information to understand conditions in the system's operating environment, and the stakeholders' requirements from the system [5].

Elicitation is important, because misunderstanding stakeholders, or in some other way missing important information, can result in the specification of the wrong system - one that fails to satisfy requirements, and/or is inconsistent with the conditions in its operating environment (e.g., it does not comply with applicable legislation). *Elicitation often involves communication with stakeholders,* through, for example, structured, semi-structured, or unstructured interviews, workshops, and so on [2, 3]. Hereafter, we write *interviews* to refer to any form of direct communication with stakeholders, which

M. Jarke et al. (Eds.): CAiSE 2014, LNCS 8484, pp. 180–195, 2014.

is done in order to elicit information. Interviews provide invaluable information through verbal and nonverbal communication.

General Issue - How to Uncover Important Context Defaults during Interviews? A difficulty when doing interviews, is that the business analyst and stakeholders have different backgrounds, experiences of existing systems, and expectations from the future system. They will come into interviews with different assumptions about the environment, requirements, and system-to-be. In itself, it is not a problem that different stakeholders hold different assumptions. It becomes a problem if some of their key assumptions remain implicit in interviews. If, instead of remaining hidden, some of these assumptions were known, then this could have helped with, for example, requirements inconsistencies, stakeholder negotiations, or the identification of other requirements.

A more technical way to see this, is to look at it through the notion of non-monotonic reasoning in artificial intelligence [6–9]: when the business analyst is doing elicitation interviews, she is asking the stakeholder questions; the stakeholder's thinking before answering could be - roughly speaking - seen as an inference that the stakeholder makes on the basis of her defaults (statements that can be rejected when there is new information) and her certain knowledge (statements which remain relevant despite of any new information) [8]; the stakeholder's answer are the conclusion of her reasoning process. If we see things this way, then it can be useful for the requirements engineering to try to reveal at least some of the stakeholder's defaults, in order to understand the requirements better, discuss other requirements, or otherwise.

This is, for RE research, the issue of how to make sure that elicitation interviews reveal as much as feasible of the defaults that may be important for RE. It is not a new research issue. Any contribution on how to prepare elicitation interviews, is also inevitably interested in how to use these interviews to elicit as much as feasible of the important information for RE [4, 10–13]. An approach to this issue that has not received attention, consists of trying to understand what domain-independent categories of information the stakeholders tend to talk spontaneously about during interviews, and which others tend to remain implicit. The latter group are the defaults mentioned above.

Contributions - Map of Elicitation Interview Topics, and their Relative Importance: The contributions of this paper are the so-called *Elicitation Topic Map (ETM)*, a list of topics to discuss in elicitation interviews, and indications of the relative importance of these topics. Topic importance reflects our measure of the stakeholders' tendency to share spontaneously the information on these topics: a topic is more important if we observed, in our sample of stakeholders, that they were more willing to share information about it spontaneously. This does not mean that less important topics in the ETM are less important for the analysts: it simply means that fewer stakeholders would spontaneously share information on them; if the analyst needs information on lower importance topics, she will have to be proactive in finding that information (for example, the analyst would need to stimulate stakeholders to discuss those topics).

Overview of Research Methodology: The ETM was produced through three phases of research. It is easier to understand the rationale for them, by starting from the third and last phase. The goal of the third phase was to evaluate topic importance. This was done by selecting professionals, all of whom had acted as stakeholders in RE projects, and

sending each of them the same set of 30 topics. We asked each individual to evaluate, for each topic, if she would share information on it spontaneously, or only if asked.

In order to have the 30 topics to evaluate, the second phase of research focused on identifying the topics. They were identified through interviews with business analysts, coming from five RE and systems engineering projects done in Belgian small and medium size businesses; projects differed in terms of number of participants (from 15 to 150) and in terms of the system domain (pharmacology, finance, etc.). To prepare interviews done in the second phase, our first phase consisted of adapting our past research on a generic model of context [14, 15]. Our model of context suggested groups of topics, without suggesting specific ones.

Organization: This paper is organized as follows. Section 2 introduces basic terminology and relates it to standard RE terminology. Sections 3–5 present the three phases of research. Section 6 presents the raw data and the analysis technique we applied to it to produce ETM. Section 7 presents hypotheses for future research in RE, that ETM suggests. Section 8 overviews related work, and Section 9 summarizes our conclusions.

2 Baseline

We start from the observation that there is explicit and implicit information when doing an elicitation interview. *Explicit information* is that which the stakeholder shared with the business analyst who did the interview. *Implicit information* is that which the stakeholder did not share by the end of the interview. The fact that some information is explicit or implicit does not matter for its relevance for understanding the requirements and the environment of the system-to-be. The stakeholder decides what information to share, and thereby which information will be explicit or implicit.

Stakeholder's decisions to share undoubtedly depends on many factors, such as the business analyst's questions, the stakeholder's assumptions about the system-to-be and its environment, her understanding of her role in the systems engineering process, and so on. The goal of the ETM is to influence primarily the set of questions that the business analyst asks, rather than the other factors. We see the elicitation interview as an exchange of information and questions between stakeholders and business analysts.

Although the number of stakeholders and analysts in an interview will have an influence on the content and procedure of the interview in practice, they do not influence the contributions in this paper - the ETM is not designed with a specific interview duration and number of participants in mind. This exchange can be more or less controlled; more, for example, if the analyst wishes to proceed in the exact same way with every stakeholder and in every interview, perhaps through the same list of questions. We see any interview as a conversation about a set of topics, regardless of how controlled that conversation is, or the analyst may want it to be.

In this paper, the term **Topic** designates an entity that different pieces of information refer to. A topic can be, for example, a time period (talking about the events in March 2013), a physical object (the company's product packaging), event (merger with another company), position (CEO), etc. Another key term in this paper is **Topic Set**, which refers to a set of Topics that are somehow related. For example, if there is a Topic for past events, another for current events, and a third for future events, then there can be

a Topic Set about time, which includes all the three Topics. It is important to keep in mind that Topic is not a subclass of Topic Set, and that same Topic can be in more than one Topic Set. We have also found no universal set of Topic Sets, or of Topics per Topic Set; we are reporting in this paper those Topics and Topic Sets that proved useful with regards to the issue we are interested in, namely, providing an ETM and an evaluation of Topic importance in it.

It is important to understand how the notion of Topic in this paper relates to common concepts in requirements modelling languages, such as RML [16], ERAE [17, 18], Telos [19], KAOS [20] or i* [21]. A requirements modelling language suggests concepts and relations to use, to represent information about requirements, environment, and the system-to-be. If an elicitation interview results in explicit information about key actors in the environment, and how they depend on the system-to-be to achieve some specific goals, then, for example, an i* model can be used to capture these as instances of its agent, role, and goal concepts, and its dependency relation. In a way, the concepts and relations of the language can be seen as suggesting Topics to discuss. If the language is i*, then Topics would be the agents and roles in the environment, the goals of the agents, and the dependencies between them for achieving these goals. The difference between Topics, and concepts and relations in requirements modelling languages, is that a Topic may correspond one to one to a concept or relation, or to more concepts and relations among those in the language. Our aim in defining the Topics was not to suggest an ontology for requirements modelling languages. Some languages may be able to capture the information associated to some Topics more easily than others, but that discussion is beyond the scope of this paper, and influences in no way the contributions here.

3 The First Phase: Defining Topic Sets

The purpose of the first phase of our research was to define Topic Sets. To do so, we started from the idea that all elicitation interviews can be said to be context-specific. This means that an interview is specific to a time, place, project, analyst doing the interview, stakeholder being interviewed, and so on. In other words, to say that elicitation is context-specific, is simply to say that no two elicitation interviews are alike.

The useful conclusion to draw here, from the observation that elicitation is context-specific, is that context influences the answers that stakeholders give. Therefore, if we keep the same analyst who interviews, the same stakeholder who is being interviewed, and the same questions, and change something else in the context (such as interview location, time, and so on), then we may get different explicit information from the interview. Notice that we are careful to say that we actually do not know if a change to context would in fact change the information that the stakeholder chooses to share.

Consequently, phase one involved two tasks: (i) identify Context Dimensions, that is, groups of variables which characterize the context, so that if they change, then we say that context changed from an old context to a new context; (ii) determine, through experiment, which of the Context Dimensions influence people's decision-making: namely, given some Context Dimensions, we want to identify those that have the following property: The same individual, when facing the problem in the old context would solve

it in one way, and when in the new context, would solve it in another way. We reported elsewhere our work on the two tasks above [14, 15].

To identify Context Dimensions, we drew on conceptualizations of context in philosophy [22, 23], artificial intelligence [24] and computer science. In computer science, for instance, fields like ubiquitous computing and context-awareness are particularly interested in the notion of context, and, so to speak, what context is made of (see [25, 26] for surveys). This interest has lead to some operational definitions of context (e.g., [27]). These definitions decompose context into a series of dimensions. We identified six of these Context Dimensions in our past work, and use them in this paper as Topic Sets, with each Context Dimension being a Topic Set. They are the following:

- **Items** deal with salient entities inside the context, e.g., a person, an object, etc.
- **Rules** deal with constraints in the context, e.g., laws, targets, habits, etc.
- **Localizations** deal with the position of the context in space and time;
- **Activities** deal with the set of objectives of Items, e.g., intentions, desire, etc.
- **Relationships** deal with the connections / links between Items and/or Rules.
- **Granularities** deal with the nature, quantity and level of any additional piece of information that is provided about things occurring in the context.

4 The Second Phase: Identifying Topics for Topic Sets

The output of phase one are a list of six Topic Sets. While interesting on their own, the Topic Sets are not very useful for elicitation, as they are too general. Asking questions about items, rules, localization, and so on, still are much too generic recommendations on what to discuss during interviews.

To identify Topics for Topic Sets, we selected business analysts, and did interviews with them. The aim in the interviews, was to discuss the Topic Sets, their perception of the relevance of Topic Sets, and to identify Topics that they would have, or actually had discussed with stakeholders. The resulting Topics are given in Table 2. The rest of this section describes how we found these Topics.

Participants: We had access to five systems engineering or reengineering projects, which involved professional business analysts. Projects took place at small and medium sized companies (up to 250 employees) located in Belgium and Luxembourg. When we did our study, all projects had ended in the 12 months that preceded our study. We interviewed the business analysts involved in these projects. The interviews took place at the respective companies that employed these individuals. In addition, we had access to requirements documentation produced for the projects. We chose projects so as to cover different domains and project sizes. Names of systems engineering buyers and providers remain anonymous in this paper. This was a condition to satisfy in order to gain access to project documentation and the people involved. Table 1 gives an overview of project characteristics.

Procedure: The research in this stage was interpretative. As suggested in [28], it was mostly based on interviews and project documentation. The interviews were semi-structured, in that the goal in each interview was to discuss all Topic Sets identified

Table 1. Summary of project characteristics

Name	# Stakeholder	Industry	Description
PP	60	Pharmacology	Reporting system for customers' feedback
FD	15	Finance	Implementation of a CRM system
ML	20	Accounting	Automation of accounting dashboards
AP	90	Communication	Design of a BI system
BD	100	ICT	Scheduling tool for human resources

in phase one. At any time during an interview, subjects were free to mention any aspect outside the scope of the Topic Sets, or challenge the Topic Sets. The process was iterative: we analyzed documents generated during the project, and asked questions to analysts, when some aspects emphasized during the interviews did not correspond to observations in the documentation. Such iterations happened up to three times (three interviews and documentation analyses, for each analyst). An interview typically involved three parts, each dealing with particular types of questions:

- Overall discussion with direct references to Topic Sets, e.g., "Do you think it is relevant to discuss Rules during an interview with a stakeholder?";
- Specific discussion about what Topics might be in each Topic Set, e.g., "What aspects related to Rules would you want to elicit?" or "Do you consider culture of the company is a relevant aspect to be discussed with stakeholders when you want to collect information about Rules";
- Concluding discussion with broader questions such as, e.g., "Do you see other aspects that we did not mention during this interview?".

Results: The result of phase two is a list of 30 Topics, organized by Topic Set. They are shown in Table 2. The limit of 30 Topics was decided taking into account the largest set of Topics on which we could work and for which methodological concerns (in terms of validity, data collection and treatment) remained manageable. Hereafter, we refer to these Topics by mentioning the identifier they have in Table 2. For example, if we write I2, we are referring to the Topic of Objects that could be related to the system, as shown in Table 2. Some of the Topics identified during interviews have been removed from the final list, e.g., "Important financial ratios" has been identified as a Topic in the ML project (see table 1), but was rejected because it dealt with aspects that are only relevant in the scope of an accounting reporting system. Similarly, "Assignments from management team about ergonomics" has been rejected because too precise, and partially redundant with R4.

5 Third Phase: Evaluating Topic importance

The goal of phase three was to evaluate if system stakeholders would share spontaneously or not the information about the Topics identified in phase two.

Our premise is that if our data suggests that stakeholders tend to spontaneously share information about a Topic, then that topic is likely to produce explicit information in elicitation interviews. If data suggests that stakeholders do not tend to spontaneously share information about a Topic, then this information will remain implicit in elicitation interviews, unless the business analyst asks the stakeholders about it. The rest of this

Table 2. A list of Topics, by Topic Sets

Item	I1. Actors who are going to use the system I2. Objects that could be wired to the IS I3. Other systems that are in use in the firm I4. Inputs and outputs expected of the system I5. Units/structure that compose the firm	**Localization**	L1. Place where the system will be used L2. Repetitive trends in the firm L3. Frequency of recurring events in the firm L4. Recurring events in the firm L5. History and evolution of the firm
Activity	A1. Core business of the firm A2. Reason why the company needs the IS A3. Purpose of the IS, what it is going to do A4. Goals assigned to you and the colleagues A5. Vision and strategy of the firm	**Connection**	C1. Type of relations between colleagues C2. Power of agents who are going to use the IS C3. IS criticality for people of the firm C4. Strength of relationships between colleagues C5. Connection between Requester and Provider
Rule	R1. Laws or regulations applying to the firm R2. Norms/guidelines/standards in the firm R3. Habits, traditions or culture of the firm R4. Recommendations from the management R5. Best practices that apply to the firm	**Granularity**	G1. Atmosphere in the company G2. Legal or financial status of the firm G3. Relevant monitoring metrics of the firm G4. Synergies inside the firm G5. Special facts about the firm

section describes how we collected the data. Section 6 discusses the conclusions that can be drawn from that data, and presents the ETM.

Participants: Participants in phase three are 40 people. Data were initially collected from 51 people, but we rejected answers from those with no experience as stakeholders of IT project. The target group for the survey was defined by randomly selecting people from the alumni's network of the University of Namur. Stakeholders from companies described in Table 1 were also invited to take part in the survey.

Procedure: Data collection took the form of an online survey. Subjects were asked to recall the last project in which they were involved as stakeholders who had been interviewed by business analyses. More precisely, subjects were asked to remember the beginning of the project, when they first got interviewed by a business analyst or equivalent (hereafter BA). Two series of questions were then submitted to subjects. A first series was interested in the Topics themselves. Questions took the following form: "During an interview with the business analyst, would you mention X" where X is to be replaced by one Topic, e.g., the first question of the series takes X= "actors that are going to use the system-to-be (e.g., employees, customers, suppliers, other companies, ...)". Subjects were asked, for each possible X in Topics listed in Table 2, whether they would discuss it with the BA. For each question, the subject had the choice between:

- "A: I would discuss this aspect even if not asked by the BA"
- "B: I would discuss this aspect only if I was asked to do so by the BA"

We interpret A as suggesting that the subject would spontaneously share information on the Topic. We interpret B as suggesting that the information on the Topic would remain implicit, unless the BA asks questions about it. We acknowledge that there could have been more alternatives, e.g., "C: I would be reluctant to discuss this aspect even if asked by the BA". Yet, given the exploratory orientation of this study, we decided to stick to a binary scale. This enables to stay consistent with our initial explicit/implicit distinction [14] and keep simple and easily interpretable results. We refer to the resulting set of answers as Topic evaluation.

In the second part of the questionnaire, subjects were asked to evaluate how frequently, in their own experience, the Topic Sets are discussed with BAs during interviews. In this second section, no Topics are mentioned, and subjects are asked to answer,

Table 3. Result of our Quantitative Study

(a) Topics Evaluations

	I1	I2	I3	I4	I5	R1	R2	R3	R4	R5	L1	L2	L3	L4	L5
A	36	23	24	32	19	12	21	15	21	23	26	9	18	19	5
B	4	17	16	8	21	28	19	25	19	17	14	31	22	21	35
	A1	**A2**	**A3**	**A4**	**A5**	**C1**	**C2**	**C3**	**C4**	**C5**	**G1**	**G2**	**G3**	**G4**	**G5**
A	36	38	35	19	13	7	25	25	8	6	6	5	16	17	4
B	4	2	5	21	27	33	15	15	32	34	34	35	24	23	36

(b) Topic Sets Evaluations

	I	R	L	A	C	G
Never	0	1	2	1	1	9
Very Rarely	2	3	3	0	3	7
Rarely	1	8	5	5	7	8
Occasionally	9	20	16	13	20	12
Very Frequently	26	7	13	18	8	3
Always	2	1	1	3	1	1

considering the Topic Groups from a general point of view. Given our objective to measure frequency, a six-level Likert scale of frequency was proposed to subjects: "Never", "Very Rarely", "Rarely", "Occasionally", "Very Frequently" or "Always". We choose a scale with more than two levels (unlike Topic evaluations) because Topic Sets are more generic and thereby less concrete to the stakeholders. We refer to the resulting set of answers as Topic Sets evaluation.

Results: The collected data are summarized in Table 3a for *Topics evaluations*, and Table 3b for *Topic Sets evaluations*. Results are presented under the form of contingency tables, given that all the variables that we used in our survey are categorical. Numbers reported in the tables are occurrences, e.g., from Table 3a, we learn that I1 has been evaluated as being explicit by 36 of our stakeholders (Answer A), while 4 of them evaluated that same Topic as being implicit (answer B). Heads of the columns are the identifiers from Table 2. We use hereafter the notation *CT=X * Y* to define a contingency table formed by the crossing of the dimension X by the dimension Y. For instance, the contingency table presented in Table 3a would be noted *CT= Topic evaluations*Topics*, while the one in Table 3b would be described by *CT= Set evaluations*Sets*.

6 Data Analysis and the Elicitation Topic Map

We applied Correspondence Analysis (CA) to the data collected in phase three. CA is conceptually similar to Principal Component Analysis: it aims to summarize within two or three dimensions most of the variance of a data set. CA is particularly useful in the scope of our study because it provides a graphical representation for the contingency tables we built from collected answers. Such displays are convenient for identifying patterns in data. CAs were performed with the R package FactoMineR [29]. This section describes the CAs we performed to analyze our data. Next section presents some hypotheses we draw from these analyses in combination with previous qualitative study.

6.1 Analysis of Topics: The Elicitation Topic Map

The most significant output from our quantitative study is the ETM. ETM is obtained from a CA performed on the data presented in Table 3a, i.e. with *CT= Topic evaluations*Topics*. Result of the CA is presented in Figure 1. The graph shows the distances between Topics, and distances between Topics and some Points of Interest (bold text). Points of Interest (PIs) can be seen as the representation, on the diagram, of stakeholders' behaviour regarding the sharing of information: one point represents spontaneous

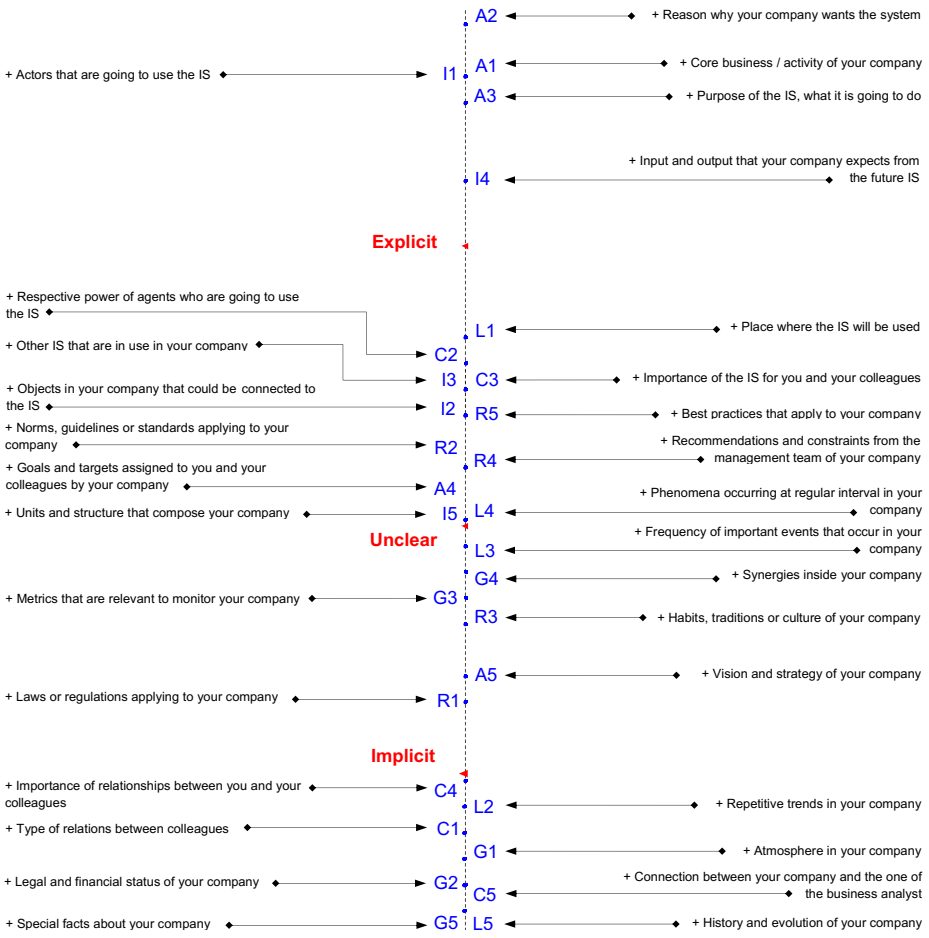


Fig. 1. The Elicitation Topic Map (ETM)

sharing (the label Explicit in Figure 1), another one (Implicit in Figure 1) the tendency not to spontaneously share the information on the topic.

Read the graph as follows: the closer a Topic is to a PI, the more it is associated by our stakeholders to the related sharing behaviour. For instance, L1 can be considered to be an explicit Topic, because it is relatively close to the Explicit PI. Yet, it is less explicit than A1 or A2, because the latter are at a larger distance from Implicit PI.

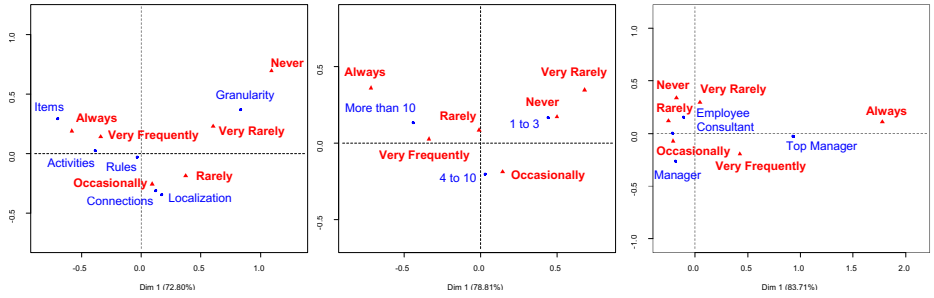
ETM is helpful during elicitation in that it provides indications about the risk of omissions of certain Topics. For example, observe that A5, and A4 to a lesser extent, are closer to the Implicit PI, i.e. they are associated to implicit sharing behaviour. This does not mean that they are not relevant to RE, e.g., understanding the strategy and vision of the company may be critical to make appropriate specification design decisions. However, it means that stakeholders are likely not to mention these Topics spontaneously during interviews. Consequently, the BA might decide to prepare her interview with

questions that focus specifically on understanding the vision, strategy, and targets of the business. It also suggests that it may be useful to the BA to prepare for these interviews by researching the vision, strategy, and targets that the business had already publicly announced in press releases, annual reports, and such.

6.2 Analysis of Topic Sets

Rather than working on Topic evaluations, we now look at the data on Topic Sets evaluations; that data is Table 3b. The mechanisms for presenting and reading the CAs stay the same as for the preceding section: axes are abstract dimensions built to represent the variance within our data set, and are interpreted in the next paragraph.

Figure 2a presents the CA on Table 3b, with *CT= Set evaluations*Sets*. We observe that Activities and Items topics are very close to the *Always* and *Very frequently* PIs. This is interpreted as the fact that our stakeholders tend to spontaneously share information on Topics in these Topic Sets. In sharp contrast, Granularity is close to *Very Rarely* and *Never* answers, thereby suggesting implicit behaviour. Connections, Localization, and to a lesser extent Rules are associated with Occasionally and Rarely answers. Figure 2 can be used in the same way as the ETM. It provides BAs with indications about the expected sharing behaviour of stakeholders toward the Topic Sets, e.g., Figure 2a suggests that it may require more effort to elicit Localizations than Items.



(a) Topic Sets (b) Experience (c) Profile

Fig. 2. Correspondance Analysis on Topic Sets Evaluations

6.3 Analysis of Experience and Profile

We now focus on the analysis of Experience (i.e. the number of projects in which the stakeholder has been involved) and Profile (i.e. the position that the stakeholder was holding in most of these projects). These two characteristics are studied because they are easily identifiable by BAs at the beginning of an interview. Again, the mechanisms for presenting and reading the CAs stay the same as for the ETM section. Our aim here is not to provide a detailed discussion of such characteristics, but rather to illustrate their potential impact. Further research could however go on the investigation of other stakeholders' characteristics that influence sharing of Topics and Topic Sets.

Experience: A CA for the experience of the stakeholder is presented in Figure 2b, and is computed from *CT=Experience*Sets Evaluations*. In our survey, we use three different levels: people who participated from 1 to 3 projects, from 4 to 10 projects, and finally those with more than 10 projects. The analysis suggests that experienced stakeholders are associated with *Very Frequently* to *Always* answers. This suggests an explicit sharing behaviour about Topics. Stakeholders with smaller experience selected more *Very Rarely* and *Never* answers, while stakeholders which took part to 4 to 10 projects favour the *Occasionally* answer.

Profile: A CA for the profile of the stakeholder is presented in Figure 2c. It is computed from *CT=Profile*Sets Evaluations*. We use four groups of profiles: employees (i.e. working for the buyer with negligible responsibilities in the project), consultants (i.e. people from outside the company helping on the project), managers (i.e. people with some responsibilities in the project) and top managers (i.e. CEO/direction of the buying company). The impact of profile on sharing behaviour is less evident, as the distances between our data points (i.e. our profiles) are smaller than in previous figures. It is still possible to observe that stakeholder with more responsibilities - managers and top managers - are more often associated with *Always*, *Very Frequently* and *Occasionally* PIs. On the contrary, stakeholders with less responsibilities - employees and consultants - are more often associated with *Rarely*, *Very Rarely* and *Never* PIs.

7 Implications: Hypotheses about Topic Importance

The ETM we presented in this paper is based on the samples we used, and it is hard to claim much generality to it. In terms of practice, it can be used as a checklist that has the added benefit of suggesting how likely information on some Topics will be shared spontaneously by the stakeholders during elicitation interviews.

From the perspective of research, ETM and the data from our samples suggest a number of hypotheses about information sharing behavior of stakeholders during elicitation interviews. We believe that it is worth doing further empirical research into these hypotheses. We have not yet completed the empirical research to validate these hypotheses. We therefore present them as interesting research issues that may be of interest to the RE community.

The approach here is exploratory: we observe patterns of answers, and then suggest hypotheses that could explain these patterns. The hypotheses are always about sharing behaviour of stakeholders during an interview with a BA, in the scope of an IT project. They should not be considered outside these particular settings. They are to be read as potential explanations why stakeholders behave differently toward different Topics.

7.1 Some Overall Hypotheses about Topic Importance

We are interested here in hypotheses that can be formulated regardless of the Set to which a Topic belongs. Such hypotheses are called overall hypotheses, and are usually dealing with some general characteristics of Topics. In other words, we expect these

hypotheses to hold for any new Topic that is added to the ETM, whatever the Topic Sets to which it may belong. Some overall hypotheses are:

- Information on Topics dealing with information systems (e.g., A3, L1, I3) are spontaneously discussed;
- Information on Topics that pertain to information that stakeholders encounter on a daily basis (e.g., A1, G3, R5) are spontaneously discussed;
- Information on Topics dealing with concrete concepts (e.g., I4, R2, L4) (as opposed to abstract concepts, e.g., C1, G1, A5) are spontaneously discussed.

These three hypotheses (and their opposites), if validated, could be used by interviewers as guidelines for understanding where to seek information that is not represented in the ETM. For instance, an BA may be interested in "The strengths/weaknesses of the firm (SWOT)", which is not represented in the ETM. Using previous hypotheses, she could estimate the Topic is likely to remain implicit during an interview with a BA, because it does not refer to any information system, and deals with abstract concepts. Hence, the BA could decide to include questions in her interview that focus on collecting sufficient information about that supposedly implicit Topic.

7.2 Some Specific Hypotheses about Topic Importance

Some hypotheses can also be suggested, that only apply within a particular Topic Set. The interest of such specific hypotheses for BAs is basically the same as for overall hypotheses. The main difference is that their usage are restricted to Topics existing within the related Topic Set. Some examples of specific hypotheses are listed below. It is important to note that the latter does not list all the possible hypotheses that can be suggested from our results: it simply lists some of the most evident ones.

- Rules that are dictated by the business (e.g., R2, R4, R5) are made explicit;
- Activities about how a business runs (e.g., A4, A5) are kept implicit;
- Localizations that suggest some distance (e.g., L2, L5) are kept implicit;
- Items capable of accomplishing some tasks (e.g., I1, I4) are made explicit;
- Connections involving human relationships (e.g., C1, C4) are kept implicit;
- Granularities with coarse grain (e.g., G5, G1, G2) are kept implicit.

8 Related Work

Importance of context (i.e. environment, domain) is hardly new to RE. Contextualism - which claims that peculiarities of context must be understood before the requirements can be derived - is often presented as an alternative design philosophy to systems design [30]. Papers like [31] - in which it is argued that the machine is to be considered within its environment and cannot be dissociated from it - or [32] - presenting ethnographic analysis as valuable to RE - are further evidence of the importance of context to RE.

As already discussed in this paper, domain modelling languages also emphasize the importance of context in RE [16, 18–21]. This importance has also been highlighted in the NATURE research project, e.g., [33] stresses the importance of a representation

dimension in RE, which copes with the tools (formal or not) that can be used to express knowledge about the system, while [34] propose a conceptual model to support the documentation of domain theories. More recently, authors have emphasized the importance of relating requirements to context. Some emphasize the importance of context and empirical validation of RE models as a direction for future research to accelerate the transfer of research results into RE practice [35]. Others even identify context study as an important research area on which RE should re-focus [36]. Modelling the domain requires information to be collected, and hence elicited. This has also been the center of attention in RE. Efforts have been devoted to the definition of elicitation methods that provide ways for acquiring contextual information. From Contextual Inquiry [37] to Inquiry Cycle [38] , context is put at the center of the acquisition effort. Other approaches indirectly account for the context of use of a system during elicitation. CREWS [39] for instance suggests that elicitation can be guided by the use of scenarios and use-cases. SCRAM [40] also positions scenarios as an important tool for RE. Alternatively, several viewpoints can be adopted to cover different concerns related to a system and therefore support completeness of elicitation (e.g., [41, 42]).

The question of how stakeholders behave during elicitation when being interviewed about context has been the center of less attention in RE. Some research has been devoted to the risks related to stakeholders' behaviour during interviews, e.g., personal, social or cognitive factors, and suggest ways to handle those risks [43]. A framework for the communication issues during elicitation has even been proposed [44]. None of these studies handles the distinction between implicit and explicit information. Still, the existence of implicit information is recognized in RE - through for instance concepts such as Tacit knowledge [45, 46] or Implicit Requirements [47, 48] - and should be accounted for during elicitation.

9 Conclusions and Limitations

In this paper, we discussed the importance of distinguishing between the information stakeholders have that is made explicit during interviews, and the information that they keep implicit. Such distinction brought us to the question of how to discover the implicit information that stakeholders may have. As an answer, we introduced the ETM, a list of RE relevant Topics that are mapped by order of importance. In this paper, importance is understood from the point of view of stakeholders, and express the likelihood of a topic to be discussed explicitly. To build the ETM, we used a combination of a qualitative study (to identify Topics) and quantitative study (to determine the importance of Topics). The ETM enabled us to formulate a set of 9 hypotheses about the sharing behaviour of stakeholders during interviews. Ways for further research are clear: new Topics should be added to the list, and larger-scale validations/replications of already proposed Topics are required, so as to make the ETM more representative. Moreover, hypotheses suggested in this paper, if validated, can make synergies for achieving more complete interviews, and hence, perhaps, systems that fit their requirements better.

Limitations in our study should be kept in mind when using our results. Threats to validity exist, e.g., non-response error, small sampling and selection bias, among others are potential threats to the validity of our study. Also, answers are based on what people

say that they do rather than on a direct observation. These threats might introduce bias to our results, but do not hold us back from drawing relevant preliminary results.

References

1. Christel, M.G., Kang, K.C.: Issues in requirements elicitation. Technical Report CMU/SEI-92-TR-12 ESC-TR-92-012 (1992)
2. Goguen, J.A., Linde, C.: Techniques for requirements elicitation. In: Proc. IEEE International Symposium on Requirements Engineering, pp. 152–164 (1993)
3. Zowghi, D., Coulin, C.: Requirements Elicitation: A Survey of Techniques, Approaches, and Tools. In: Engineering and Managing Software Requirements, pp. 19–46. Springer, Heidelberg (2005)
4. Davis, A.M., Dieste, O., Hickey, A.M., Juristo, N., Moreno, A.: Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review. In: 14th International Requirements Engineering Conference, pp. 179–188. IEEE (September 2006)
5. Zave, P.: Classification of research efforts in requirements engineering. ACM Computing Surveys (CSUR) 29(4), 315–321 (1997)
6. McDermott, D., Doyle, J.: Non-monotonic logic I. Artificial intelligence 13(1), 41–72 (1980)
7. McCarthy, J.: Circumscription - a form of non-monotonic reasoning. Artificial Intelligence 13(1), 27–39 (1980)
8. Reiter, R.: A logic for default reasoning. Artificial Intelligence 13(1-2), 81–132 (1980)
9. Moore, R.: Semantical considerations on nonmonotonic logic. Artificial Intelligence 25(1), 75–94 (1985)
10. Sampaio do Prado Leite, J.C., Gilvaz, A.P.: Requirements Elicitation Driven by Interviews : The Use of Viewpoints. In: IWSSD. (1996) 85–94
11. Lecoeuche, R., Mellish, C., Robertson, D.: A framework for requirements elicitation through mixed-initiative dialogue. In: Proc. 3rd IEEE International Conference on Requirements Engineering, pp. 190–196 (1998)
12. Davey, B., Cope, C.: Requirements Elicitation - What' s Missing? Information Science and Information Technology 5 (2008)
13. Yamanka, T., Komiya, S.: A Method to Navigate Interview-driven Software Requirements Elicitation Work. WSEAS Transactions on Information Science and Applications 7(6), 784–798 (2010)
14. Burnay, C., Jureta, I.J., Faulkner, S.: Context Factors: What they are and why they matter for Requirements Problems. In: Proc. 25th International Conference on Software Engineering and Knowledge Engineering, pp. 30–35 (2013)
15. Burnay, C., Jureta, I.J., Faulkner, S.: Context-Driven Elicitation of Default Requirements: an Empirical Validation. CoRR abs/1211.2 (November 2013)
16. Greenspan, S., Mylopoulos, J., Borgida, A.: Capturing more world knowledge in the requirements specification. In: Proc. 6th International Conference on Software Engineering, pp. 225–234 (1982)
17. Dubois, E., Hagelstein, J., Lahou, E., Ponsaert, F., Rifaut, A.: A knowledge representation language for requirements engineering. Proceedings of the IEEE 74(10), 1431–1444 (1986)
18. Hagelstein, J.: Declarative approach to information systems requirements. Knowledge-Based Systems 1(4), 211–220 (1988)
19. Mylopoulos, J., Borgida, A.: Telos: Representing knowledge about information systems. ACM Transactions on Information Systems 8(4), 325–362 (1990)
20. Dardenne, A., Van Lamsweerde, A., Fickas, S.: Goal-directed requirements acquisition. Science of Computer Programming 20, 3–50 (1993)

21. Yu, E.S.: Towards modelling and reasoning support for early-phase requirements engineering. In: Proc. 3rd International Symposium on Requirements Engineering, pp. 226–235 (1997)
22. Gauker, C.: Zero tolerance for pragmatics. Synthese 165(3), 359–371 (2008)
23. Stalnaker, R.: On the Representation of Context. Journal of Logic, Language and Information 7(1), 3–19 (1998)
24. McCarthy, J., Buvac, S.: Formalizing context (expanded notes) (1997)
25. Dourish, P.: What we talk about when we talk about context. Personal and Ubiquitous Computing 8(1), 19–30 (2004)
26. Baldauf, M.: A survey on context-aware systems. International Journal of Ad Hoc and Ubiquitous Computing 2(4), 263–277 (2007)
27. Lenat, D.: The Dimensions of Context-Space. Technical Report 512 (1998)
28. Walsham, G.: Doing interpretive research. European Journal of Information Systems 15, 320–330 (2006)
29. Lê, S., Josse, J., Husson, F.: FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software 25(1), 1–18 (2008)
30. Potts, C., Hsi, I.: Abstraction and context in requirements engineering: toward a synthesis. Annals of Software Engineering 3(1), 23–61 (1997)
31. Jackson, M.: The meaning of requirements. Annals of Software Engineering 3(1), 5–21 (1997)
32. Viller, S., Sommerville, I.: Social analysis in the requirements engineering process: from ethnography to method. In: Proc. IEEE International Symposium on Requirements Engineering, vol. 6, pp. 6–13 (1999)
33. Pohl, K.: The three dimensions of requirements engineering. In: Advanced Information Systems Engineering (1993)
34. Jarke, M., Pohl, K., Jacobs, S., Bubenko, J., Assenova, P., Holm, P., Wangler, B., Rolland, C., Plihon, V., Schmitt, J.R., Sutcliffe, A., Jones, S., Maiden, N., Till, D., Vassiliou, Y., Constantopoulos, P., Spanoudakis, G.: Requirements engineering: an integrated view of representation, process, and domain (1993)
35. Cheng, B.H., Atlee, J.M.: Research directions in requirements engineering. In: Future of Software Engineering (FUSE), pp. 285–303. IEEE Computer Society (2007)
36. Jarke, M., Loucopoulos, P., Lyytinen, K., Mylopoulos, J., Robinson, W.: The brave new world of design requirements. Information Systems 36(7), 992–1008 (2011)
37. Holtzblatt, K., Jones, S.: Contextual inquiry: A participatory technique for system design. In: Participatory design: Principles and Practice, pp. 180–193 (1993)
38. Potts, C., Takahashi, K., Anton, A.: Inquiry-based requirements analysis. IEEE Software 11(2), 21–32 (1994)
39. Maiden, N.: CREWS-SAVRE: Scenarios for acquiring and validating requirements. Automated Software Engineering 5(4), 419–446 (1998)
40. Sutcliffe, A., Ryan, M.: Experience with SCRAM, a scenario requirements analysis method. In: Proc. 3rd IEEE International Conference on Requirements Engineering (1998)
41. Mullery, G.P.: CORE - A method for controlled requirement specification. In: Proc. 4th International Conference on Software Engineering, pp. 126–135 (1979)
42. Easterbrook, S.: Domain Modelling with Hierarchies of Alternative Viewpoints. In: First IEEE International Symposium on Requirements Engineering (January 1993)
43. Cohene, T., Easterbrook, S.: Contextual risk analysis for interview design. In: Proc. 13th International Conference on Requirements Engineering, pp. 95–104 (2005)

44. Coughlan, J., Lycett, M., Macredie, R.D.: Communication issues in requirements elicitation: a content analysis of stakeholder experiences. Information and Software Technology 45(8), 525–537 (2003)
45. Goguen, J.A.: Formality and Informality in Requirements Engineering. In: Proc. 4th International Conference on Requirements Engineering, pp. 102–108 (1996)
46. Stone, A., Sawyer, P.: Identifying tacit knowledge-based requirements. IEEE Proceedings-Software 153(6), 211–218 (2006)
47. Singer, L., Brill, O.: Utilizing Rule Deviations in IT Ecosystems for Implicit Requirements Elicitation. In: Managing Requirements Knowledge (MARK), pp. 22–26 (2009)
48. Daramola, O., Moser, T., Sindre, G., Biffl, S.: Managing Requirements Using Semantic Case-Based Reasoning Research Preview. In: Requirements Engineering: Foundation for Software Quality, pp. 172–178 (2012)