Evaluating Focus Theories for Requirements Elicitation

Contact author:
Renaud Lecœuche
Room E17
Department of Artificial Intelligence
80 South Bridge
Edinburgh EH11HN
Scotland
e-mail: renaudl@dai.ed.ac.uk

Evaluating Focus Theories for Requirements Elicitation

Renaud Lecœuche[*†]
renaudl@dai.ed.ac.uk

Dave Robertson[†]
dr@dai.ed.ac.uk

Catherine Barry[*]
greboval@insa-rouen.fr

Chris Mellish[†]
chrism@dai.ed.ac.uk

Abstract

Interactive tools for controlling requirements elicitation are normally driven by rules of specification construction, rather than by conventions of human dialogue. On the other hand, theories of dialogue tend to ignore the interaction with constraints imposed by the domain of application. Various competing theories of dialogue exist but there have been few comparative studies of their use in non-trivial tasks. This paper presents a comparative study of the use of focus theories for requirements elicitation which required us to be precise about our interpretation of our chosen focus theories and to develop an innovative means of empirical testing for them.

1 Introduction

Research on requirements elicitation has been mostly concerned with finding ways of representing the requirements formally. Comparatively, little attention has been paid to the elicitation process (Rolland 1994). Therefore, little guidance is offered to the users of requirements elicitation systems. This is particularly true for requirements elicitation through natural language dialogue. Natural language is inherently ambiguous (Meyer 1985) and therefore prone to introduce errors in a specification. Moreover formal methods have matured and are making some inroads in requirements engineering (Delisle and Garlan 1990; Hall 1990; Spivey 1990; Bowen 1995). Nevertheless, most people are still not accustomed to deal directly with formal notations. Providing them with a natural language interface to a formal system seems therefore a promising idea. A majority of such current interfaces are based on a translation approach which partly hides the complexity of the underlying representation framework (Fuchs and Schwitter 1996; Nanduri and Rugaber 1996). Most of these approaches are based on subsets of natural language which can be automatically interpreted and translated into a formal representation (see Mich (1996) for an attempt at full natural language treatment). Using a subset of natural language also helps in keeping the language simple and concise (Sommerville and Sawyer 1997, good pratice 6.2). However translation approaches provide little guidance to those who write the specifications. Because the

[*]PSI-LIRINSA, I.N.S.A., Place Emile Blondel, 76130 Mont-Saint-Aignan, France
[†]Department of Artificial Intelligence, 80 South Bridge, Edinburgh EH1 1HN, Scotland

translation is one-way, with no or minimum feedback, systems cannot provide help or guidance to their users. Since some elicitation systems can offer some support to their users in incrementally defining the specifications (Darimont 1995; Souquières 1993; Souquières and Lévy 1996), the translation approach appears too restrictive (see however Rolland and Ben Achour (1998) for an attempt at providing guidance with a translation approach).

We have been developing a requirements elicitation system using a natural language dialogue, with a particular emphasis on providing guidance in the dialogue. The system is mainly composed of an elicitation module managing the requirements and a dialogue manager managing the dialogue with users.

There are three aspects that are usually taken into account when managing dialogues: the informational aspect, the intentional aspect (Moore and Paris 1993) and the attentional aspect (Grosz and Sidner 1986). We present them here in an example about the design of WWW sites. This domain is the application domain of our system. The system has some knowledge of how WWW sites are built and helps a user design his or her own site.

Example 1

U: I want a link between the publication page and the home page.
S: (a) Do you want a link to or from the home page ?
(b) Let's consider other navigational devices.
(c) Do you want to link the publication page to other pages ?

The informational aspect deals with the information contained in a message. In example 1 where a user expresses a requirement, the information is that the two pages must be related. (A complete elicitation dialogue is presented in appendix C.) Checking the correctness of the information conveyed is difficult. We should ensure that it is consistent and complete among other factors (Davis et al. 1993). In the example, the information is not complete. Because links are oriented, we need to know if the link is going from the publication page to the home page or vice-versa. In that case, the system could reply with sentence (a). Since making correctness checks is complex, we believe this should be done by the dedicated elicitation module rather than by the dialogue manager. The dialogue manager will just pass on the information and report back eventual problems.

The intentional aspect deals with the intention driving the utterances in the dialogue. In example 1, the intention might be to speak about navigation through the pages. In that case the system could provide guidance by proposing, as in sentence (b), to speak about other navigational devices such as a navigator or site map. However, it is not clear whether taking into account intentions offers benefits for an elicitation system. Given the difficulty of recognising intentions, other methods could be more effective (Cohen et al. 1990). For example, if a link is classified as a navigational device along with a navigator and site map, it could be a good idea to mention them whenever we speak about links, irrespective of the possible intentions of the speaker (see however Si-Saïd and Rolland (1998) for guidance based on intention tracking).

The attentional aspect deals with what people are paying attention to at a particular point in the dialogue. As shown above, this aspect may serve to direct the users' attention to related topics they may want to speak about. This aspect can also be used to restrict what can be spoken about. For example, the system could try with sentence (c) to know everything about the publication page now that it has been mentioned. Elicitation modules are able to deal with large specifications where problems can be detected in several places. However users have a limited attentional capability (Miller 1956). Taking the attentional aspect into account leads to restricting the dialogue to what users can deal with at the time.

Based on these remarks, we built a dialogue manager for an elicitation system primarily concerned with ensuring a good attentional aspect for the interaction. The informational aspect is dealt with by the elicitation module. The intentional aspect is difficult to take into account and is ignored in our system.

In our system, guidance is provided by focus rules which define how the dialogue can evolve (Lecœuche et al. 1998). Three questions need to be answered to achieve this goal: (1) On which attentional theory should we base the dialogue manager ? (2) How can it be implemented ? (3) Does it provide benefits over other techniques ? Although we answer these questions in the requirements engineering domain, our results have wider implications. The analysis and comparison of the focus rules carried out can be re-used in other domains. The evaluation and development method we propose can also be adapted for other purposes.

In this paper, we first present in section 2 the architecture of our system. Then we describe in depth in section 3 formalisations of the focus theories we use. These formalisations provide rules constraining the dialogue evolution. An example of how these rules are used is then given in section 4. In section 5, we present an innovative method of developing and evaluating our system. We describe how we used this method to develop and evaluate the theories presented and compare them. Finally we conclude in section 7.

2 System Architecture

In this section we present the general architecture of our system and we describe its main components.

The domain in which the system is used is the construction of WWW sites for research groups. The system has some notions of what a research group involves (researchers, projects, publications) and how WWW sites are designed (site, page, link)[1]. It then tries to determine how a particular research group is to be presented. The result is a specification of a WWW site showing how the information is distributed among pages and how it is possible to navigate between them. Although we base our presentation on this domain, the system can be used in other domains.

We now describe the main components of the system. Its working will be explained in more detail in section 4. The general architecture of the system is shown in figure 1.

The general system architecture is composed of the following data stores:

Domain Specific Specification Model This contains the domain specific knowledge of the system about what can be elicited. This model is used during the elicitation process to direct the system. It is also used for correctness and completeness checking.

In our system, this model is based on an ER model (Chen 1976) augmented with two additional types of constraints (see table 1):

[1] A site is a set of pages, one of them being a home page. Pages can be linked together. Several other entities exist in our system but we do not present them here.

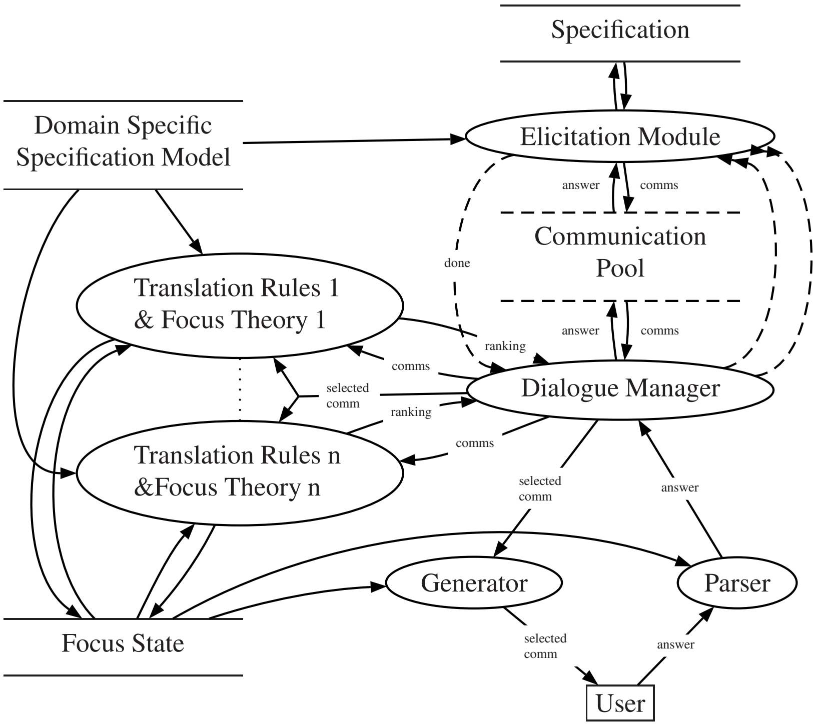

Figure 1: System architecture and transformation diagram (see Ward (1986) for a description of the notation used)

- Cardinality restrictions on types can be specified. This enables us to define the minimum and maximum numbers of entities of a particular type.
- Cardinality restrictions on relationship roles can be specified. A role is the function played by entities in a relationship. For example, in figure 2 presenting part of the model constraining the creation of a research group and site, the relation `involve` has two roles, namely, `involved in` and `involving`. The latter can be filled by either `publication set` entities or `researcher set` entities. When a role can be filled by several types of entities, the minimum and maximum numbers of rolefillers for the role can be specified. This is especially useful when roles can filled by either of two entities but not both (in this case the maximum cardinality for the role is 1). In all cases:

$$rolecardmin(Role,N) \geq \sum_{T_e} 1_{\cdot rolefiller(T_r,Role,T_e)} \times cardmin(T_r,Role,N)$$

$$rolecardmax(Role,N) \leq \sum_{T_e} 1_{\cdot rolefiller(T_r,Role,T_e)} \times cardmax(T_r,Role,N)$$

Expression	Interpretation
Model level	
$entityset(T_e)$	T_e is a set of entities (of that type)
$relationset(T_r)$	T_r is a set of relations (of that type)
$cardmin(T_e,N)$	The minimum cardinality of T_e is N
$cardmax(T_e,N)$	The maximum cardinality of T_e is N
$attribute(T_e,A,T_a)$	Entities of type T_e have an attribute A of type T_a
$role(T_r,Role)$	Relations of type T_r have a role $Role$
$rolefiller(T_r,Role,T_e)$	Entities of type T_e can be used to fill the role $Role$ of relations of type T_r. More than one type of rolefiller can be specified
$cardmin(T_r,Role,N)$	The minimum cardinality of the role $Role$ for relation of type T_r is N
$cardmax(T_r,Role,N)$	The maximum cardinality of the role $Role$ for relation of type T_r is N
$rolecardmin(T_r,Role,N)$	The minimum cardinality of role $Role$ for relation of type T_r is N
$rolecardmax(T_r,Role,N)$	The maximum cardinality of role $Role$ for relation of type T_r is N
Instance level	
$entity(E,T_e)$	Entity E is of type T_e
$attributevalue(E,A,V)$	Entities E has an attribute A of value V
$relation(R,T_r)$	Relation R is of type T_r
$rolefillervalue(R,Role,E)$	Entity E fills the role $Role$ of relation R

Table 1: Entity-Relationship (ER) Model notation

where $1._{p(x_1,\ldots,x_n)} = 1$ if $p(x_1,\ldots,x_n)$ is true and 0 otherwise. These equations mean that the role cardinalities should represent real constraints compared to the constraints already imposed on the model by the relation cardinality restriction. The equalities hold if not specified otherwise.

The part of the model constraining the creation of a research group and site presented in figure 2 is detailed in table 9 in appendix A. (In particular, the cardinality restrictions have been omitted from the figure.)

This example shows the use of a role cardinality restriction to constrain sites and pages to present at least one entity, which can either be a research group, a researcher set or a publication set.

Specification This represents the requirements that have been elicited.

In our system, the requirements are represented in an ER model (see bottom of table 1). For example, the fact that a research group is presented by a site would be represented as follows:

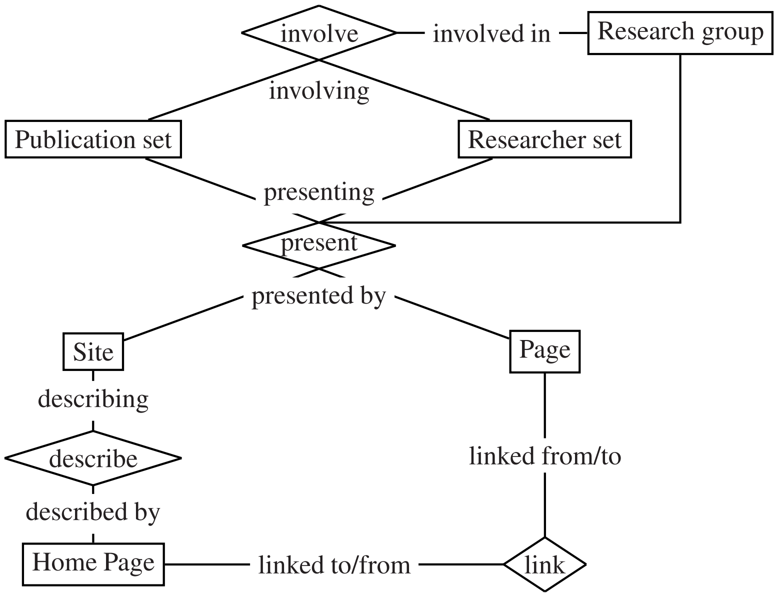

Figure 2: Example of ER model – graphical notation

```
entity(rg1, research group)
entity(s1, site)
relation(p1, present)
rolefillervalue(p1, presented, rg1)
rolefillervalue(p1, presented by, s1)
```

The specification can also contain some "hypothetical" entities and relations that the elicitation system assumes will be created but have not been yet asserted by the user. These entities and relations are marked by a question mark.

An ER model may not be the better way of representating a specification in general. In particular dynamic behaviour is difficult to represent with an ER model. Other aspects of the specification such a rationale, sources and dependencies are not represented either. However, we chose an ER model for its simplicity and because it is well suited for describing static domains such as the one dealt with here (Davis 1988, 1993; Wieringa 1996).

Communication Pool This contains the communications to be output by the system and the answers provided by the users.

In our system, the elicitation module and the dialogue manager interact by means of a predefined set of communication types. The elicitation module puts communications of these types in the communication pool and the dialogue manager contributes (eventual) answers to the communications as they become available. Each communication has one *main subject* which is either a thing to be presented or to be created, and possibly some other *subjects* which provide some context for the communication. The notation used is `communication_type(mainsubject | other subjects)`. Communications can be of two types:

- Presentation communications are used to inform users of an action made by the elicitation module, such as the creation of an entity or a relation. They can also indicate dialogue actions such as the change to a new topic. For example, `presentation(rg1)` is a communication presenting research group `rg1`.
- Question communications are used to prompt users for information. Questions serve to fill in the specification. For example, `question(?s1 | p1, presented by)` is a question about the identity of `?s1` with the context information, given after the vertical bar, that `?s1` is the rolefiller of role `presented by` of relation `p1`.

Focus State This data store is used by the focus theories to keep track of the focus evolution. The nature of the information stored depends on the theory (see section 3).

These data stores are used by the following programs:

Elicitation Module Depending on the current state of the specification and on the domain specific specification model, the elicitation module generates possible communications and interprets answers by the users. Depending on the answer of the users, it modifies the specification. It can also remove communications from the communication pool. This gives much flexibility to the elicitation system in its communication decisions, but makes the task of the dialogue manager more complex.

In our system, the elicitation system outputs all the questions it can and makes all the modifications possible given the users' answers. The general algorithm driving the module is the following.

- Create all entities and relations that should exist. This depends on the minimal cardinalities specified in the domain specific specification model.
- Add presentation communications to the communication pool presenting the things that are modified in the specification. In particular, all entity and relation creations are presented as well as relations' new role fillers.
- Add question communications to the communication pool asking for missing information, such as the creation of new entities or relations, and rolefillers.
- Interpret answers to question communications and modify the specification accordingly.
- Remove question communications from the communication pool that would lead to a violation of constraints. This depends on the maximal cardinalities specified in the domain specific specification model and on the modifications made to the specification since the communication has been placed in the communication pool.

Dialogue Manager This program calls the focus theories and selects a communication from the communication pool to output based on their recommendation. Several focus theories can be used together. For example a focus theory can check for the overall coherence of the dialogue (we will call such a theory a "global focus theory") while another focus theory checks for sentence-to-sentence coherence (we will call such a theory a "local focus theory"). The dialogue manager then gives the selected communication to the Natrual Language Interface (NLI) composed of a natural language generator and a natural language parser. This dialogue manager differs from other dialogue managers using focus rules (Huang 1994; McKeown 1985; Reed and Long 1997a) by the fact that it cannot plan reliably ahead in the text since the communication pool may be modified at any time by the elicitation module. If the dialogue manager was planning the output by relying on having several communications in the communication pool, there would always be a risk for example that a communication which was part of the output plan be removed.

In our system, the dialogue manager has access to different focus strategies that can be "plugged-in" the system. The dialogue manager calls them in a predefined order. (This order can however be changed at run-time by the system). Theory 1 proposes a first partial ranking which is then refined by theory 2 and so on until the last theory N. The highest ranked communication is then selected and output. Using this method we are able to generate texts based only on the local context, without planning (Sibun 1992), by using local focus theories but we are also able to ensure global coherence by adding global focus theories (see section 3). Therefore we can adapt the level of text structure depending on the need of the dialogue. This of course depends on the communications available in the communication pool and therefore to a certain extend to the elicitation module (see section 4).

In the case where focus theories are unable to distinguish between competing communications, presentation communications are prefered over question communications. This is because it seems better to provide information before asking questions. If communications are still competing, one is selected at random.

Generator The role of the generator is to output the communications in natural language.

In our system, the output is based on templates which can be slightly adapted depending on the context (Reiter 1995). For example, the generator makes use of the focus theories results (if available) to make decisions about pronominalisation, i.e., the use of pronouns such as "it", and clue words selection, i.e., the use of words such as "now" and "anyway".

Parser The role of the parser is to interpret the answers of the users based on a controlled language. It transforms them into a form that can be used by the dialogue manager and elicitation module.

In our system, the parsing is based on a set of templates. The templates are divided into three groups:

- Question answering templates enable users to answer questions asked by the system. An example of a sentence that a user could enter is: "Research group 1 is presented by a new site".
- Automatic question answering templates enable users to set the NLI so that it automatically answers questions without outputing them. An example of a sentence that a user could enter is: "Every new page is linked to the homepage".

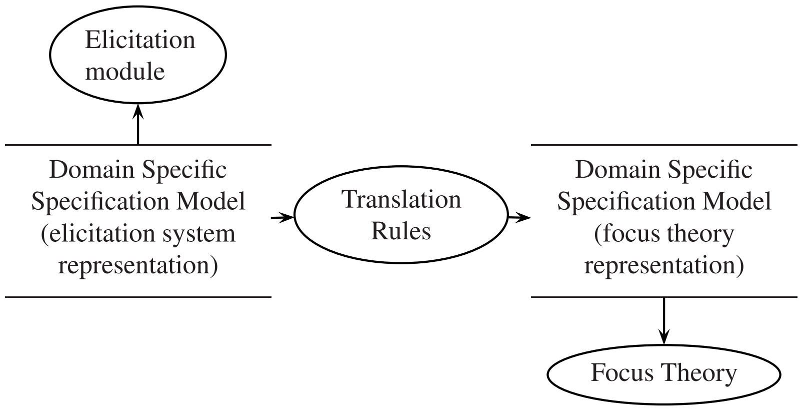

Figure 3: Translation rules

- Dialogue directing templates enable users to shift the focus of the dialogue to a new topic. (This capability depends on the focus theories used.) An example of a sentence that a user could enter is: "I want to speak about the research group".

The parser also performs some anaphora resolution, i.e., finds the nouns pronouns refer to, based on the available focus information.

Translation Rules & Focus Theory The theories rank the available communications in order to ensure a coherent dialogue. In order to perform this task, some focus theories need to access domain knowledge stored in the domain specific specification model. The translation rules bridge the gap between the representation used for the domain specific specification model and the representation used by the focus rules (see figure 3). Not all focus theories use the domain specific specification model. In that case the theories do not need translation rules.

Once a communication has been selected, the focus theories are called again by the dialogue manager to update the focus state.

The theories we have been using in our system are presented in-depth in section 3.

The various elements presented above are activated in a cycle shown in figure 4 and described below:

1. The elicitation module is activated by the dialogue manager. It then modifies the specification and modifies the communication pool if needed. For example, at the start of the elicitation process, the system creates an entity representing a research group and puts a communication presenting it in the communication pool. The eliciation module then sends a signal to the dialogue manager indicating the end of the modifications. It is then de-activated.

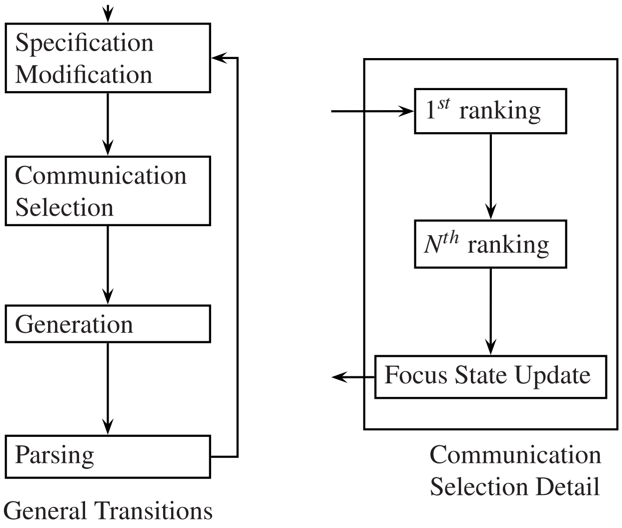

Figure 4: System state diagram

2. The dialogue manager selects a communication. This is done by asking the focus theories to rank the communications (see detail in figure 4). In the case of the communication created above, it would be selected since this is for now the only one in the communication pool. Once a communication has been selected, the focus state is updated.

3. The generator outputs the communication. The generator would output here something like “Let’s speak about the research group”.

4. The parser interprets the user’s answer. In this case the user could simply acknowledge the communication or try to redirect the dialogue by providing new information.

The dialogue manager plays a pivotal role in this cycle by activating the other elements when needed (see figure 1)

3 Formalising focus theories

The attentional aspect of dialogues is the subject of numerous theories called focus theories. However, most of these theories are not stated clearly enough to be easily implemented. In this section we present the formalisation of two theories. The first theory is a global focus theory organising the dialogue at a high level. The second theory is a local theory organising the dialogue on a local basis.

An extended example of the use of these theories is given in section 4.

Expression	Interpretation
$\mathcal{F}_{i,t}$	A focus space when t exchanges have been carried out
$active(\mathcal{F}_{i,t})$	Focus space $\mathcal{F}_i$ is active when t exchanges have been carried out
$controlling(\mathcal{F}_{i,t}, \mathcal{F}_{j,t})$	Focus space $\mathcal{F}_i$ is controlling focus space $\mathcal{F}_j$ when t exchanges have been carried out
$closed(\mathcal{F}_{i,t})$	Focus space $\mathcal{F}_i$ is closed when t exchanges have been carried out
$\mathcal{C}_u$	$\mathcal{C}_u$ is a communication, i.e., a message to be output. It makes the dialogue evolve from time t to $t+1$.
$subject(\mathcal{C}_u,X)$	X is a subject of communication $\mathcal{C}_u$
$mainsubject(\mathcal{C}_u,X)$	X is the main subject of communication $\mathcal{C}_u$. The communication main subject is a communication subject
$dir_t(D_i,X_1,X_2)$	X_1 is in direct relation D_i with X_2 at time t
$spe_t(S_i,X_1,X_2)$	X_1 is in specialisation relation S_i with X_2 at time t
$gen_t(G_i,X_1,X_2)$	X_1 is in generalisation relation G_i with X_2 at time t
$rel_t(R_i,X_1,X_2)$	X_1 is related to X_2 by R_i, other than by direct relation, specialisation or generalisation at time t
$T \in \Sigma$	The term T appears in set Σ
$T \notin \Sigma$	The term T does not appear in set Σ
$\neg P$	P is false
$P_1 \wedge P_2$	P_1 and P_2 are true
$P_1 \rightarrow P_2$	P_1 implies P_2

Properties
$\forall t.dir_t(D_i,X_1,X_1)$
$\forall t.(X_1 \in \mathcal{F}_{i,t} \wedge X_2 \in \mathcal{F}_{i,t} \rightarrow dir_t(D_i,X_1,X_2))$
$\forall t.(spe_t(S_i,X_1,X_2) \rightarrow gen_t(G_j,X_2,X_1))$

Table 2: Notation used to formalise the global focus rules

3.1 Global Theory

Our formalisation is based on a simplification of Reichman's theory (Reichman 1978, 1981, 1984). This theory is a global focus theory, i.e., a theory organising dialogues at a high level. The basic idea is that the dialogue is supported by a sequence of changes to the focus space set, $\mathcal{S}$. A focus space represents the information in focus during part of the dialogue, i.e., it contains all the things to which participants in a dialogue are attending at a certain point in a dialogue. If $\mathcal{S}_0 = \{\mathcal{F}_{1,0}\}$ is the initial focus space set, only containing the initial focus space $\mathcal{F}_{1,0}$ at the beginning of the dialogue, then the set when t exchanges have been carried out in the dialogue is $\mathcal{S}_t = \{\mathcal{F}_{1,t}, \ldots, \mathcal{F}_{n,t}\}$ where each $\mathcal{F}_{i,t}$ contains some of the things spoken about in the dialogue. We may need to create several focus spaces, indexed by i, since the dialogue may deal with more than one topic. Focus spaces have different activation levels (Reichman 1985, p. 54):

Active This is the space to which current communications are added.

Controlling These are the spaces expected to become active again when the space they control is closed.

Closed These are the spaces which have been dealt with and are not expected to be returned to.

The dialogue evolves when communications are output. Communications cause things to be included in focus spaces. They can also cause focus space activations to be modified and/or focus spaces to be created. Each communication has one *main subject*, i.e., the most important thing the communication is about, and possibly some other *subjects*, i.e., the other things the communication mentions. The exact nature of the communications depends on the application domain.

The focus rules define the possible ways the dialogue may develop and the constraints to satisfy for choosing a particular way. Associated with an ordering, they represent what we are expecting to say next in the dialogue.

We define six focus rules, each representing an alternative way of continuing the dialogue. Four rules create new spaces and two (pop and reopening) change the activation levels of existing spaces. The four creating moves are differentiated along two criteria: (1) moves that keep/do not keep the current focused entities in focus (2) moves that are expected/not expected to shift back to the current focus space.

The moves are based on three relations between the things that can be mentioned in the dialogue:

Direct relation There is a direct relation between two things if they are closely related and can be mentioned in the same focus space.

Specialisation relation There is a specialisation relation between two things if one of them is dependent on the other. In that case the second thing can be discussed in the perspective of the controlling one.

Generalisation relation This is the inverse relation to the specialisation relation.

The exact nature of these relations depends on the application domain. Examples of translation between domain relations and these three relations are given in section 3.1.2.

Focus rules only indicate what changes in the focus spaces. Everything else stays the same, and in particular $\forall i, \forall t, \forall X.(X \in \mathcal{F}_{i,t} \rightarrow X \in \mathcal{F}_{i,t+1})$. This means in particular that nothing can be removed from a focus space. All variables are assumed to be existentially quantified unless explicitly universally quantified. The focus space activation levels are also kept unchanged, unless stated otherwise. The notation used to represent the rules and some properties are shown in table 2.

The interpretation of rule 1 for example is then the following: If the current active focus space is $\mathcal{F}_i$ and if we can select a communication $\mathcal{C}_u$ such that its main subject is in direct relation with an element of $\mathcal{F}_i$ and this relation is itself a subject of $\mathcal{C}_u$, then we can perform a no change focus move. This move has the effect of putting all the subjects of $\mathcal{C}_u$ in space $\mathcal{F}_i$.

Rule 1 (No change) *The focus space does not change and new information is added to it.*

$$\left(\begin{array}{r} active(\mathcal{F}_{i,t}) \wedge \\ X_1 \in \mathcal{F}_{i,t} \wedge \\ mainsubject(\mathcal{C}_u, X_2) \wedge \\ dir_t(D_i, X_1, X_2) \wedge \\ subject(\mathcal{C}_u, D_i) \end{array}\right) \rightarrow \forall X.(subject(\mathcal{C}_u, X) \rightarrow X \in \mathcal{F}_{i,t+1})$$

Rule 2 (Resetting) *A new focus space is created.*

$$\left(\begin{array}{r} active(\mathcal{F}_{i,t}) \wedge \\ X_1 \in \mathcal{F}_{i,t} \wedge \\ mainsubject(\mathcal{C}_u, X_2) \wedge \\ gen_t(G_i, X_1, X_2) \wedge \\ subject(\mathcal{C}_u, G_i) \end{array}\right) \rightarrow \left(\begin{array}{l} \neg active(\mathcal{F}_{i,t+1}) \\ \wedge\, closed(\mathcal{F}_{i,t+1}) \\ \wedge\, active(\mathcal{F}_{n+1,t+1}) \\ \wedge\, \forall X.(subject(\mathcal{C}_u, X) \\ \quad \rightarrow X \in \mathcal{F}_{n+1,t+1}) \end{array}\right)$$

Rule 3 (Additive) *A new focus space is created. It is controlled by the current active space. Entities in the current active space are copied to the new space.*

$$\left(\begin{array}{r} active(\mathcal{F}_{i,t}) \wedge \\ X_1 \in \mathcal{F}_{i,t} \wedge \\ mainsubject(\mathcal{C}_u, X_2) \wedge \\ spe_t(S_i, X_1, X_2) \wedge \\ subject(\mathcal{C}_u, S_i) \end{array}\right) \rightarrow \left(\begin{array}{l} \neg active(\mathcal{F}_{i,t+1}) \\ \wedge\, active(\mathcal{F}_{n+1,t+1}) \\ \wedge\, controlling(\mathcal{F}_{i,t+1}, \mathcal{F}_{n+1,t+1}) \\ \wedge\, \forall X.(X \in \mathcal{F}_{i,t} \rightarrow X \in \mathcal{F}_{n+1,t+1}) \\ \wedge\, \forall X.(subject(\mathcal{C}_u, X) \\ \qquad \rightarrow X \in \mathcal{F}_{n+1,t+1}) \end{array}\right)$$

Rule 4 (Generating) *A new focus space is created. Entities in the current active space are copied to the new space.*

$$\left(\begin{array}{r} active(\mathcal{F}_{i,t}) \wedge \\ X_1 \in \mathcal{F}_{i,t} \wedge \\ mainsubject(\mathcal{C}_u, X_2) \wedge \\ rel_t(R_i, X_1, X_2) \wedge \\ subject(\mathcal{C}_u, R_i) \wedge \end{array}\right) \rightarrow \left(\begin{array}{l} \neg active(\mathcal{F}_{i,t+1}) \\ \wedge\, closed(\mathcal{F}_{i,t+1}) \\ \wedge\, active(\mathcal{F}_{n+1,t+1}) \\ \wedge\, \forall X.(X \in \mathcal{F}_{i,t} \rightarrow X \in \mathcal{F}_{n+1,t+1}) \\ \wedge\, \forall X.(subject(\mathcal{C}_u, X) \\ \qquad \rightarrow X \in \mathcal{F}_{n+1,t+1}) \end{array}\right)$$

Rule 5 (Pop) *A controlling space becomes active again.*

$$\left(\begin{array}{r} active(\mathcal{F}_{i,t}) \wedge \\ controlling(\mathcal{F}_{j,t}, \mathcal{F}_{i,t}) \end{array}\right) \rightarrow \left(\begin{array}{l} \neg active(\mathcal{F}_{i,t+1}) \\ \wedge\, closed(\mathcal{F}_{i,t+1}) \\ \wedge\, active(\mathcal{F}_{j,t+1}) \end{array}\right)$$

Rule 6 (Digressing) *A new focus space is created. It is controlled by the current active space.*

$$\left(\begin{array}{r} active(\mathcal{F}_{i,t}) \wedge \\ mainsubject(\mathcal{C}_u, X_1) \wedge \\ \forall j.(X_1 \notin \mathcal{F}_{j,t}) \end{array}\right) \rightarrow \left(\begin{array}{l} \neg active(\mathcal{F}_{i,t+1}) \\ \wedge\, active(\mathcal{F}_{n+1,t+1}) \\ \wedge\, controlling(\mathcal{F}_{i,t+1}, \mathcal{F}_{n+1,t+1}) \\ \wedge\, \forall X.(subject(\mathcal{C}_u, X) \\ \qquad \rightarrow X \in \mathcal{F}_{n+1,t+1}) \end{array}\right)$$

Rule 7 (Reopening) *An old space becomes active again.*

$$\left(\begin{array}{r} active(\mathcal{F}_{i,t}) \wedge \\ closed(\mathcal{F}_{j,t}) \wedge \\ mainsubject(\mathcal{C}_u, X_1) \wedge \\ X_1 \in \mathcal{F}_{j,t} \wedge \\ \forall R.(\neg \exists X.(X \in \mathcal{F}_{i,t} \wedge \\ (dir_t(R, X, X_1) \vee spe_t(R, X, X_1) \vee \\ gen_t(R, X, X_1) \vee rel_t(R, X, X_1)))) \end{array}\right) \rightarrow \left(\begin{array}{l} \neg active(\mathcal{F}_{i,t+1}) \\ \wedge\, closed(\mathcal{F}_{i,t+1}) \\ \wedge\, \neg closed(\mathcal{F}_{j,t+1}) \\ \wedge\, active(\mathcal{F}_{j,t+1}) \\ \wedge\, \forall X.(subject(\mathcal{C}_u, X) \\ \qquad \rightarrow X \in \mathcal{F}_{j,t+1}) \end{array}\right)$$

Although the main characteristics of Reichman's approach are conserved, the theory presented above is a simplified version of the original approach. In particular, fewer types of focus spaces and fewer types of relations between focus spaces are used. Also, each focus space contains less information than in Reichman's theory or Hitzeman and Poesio's approach (Hitzeman and Poesio 1998). On the other hand, this theory is richer than Grosz and Sidner's approach (Grosz and Sidner 1986) since we distinguish between normal evolution of the dialogue and digressions. We also keep track of closed spaces. This balance between simplicity and complexity seems adequate for our purposes.

3.1.1 Ordering of the rules

Several rules may have their preconditions satisfied at the same time. We prefer then to apply the rule that maintains the focus if possible, or minimises its movement. Changing focus, i.e., changing the entities to which participants are paying attention, does require some cognitive processing. Therefore, minimising focus movement reduces the cognitive overhead needed to understand the dialogue. We minimise the focus movements by presenting general concepts before specialised ones and by avoiding references to unrelated concepts. Rules are therefore applied in the following preference order:

> no change $>$ resetting $>$ additive $>$ generating $>$ pop $>$ digressing $>$ reopening.

This means that we first try to find a communication to be output which would allow a "no change" move. Then, if such a communication does not exist, we would try to find a "resetting" communication and so on until a communication is found.

There is an exception to this ordering. If the main subject of a possible communication to output is already a member of a controlling space then a pop is the preferred move. This avoids reintroducing concepts that we are expecting to return to later in the dialogue. An example of the use of this theory are given in section 4.

3.1.2 Translation rules

Because the focus rules we presented in section 3.1 are based on generic concepts such as specialisation or generalisation which are not necessarily used by the elicitation module, we need to bridge the gap between the representation used by the elicitation module and the representation used by the focus rules (Kittredge et al. 1991). This is done by "translation rules". These rules also allow the use of this focus theory in different domains.

We have created a set of translation rules to map the ER representation used by the elicitation system (see section 2) to the generic concepts used by our formalised focus theory. We present in table 10 in appendix B a simplified version of these rules. The main idea behind these rules is to allow relations and entities that are related to the current focus and cannot be accessed easily otherwise to be mentioned immediately. (Their actual selection will depend on the current focus state and the focus rules.) This is equivalent to McKeown's rules (McKeown 1985, p. 64). Relations "branching out" from the same entity are considered as specialisations. This is equivalent to Grosz and Sidner's dominance relation (Grosz and Sidner 1986). The other relations are not particularly marked with regard to focus moves. (An example of the result of the translation rules application is given in figure 5 in section 4). Basing the translation rules on the structure of the ER model has the advantage of not being dependent on the semantics of the relations which is a major problem for some theories (Hahn and Strube 1997; Polanyi 1985; Schank 1977). We are still dependent on the semantics of the ER model. However this dependency is at a higher level than domain specific ones. It is therefore usually easier to adapt the system to new domains. This advantage can however be a drawback for some domains where the structure of the model is less well related to the structure of the dialogue than for our application domain. For example, attributes are treated as independent in our system. This could be a problem if attributes can be grouped into clusters. For example some attributes could deal with colour of objects (background colour, border, colour, etc.) while others could deal with their size (length, height, etc.). Mixing these attributes would make the text less coherent than if they were discussed in two clusters. In this case, translation rules would have to be more domain-dependent to classify the attributes in their clusters. Reaching a balance between domain dependence and translation power is a difficult task.

Changing the translation rules gives to the theory a more local or more global aspect. If more direct relations are generated, entities and relations will be grouped in a few large focus spaces. If more specialisation relations are generated, entities and relations will be grouped in numerous small focus spaces. In our case, the former case is preferred as local focus decisions are the task of a local focus theory (see section 3.2).

3.1.3 Surface phenomena

Focus theories do not only help in organising the dialogue but also indicate how this organisation can be made explicit in the dialogue. For example, creating a new focus space can be marked in the dialogue by a cue word such as "now". Because the focus theory organises the dialogue into focus spaces, it is also possible to know when a topic is returned to as a normal part of the dialogue (as opposed to re-opening a closed focus space). In that case, the natural language generator can provide useful cues as shown in example 2 where it explicitly redirects the dialogue. (This use of cue words is also made in Cohen (1984, 1987); Reed and Long (1997b).)

Example 2

S: *Let's come back to our previous topic*
Researcher set 7 is involved in research group 1

3.2 Local theory

Our formalisation is based on a simplification of the centering theory (Grosz et al. 1995). This theory is a local focus theory, i.e., a theory organising dialogues on a local basis. The basic idea is that each sentence introduces things into the dialogue. The next sentence then selects one of these things as the topic of the dialogue. For example, assuming I just said: "I bought a new hat at the shop round the corner yesterday", if the next sentence is: "I am going to a wedding" then the dialogue is about me but if the next sentence is: "It is nice, isn't it" then the dialogue is about the hat.

Each sentence in the dialogue is therefore associated with some "forward-looking centres", i.e., the things introduced by the sentence, and to a "backward-looking centre", i.e., the topic selected by the sentence from the previous sentence's forward-looking

Expression	Interpretation
$Cb_t(X)$	X is the backward-looking center at time t
$Cf_t(X_1,\ldots,X_n)$	$X_1,\cdots,X_n$ are the forward-looking centers at time t
$Cp_t(X)$	X is the preferred centre at time t
$Cb_t(\mathcal{C}_u,X)$	X would be the new backward-looking centre if communication $\mathcal{C}_u$ was output at time t
$Cf_t(\mathcal{C}_u,X_1,\ldots,X_n)$	$X_1,\cdots,X_n$ would be the new forward-looking centres if communication $\mathcal{C}_u$ was output at time t
$Cp_t(\mathcal{C}_u,X)$	X would be the new preferred centre if communication $\mathcal{C}_u$ was output at time t
$\mathcal{C}_u$	$\mathcal{C}_u$ is a communication, i.e., a message to be output. It makes the dialogue evolve from time t to $t+1$.
$subject(\mathcal{C}_u,X)$	X is a subject of communication $\mathcal{C}_u$
$mainsubject(\mathcal{C}_u,X)$	X is the main subject of communication $\mathcal{C}_u$. The communication main subject is a communication subject
$T \in \Sigma$	The term T appears in set Σ
$T \notin \Sigma$	The term T does not appear in set Σ
$\neg P$	P is false
$P_1 \wedge P_2$	P_1 and P_2 are true
$P_1 \rightarrow P_2$	P_1 implies P_2
Properties	
$\forall t.(Cp_t(X) \wedge Cf_t(X_1,\ldots,X_n) \rightarrow \exists! i \setminus X = X_i)$	The preferred centre is one the forward-looking centres

Table 3: Notation used to formalise the local focus rules

centres. Forward-looking centres are ranked. The most highly ranked centre is called the preferred centre and is supposed usually to become the backward-looking centre of the next sentence, although this is of course not necessary. The ranking of the forward-looking centres depends on the importance of the things introduced.

When a new communication is output, the new backward-looking centre is the most highly ranked element of the current forward-looking centres realised in the new sentence. The new forward-looking centres are the things introduced by the new sentence.

The theory is composed of four rules. These rules and the ordering associated with them constrain how the dialogue may evolve.

The rules are again only indicating what changes in the dialogue state. The notation used to represent the rules and some properties which always hold are shown in table 3.

The interpretation of rule 8 for example is then the following: If the current backward-looking centre is B and if we can select a communication $\mathcal{C}_u$ such that its backward-looking centre with respect to the current dialogue state and its preferred centre are both equal to B then we can perform a continuation focus move. This move has the effect of changing the current forward-looking centres to those of $\mathcal{C}_u$ and the preferred centre to that of $\mathcal{C}_u$.

Rule 8 (Continuation) *The new backward-looking centre is the same as for the previ-*

ous sentence and is also the new preferred centre.

$$\left(\begin{array}{r} Cb_t(B) \wedge \\ Cb_t(C_u, NewB) \wedge \\ Cp_t(C_u, NewP) \wedge \\ Cf_t(C_u, NewF) \wedge \\ B = NewB = NewP \end{array}\right) \rightarrow \left(\begin{array}{l} Cf_{t+1}(NewF) \\ \wedge\, Cp_{t+1}(NewP) \end{array}\right)$$

Rule 9 (Retaining) *The new backward-looking centre is the same as for the previous sentence but it is different from the new preferred centre.*

$$\left(\begin{array}{r} Cb_t(B) \wedge \\ Cb_t(C_u, NewB) \wedge \\ Cp_t(C_u, NewP) \wedge \\ Cf_t(C_u, NewF) \wedge \\ B = NewB \wedge \\ \neg NewB = NewP \end{array}\right) \rightarrow \left(\begin{array}{l} Cf_{t+1}(NewF) \\ \wedge\, Cp_{t+1}(NewP) \end{array}\right)$$

Rule 10 (Smooth-shift) *The new backward-looking centre is not the same as for the previous sentence but is also the same as the new preferred centre.*

$$\left(\begin{array}{r} Cb_t(B) \wedge \\ Cb_t(C_u, NewB) \wedge \\ Cp_t(C_u, NewP) \wedge \\ Cf_t(C_u, NewF) \wedge \\ \neg B = NewB \wedge \\ NewB = NewP \end{array}\right) \rightarrow \left(\begin{array}{l} Cf_{t+1}(NewF) \\ \wedge\, Cp_{t+1}(NewP) \\ \wedge\, Cb_{t+1}(NewB) \end{array}\right)$$

Rule 11 (Rough-shift) *The new backward-looking centre is not the same as for the previous sentence and is different from the new preferred centre.*

$$\left(\begin{array}{r} Cb_t(B) \wedge \\ Cb_t(C_u, NewB) \wedge \\ Cp_t(C_u, NewP) \wedge \\ Cf_t(C_u, NewF) \wedge \\ \neg B = NewB \wedge \\ \neg NewB = NewP \end{array}\right) \rightarrow \left(\begin{array}{l} Cf_{t+1}(NewF) \\ \wedge\, Cp_{t+1}(NewP) \\ \wedge\, Cb_{t+1}(NewB) \end{array}\right)$$

3.2.1 Ordering of the rules

Several rules may have their preconditions satisfied at the same time. We prefer then to apply the rule that maintains the focus if possible, or minimises its movement. Rules are therefore applied in the following preference order:

continuation $>$ smooth-shift $>$ retaining $>$ rough-shift.

This is a variation on the original ordering where retaining is preferred to smooth-shift (Brennan et al. 1987). This change has been done in order to present new things as soon as possible. It means that we first try to find a communication to be output which would allow a “continuation” move. Then, if such a communication does not exist, we would try to find a “smooth-shift” communication and so on until a communication is found.

Preferring a smooth-shift to retaining introduces new things into the dialogue as soon as possible. This is equivalent to McKeown’s rules (McKeown 1985, p. 64).

An example of the use of this theory is given in section 4.

3.2.2 Translation rules

Because the focus rules we presented in section 3.2 are based on generic concepts such as backward or forward looking centres which are not necessarily used by the elicitation module, we need again to bridge the gap between these representations.

In the case of this theory, the translation rules are particularly simple:

- $\forall t.(Cf_t(\mathcal{C}_u, X_1, \ldots, X_n) \wedge X \in X_1, \ldots, X_n \leftrightarrow subject(\mathcal{C}_u, X))$. Everything the communication mentions is a forward-looking center.
- $\forall t.(Cp_t(\mathcal{C}_u, X) \leftrightarrow mainsubject(\mathcal{C}_u, X))$. The main subject of the communication is the preferred centre. The other forward-looking centres are not particularly ordered as they only present background information.

3.2.3 Surface phenomena

The centering theory has been especially created in order to account for the use of pronouns. In our system we make use of this capability to pronominalise some of the entities during the generation of communications and to resolve some anaphora during the parsing of sentences. In particular we allow the backward-looking centre to be realised by a pronoun. This is the case for the example in example 3 where the backward-looking centre of the second sentence is researcher set 7 since it was the first sentence preferred centre.

Example 3

S: Researcher set 7 is involved in research group 1
Do you want *it* represented by a site?

---

4 Example

In this section we describe how our system selects moves based on the focus theories and the translation rules presented. A particular point of interest is the interaction between the dialogue manager and the elicitation system. The elicitation system influences the dialogue manager by adding communications to or removing communications from the communication pool. The dialogue manager influences the elicitation system by deciding at which point in the dialogue the communications are output. This is different from systems where the natural language interface is used as an input or output system but does not directly influence the way of working of the system.

Readers who are not interested in a detailed presentation of the system at work can safely skip this section.

We now present how the dialogue manager and elicitation system interact to produce example 4 and show how the specification and the focus spaces are modified[2]. (The complete dialogue is given in appendix C.) The two theories presented above are used to rank the communications. The global theory proposes an order which is refined

[2]This is a much simplified account of what really happens since our system deals with more entities and relations. Moreover, the mixed-initiative capability of the system is not shown. Nevertheless, it presents important mechanisms taking place.

Example 4

S1:	Research group 1 is a research group
	Do you want a site presenting it?
U1:	Yes
S2:	A site presents research group 1
	Which site presents research group 1 ?
U2:	A new site
S3:	Site 22 presents research group 1
	Do you want a home page describing site 22?
U3:	Yes
S4:	A home page describes site 22
	Which home page describes site 22?
U4:	A new home page
S5:	Home page 24 describes site 22
	What is the title of home page 24?
U5:	Research group
S6:	Do you want a page linked from it?
U6:	Yes
S7:	A page is linked from research group home page 24
	Which page is linked from research group home page 24?
U7:	A new page

by the local theory. The global theory has therefore precedence over the local theory in the ranking process.

S1 (Research group 1 is a research group) The elicitation module starts by creating a research group. This causes a presentation communication to be added to the communication pool in order to keep the user aware of the specification modification. This communication is then selected and presented.

Specification
entity(rg1, research group)

Communication pool
presentation(rg1)

Focus set
$\mathcal{F}_{1,0} = \emptyset$

Centres
$Cf_0()$ $Cp_0()$ $Cb_0()$

Since this is the only communication, it is selected and output.

S1 (Do you want a site presenting it?) Then the elicitation module knows that a research group can be presented by a site or a page. It therefore creates two hypothetical relations (shown by question marks) and two as-yet-unidentified entities (also shown by question marks). It then creates two question communications, `question(p1 | rg1)` and `question(p2 | rg1)`, to know which relation the user really wants. Since it also knows that a research group involves a set of

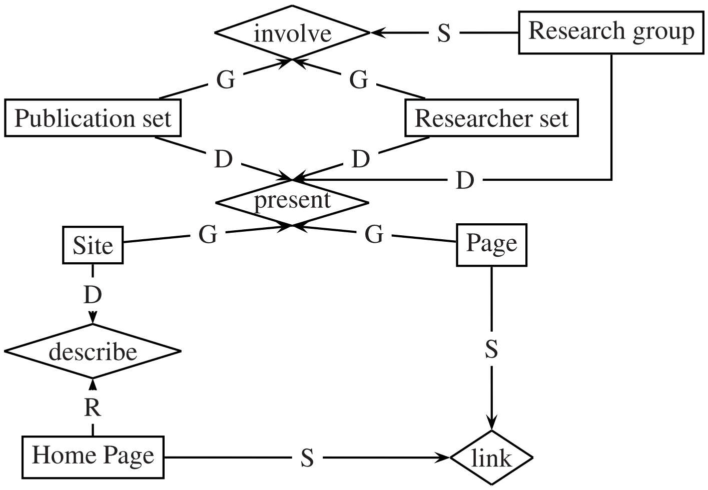

Figure 5: Direct, Specialisation, Generalisation and other relations

The figure represents the by-default direct (D), specialisation (S), generalisation (G) and other (R) relations from entities to relations (arrow direction). There is always a direct relation from relations to entities (opposite to arrow direction; translation rule D4). The relations may change as the dialogue evolves, depending on what is in focus (translation rule D2).

researchers (among other things, see table 9), it creates such a set and puts a presentation communication, `presentation(i1 | rg1, rs1)`, in the communication pool to present it. (Another set is created for the publications but we do not consider it here.)

Specification
entity(rg1, research group)
entity(rs1, researcher set)
? entity(?s1, site)
? entity(?pa1, page)
relation(i1, involve)
? relation(p1, present)
? relation(p2,present)
rolefillervalue(i1, involving, rs1)
rolefillervalue(i1, involved in, rg1)
? rolefillervalue(p1, presenting, rg1)
? rolefillervalue(p1, presented by, ?s1)
? rolefillervalue(p2, presenting, rg1)
? rolefillervalue(p2, presented by, ?pa1)

Communication pool
presentation(i1 | rg1, rs1)
question(p1 | rg1)
question(p2 | rg1)

Focus set
$\mathcal{F}_{1,1} = \{rg1\}$

Centres
$Cf_1(rg1)$
$Cp_1(rg1)$
$Cb_1(rg1)$

A specialisation relation exists between `rg1` and `i1` (translation rule S1, T_r = involve, $Role_1$ = involved in, $Role_2$ = involving, rolecardmax(involve, involving,2)) since `rg1` also involves a publication set. (The relations have been summarised in figure 5.) Since site and page are optional, dependent relations on the research group (the research group can be not presented at all but the site and the page must present something, see section 2), direct relations exist between `rg1` and `p1` and `p2` (translation rule D3, T_r = present, $Role_1$ = presented by , $Role_2$ = presenting,). Therefore a no change move can be made and one of the question communications is selected. (The presentation of `i1` involves an additive global focus move and is therefore delayed.) At the same time, both communications are associated with a retaining local move. As far as these two communications are concerned, the focus theories used here do not discriminate between them. The selection is therefore random. In this case, the question about creating the site is selected and output. The research group is realised by a pronoun, being the backward-looking centre.

S2 (A site presents research group 1) Since the user answers positively, the corresponding relation is asserted in the specification and the other question is removed from the communication pool as it would lead to a violation of the constraints of the ER model (see section 2). This shows that the dialogue manager cannot plan reliably the dialogue ahead. The elicitation system then puts a presentation communication in the communication pool indicating that a relation has just been created. It also puts a question communication in the communication pool asking which site is presenting the research group.

Specification
entity(rg1, research group) entity(rs1, researcher set) entity(?s1, site) relation(i1, involve) relation(p1, present) rolefillervalue(i1, involving, rs1) rolefillervalue(i1, involved in, rg1) rolefillervalue(p1, presenting, rg1) rolefillervalue(p1, presented by, ?s1)

Communication pool
presentation(i1 | rs1, rg1) presentation(p1 | rg1) question(?s1 | p1, presented by)

Focus set
$\mathcal{F}_{1,2} = \{rg1, p1\}$

Centers
$Cf_2(p1, rg1)$ $Cp_2(p1)$ $Cb_2(rg1)$

A direct relation exists for these two communications with `rg1` (since `p1` is in direct relation with itself and because of translation rule D4, Role = presented by, respectively). Moreover they are both associated with a smooth-shift local move. Since presentation communications are by-default preferred over question communications, the presentation communication is selected and output. The presentation of `i1` is still delayed since it involves an additive move.

Specification
entity(rg1, research group) entity(rs1, researcher set) entity(?s1, site) relation(i1, involve) relation(p1, present) rolefillervalue(i1, involving, rs1) rolefillervalue(i1, involved in, rg1) rolefillervalue(p1, presenting, rg1) rolefillervalue(p1, presented by, ?s1)

Communication pool
presentation(i1 | rs1, rg1) question(?s1 | p1, presented by)

Focus set
$\mathcal{F}_{1,3} = \{rg1, p1\}$

Centres
$Cf_3(p1, rg1)$ $Cp_3(p1)$ $Cb_3(p1)$

S2 (Which site presents research group 1?) The question communication is then itself output with a no change, continuation move.

Specification
entity(rg1, research group)
entity(rs1, researcher set)
entity(?s1, site)
relation(p1, present)
relation(i1, involve)
rolefillervalue(i1, involving, rs1)
rolefillervalue(i1, involved in, rg1)
rolefillervalue(p1, presenting, rg1)
rolefillervalue(p1, presented by, ?s1)

Communication pool
presentation(i1 | rs1, rg1)

Focus set
$\mathcal{F}_{1,4} = \{rg1, p1\}$

Centres
$Cf_4(p1, rg1)$
$Cp_4(p1)$
$Cb_4(p1)$

Given the user's answer, a new site is then created and added to the specification. A presentation communication about this new site, `presentation(p1 | s1)`, is added to the communication pool. A new question communication, `question(d1 | rg1)`, is also added to the communication pool about the possibility of describing the site with a home page (see section 2).

S3 (Site 22 presents research group 1) Direct relations exist for the two new communications (since `p1` is in direct relation with itself, and `d1` is in direct relation with `s1` by rule D3, T_r = describe, $Role_1$ = describing, $Role_2$ = described by, rolecardmin(describe, describing, 1)). They can both be associated with no change global focus move. However, the presentation communication is associated with a continuation local focus move, since it is about `p1`, while the question communication is associated with a rough-shift. The presentation communication is therefore selected with a no change, continuation move and then output. The presentation of `i1` is still delayed since it involves an additive move.

Specification
entity(rg1, research group)
entity(rs1, researcher set)
entity(s1, site)
relation(i1, involve)
relation(p1, present)
rolefillervalue(i1, involving, rs1)
rolefillervalue(i1, involved in, rg1)
rolefillervalue(p1, presenting, rg1)
rolefillervalue(p1, presented by, s1)

Communication pool
presentation(i1 | rs1, rg1)
question(d1 | rg1)

Focus set
$\mathcal{F}_{1,5} = \{rg1, p1, s1\}$

Centres
$Cf_5(p1, rg1, s1)$
$Cp_5(p1)$
$Cb_5(p1)$

S3 (Do you want a home page describing site 22?) The question communication is then selected with a no change, rough shift move and output. The presentation of `i1` is again delayed.

Specification
entity(rg1, research group)
entity(rs1, researcher set)
entity(s1, site)
?entity(?hp1, home page)
relation(i1, involve)
relation(p1, present)
?relation(describe, d1)
rolefillervalue(i1, involving, rs1)
rolefillervalue(i1, involved in, rg1)
rolefillervalue(p1, presenting, rg1)
rolefillervalue(p1, presented by, s1)
?rolefillervalue(d1, describing, s1)
?rolefillervalue(d1, described by, ?hp1)

Communication pool
presentation(i1 $\mid$ rs1, rg1)

Focus set
$\mathcal{F}_{1,6} = \{rg1, p1, s1, d1\}$

Centres
$Cf_6(d1, rg1)$ $Cp_6(d1)$ $Cb_6(d1)$

Given the answer of the user, the hypothetical relation is asserted.

S4 The same process happens for the home page which is an optional, dependent entity of the site.

S5 (What is the title of home page 24?) Since home pages have titles, the elicitation system puts a question communication in the communication pool about the title. At the same time it creates a hypothetical link with a page, since a home page can be linked to other pages. It also puts a question communication about the validity of this link.

Specification
entity(rg1, research group)
entity(rs1, researcher set)
entity(s1, site)
entity(hp1, home page)
? entity(?pa1, page)
relation(i1, involve)
relation(p1, present)
relation(d1, describe)
? relation(l1, link)
rolefillervalue(i1, involving, rs1)
rolefillervalue(i1, involved in, rg1)
rolefillervalue(p1, presenting, rg1)
rolefillervalue(p1, presented by, s1)
rolefillervalue(d1, describing, s1)
rolefillervalue(d1, described by, hp1)
? rolefillervalue(l1, linked from, hp1)
? rolefillervalue(l1, linked to, pa1)

Communication pool
presentation(i1 $\mid$ rs1, rg1) question(l1 $\mid$ hp1) question(hp1, title)

Focus set
$\mathcal{F}_{1,9} = \{rg1, p1, s1, d1, hp1\}$

Centres
$Cf_9(d1, s1, hp1)$ $Cp_9(d1)$ $Cb_9(d1)$

The title is in a direct relation with the home page (translation rule D1) whereas the link is in a specialisation relation (since there can be several linked pages, translation rule S1 applies, T_r = link, $Role_1$ = linked from, $Role_2$ = linked to, rolecardmax(link, linked to, ∞)). Because direct relations are preferred (no change moves are ranked higher than additive moves), the question about the title is selected and output.

Specification
entity(rg1, research group)
entity(rs1, researcher set)
entity(s1, site)
entity(hp1, home page)
? entity(?pa1, page)
relation(i1, involve)
relation(p1, present)
relation(d1, describe)
? relation(l1, link)
rolefillervalue(i1, involving, rs1)
rolefillervalue(i1, involved in, rg1)
rolefillervalue(p1, presenting, rg1)
rolefillervalue(p1, presented by, s1)
rolefillervalue(d1, describing, s1)
rolefillervalue(d1, described by, hp1)
? rolefillervalue(l1, linked from, hp1)
? rolefillervalue(l1, linked to, ?pa1)

Communication pool
presentation(i1 | rs1, rg1)
question(l1 | hp1)

Focus set
$\mathcal{F}_{1,10} = \{rg1, p1, s1, d1, hp1\}$

Centres
$Cf_{10}(hp1)$
$Cp_{10}(hp1)$
$Cb_{10}(hp1)$

S6 (Do you want a page linked from it?), S7 The question about an eventual link is then in competition with the presentation of `i1` since both involve an additive move. However, the question about the link is associated with a retaining shift move whereas the presentation of `i1` is associated with a rough shift move. Therefore the question communication is preferred and output. A new focus space is therefore created and all the questions related to the new page will be asked in this space.

Specification
entity(rg1, research group)
entity(rs1, researcher set)
entity(s1, site)
entity(hp1, home page)
entity(?pa1, page)
relation(i1, involve)
relation(p1, present)
relation(p1, present)
relation(l1, link)
relation(d1, describe)
rolefillervalue(i1, involving, rs1)
rolefillervalue(i1, involved in, rg1)
rolefillervalue(p1, presenting, rg1)
rolefillervalue(p1, presented by, s1)
rolefillervalue(d1, describing, s1)
rolefillervalue(d1, described by, hp1)
rolefillervalue(l1, linked from, hp1)
rolefillervalue(l1, linked to, ?pa1)
attributevalue(hp1, title, research group)

Communication pool
presentation(i1 | rs1, rg1)

Focus set
$\mathcal{F}_{1,11} = \{rg1, p1, s1, d1, hp1\}$
$\mathcal{F}_{2,11} = \{l1, p1\}$
controlling($\mathcal{F}_{1,11}$, $\mathcal{F}_{2,11}$)

Centres
$Cf_{11}(l1, hp1)$
$Cp_{11}(l1)$
$Cb_{11}(hp1)$

As can be seen from this example, the dialogue depends on the communication selection made by the dialogue manager. However this selection is constrained by the availability of communications in the communication pool. If the elicitation module follows a strict and narrow way of working, always producing a unique communication, then the dialogue manager becomes irrelevant. On the other hand, putting several communications in the communication pool entails that their treatment can be deferred for an unknown amount of time, until the dialogue manager decides to output them. As a result, the elicitation system should balance the advantages and drawbacks of putting too few or too many communications in the communication pool, and the dialogue manager should balance the advantages and drawbacks of more or less delaying the output of communications.

Global and local focus rules both constrain the evolution of the dialogue. In the previous dialogue for example, global focus delays the presentation of the researcher set while the research group is in focus. Then, when two additive moves are possible, one presenting the researcher set and the other a link to the home page, local focus constrains the dialogue to deal with the link which is more related to the then current topic.

5 Evaluation

Numerous pieces of work make use of focus theories. The theories are sometimes modified to achieve specific aims (e.g, Mittal et al. (1998); Huang (1994)). However, most of these approaches do not precisely evaluate what the contributions of the theories are. The fact that they provide improvements in the dialogue quality is assumed. In this section we present an evaluation of the impact of using focus theories. This evaluation was made during the development of the system, helping decide to which aspect of the system efforts should be devoted. We describe what the contributions of the different theories are, with an emphasis on the theories presented in section 3.

The evaluation strategy we followed can be formalised. We present its general form in section 5.2.

5.1 Experiments

We first describe the setting in which the evaluation has been done and the resulting constraints that applied to the evaluation (Sparck Jones and Galliers 1996).

In our case, we wanted to see how the focus theories influence the capability of the system to produce coherent dialogue. As shown in figure 1, the focus theories intervene through the dialogue manager by selecting a communication from the communication pool and directing it to the Natural Language Interface. Since we want to test the theories in the same conditions, we chose to use the same elicitation module and the same NLI for all tests. The elicitation system is a loose one, i.e., putting as many communications in the communication pool as possible (see section 2). It offers many choices for selection. The NLI is based on templates. These are modified to allow changes such as the use of pronouns. The resulting experiments can be seen as a black-box experiment where the influence of the theories are evaluated through the behaviour of the whole system.

We consider that users of the system are accustomed to using computers but not this particular elicitation system. Users who had no linguistics knowledge were given short explanations on the evaluation they had to perform in order to ensure they understood exactly what was asked in the experiments.

	No management[3]	One focus theory	Two focus theories
Evaluations	2, 3, 5	15, 10, 6	10, 12, 12, 10, 8

Table 4: First evaluation
Evaluations (vary from 0 (very bad) to 16 (excellent))

Using a focus theory is compared to having no management, using other theories and using other management methods (not focus based). It is not useful to implement complex focus theories if they do not provide improvements over simpler methods. Moreover, the users are not particularly interested in linguistic phenomena. Therefore, the use of the theories should make a real change so that users can feel the difference from other dialogue management strategies.

5.1.1 First experiment

The aim of the first evaluation was to determine if focus theories have an influence on dialogue quality. In order to evaluate this, we defined a criterion of dialogue quality based on dialogue properties such as ease of understanding and well-structuredness. The experiment consisted in having users use our system with and without dialogue focus rules and then filling in a questionnaire about the dialogue they had had. The criterion was then calculated by summing up the answers in the questionnaire, each question (except the first and the last which are not evaluation questions) counting from 0 up to 4 depending on the answer (the higher the user's satisfaction, the higher the mark). The questionnaire is presented in appendix D.

Eight users were asked to enter a given specification. Users had mostly to answer questions but could also "pass" them, i.e., decline to answer them at that point. They could also ask to see the current state of the specification. Other mixed-initiative possibilities were limited.

Seven of the users had never used the system before. We obtained eleven dialogues (some users producing more than one dialogue) and their evaluation (see table 4). The experiment lasted less than two hours, including a presentation of the system to the new users. We found that there was a statistically significant difference between the evaluation of a random dialogue, i.e., a dialogue where communications are picked at random in the communication pool, and the evaluation of a dialogue using focus rules (Student test with null-hypothesis 'no difference in evaluation means', $t = 4.15$; degree of freedom, $df = 9$; probability of error in rejecting null-hypothesis, $p = 0.002$).

Although this experiment let us know that focus rules may be useful, it does not show that they perform better than other dialogue management strategies such as following the elicitation module order. Testing these possibilities would require a heavy investment in users' evaluation of the system. In order to reduce this investment while being able to test other dialogue management strategies, we designed an automated evaluation procedure.

5.1.2 Automation

The aim of the automation is to replace the real users with automated users while obtaining an evaluation that remains faithful to the real users' perception of the dialogues.

[3]Communications were selected at random from the communication pool.

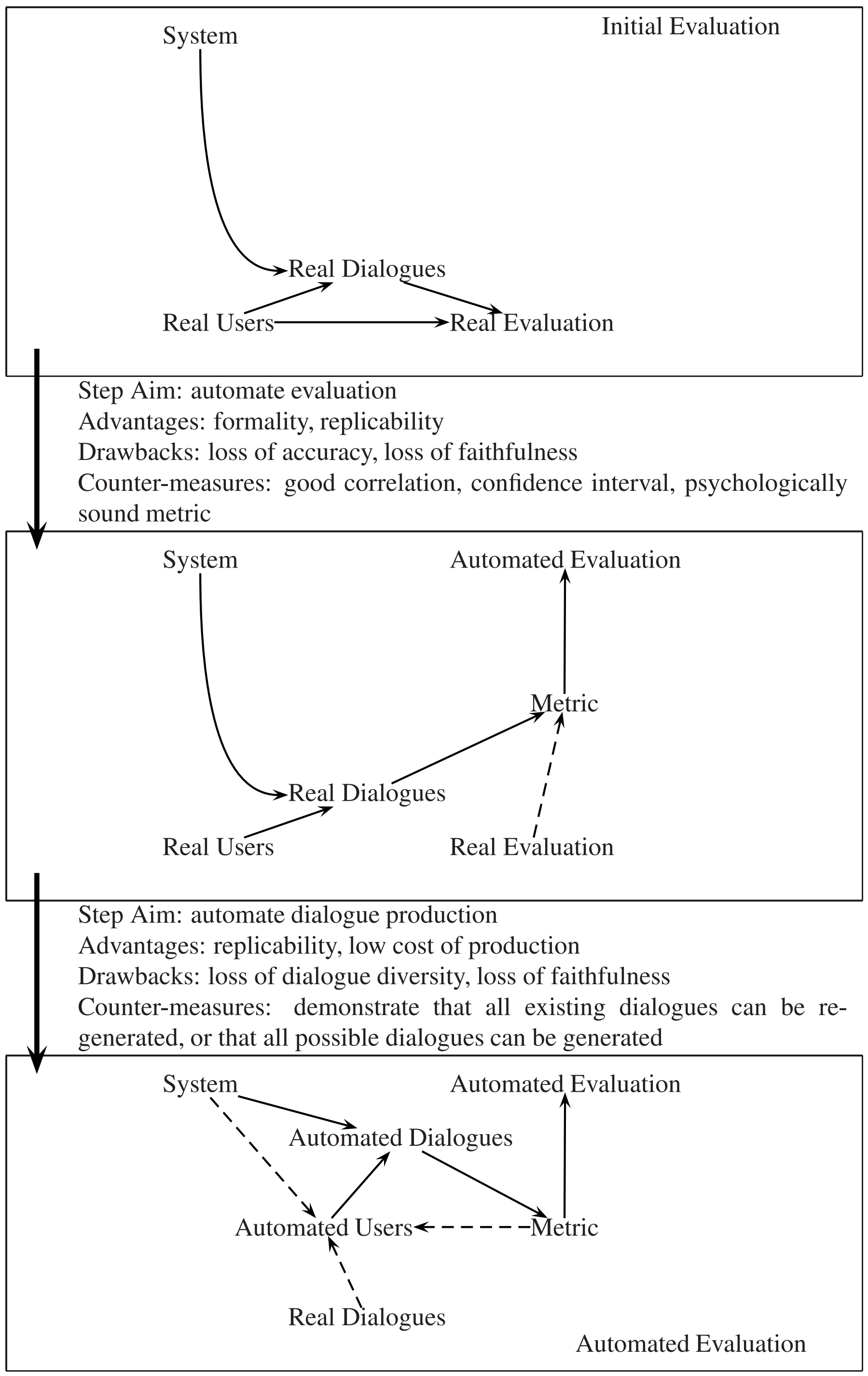

Figure 6: Automated evaluation

Solid lines show how the evaluation is carried out. Dashed lines show how automated components are obtained.

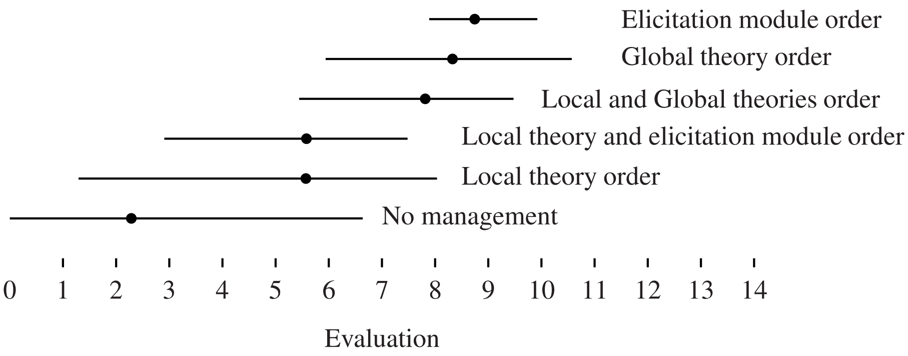

Figure 7: Evaluations (vary from 0 (very bad) to 16 (excellent)) and 90% confidence intervals

The automation of the evaluation is composed of two steps as shown in figure 6. Each step consists in automating a part of the evaluation process. Each step provides gains but involves possible losses. Counter-measures are taken to limit these losses.

The first step consists in finding a metric based on easily measurable dialogue properties which is correlated with the evaluation given by users. The idea is that the metric should allow an easy and clear separation between the strategies which provide improvements and those that do not. Using this metric we are able to evaluate automatically the quality of a given dialogue. Possible problems are a loss in accuracy and in faithfulness to the initial evaluation. To counterbalance this loss we searched for a metric well-correlated with the evaluation and psychologically motivated. Several metrics often proposed in the literature as being related to dialogue quality were recorded during the users' evaluation. We found that some obvious choices were not satisfactory in our case. For example, the length of the dialogue was not well correlated to the evaluation (r = 0.4019). The number of passed communications (r = -0.1207) and the number of times users asked to see the specification (r = -0.3595) were not good metrics either.

We found that one of the metrics accounted for 65% of the evaluation mark (correlation coefficient, r = 0.8419; error on r, $\sigma_r = 0.2$; Student test with null-hypothesis 'r = 0', t = 4.08; degree of freedom, df = 9; probability of error in rejecting null-hypothesis, p = 0.003). This metric is based on dialogue coherence: communications related to the previous ones increment the metric (with a bonus for those related to the immediately preceding sentence)[4]. The fact that this metric is based on dialogue coherence encourages us to investigate the role of focus theories more precisely. Metrics with higher correlation coefficient were found. However, they were either not statistically significant or were not psychologically sound, i.e., their results could not be easily explained.

The second step consists in producing dialogues automatically with different ver-

[4]More precisely, a communication was considered related to the previous ones if (1) it was a presentation communication or (2) its main subject was among the last seven subjects spoken about. A bonus was accorded if the communication main subject was among the last two subjects spoken about.

sions of the system and comparing them using the metric. This is done by writing automated users. The aim of these programs is to approximate the dialogues of real users well enough that the metric will produce useful evaluations. The metric divides dialogues into equivalence classes (dialogues having the same metric value at a given confidence level). Therefore, these users do not need to imitate all aspects of the real users but only produce dialogues of the same equivalence class as real users. More accurate automated users would not always provide much more accurate evaluations since the metric would often level the results off. Therefore we wrote simple and cost-effective automated users. These users know about a target specification and answer the questions asked by the system about it. In order to do this, they map the entities mentioned in the dialogue with their counterparts in the specification. The mapping is essentially done when presentation communications are processed. For example, if a research group is mentioned in the dialogue and a unique research group exists in the target specification, the two entities will be mapped. This means that the identifier of the research group in the dialogue, e.g., 1, is associated with the identifier of the research group in the automated user's internal specification. This association is stored in a "Mappings" list. Questions can then be answered by mapping the elements of the questions to their internal counterparts and then searching for an answer in the target specification. If presentation communications cannot be processed, e.g., because the system has not already presented things needed to interpret them, they are stored for later processing. If questions cannot be answered, they are passed. A possible problem with the automation is the loss of dialogue diversity due to simple automated users. We ensured this was not the case by verifying our automated users could generate dialogues equivalent in length and structure to the ones produced by real users. (Because our domain of discourse is finite due to the generator and parser limitations, we could also have shown that our users are able to produce all possible dialogues.) In order to achieve this we had to limit the memory span of our automated users so that they would forget things as human do. These users keep a list of things in focus. The list has a limited storage capacity. New things put in the list may therefore displace those already in the list, on a First-In-First-Out (FIFO) basis. If a question is about things that are not in the list, the automated users will answer or pass the question with some probability. The algorithm driving the automated users is described in table 5.

Finally we can run the system with the automated users and obtain an evaluation of new dialogue management strategies. (A sample of a dialogue generated automatically is given in appendix C.) The results given by this approach are not very accurate since the use of a metric and automated users introduces approximations. However, these losses in accuracy can be partially evaluated and counterbalanced. For example, by considering confidence intervals rather than point evaluations, we can get a better opinion on the real differences between strategies. Confidence intervals can be computed from the uncertainty on the automated users' evaluation (due to the random element) and the uncertainty on the metric (due to the imperfect correlation with a real evaluation).

Two main assumptions are made during the automation process:

1. A well-correlated and psychologically sound metric can be found. Numerous metrics have been studied and evaluated in various experiments. However, this may still be difficult for dialogue aspects on which little research has been carried out.

2. Automated users imitating real users can be written. This is usually possible since the system under study can only accept a limited number of answers. How-

```
Unprocessed Mappings ← ∅ {contains the presentation communications that could
not be processed immediately}
Mappings ← ∅ {contains the mapping between dialogue entities and their internal
counterparts}
Focus list ← ∅ {contains the things spoken about recently and acts as a FIFO}
while there is a communication to process coming from the dialogue manager do
  if this is a presentation communication then
    if it can be interpreted using Mappings then
      add any new mappings to Mappings
      add its subjects to the focus list
      repeat
        select a communication from Unprocessed Mappings
        if it can be interpreted using Mappings then
          add any new mappings to Mappings
          remove the communication from Unprocessed Mappings
        end if
      until no communication can be removed
    else
      add the communication to Unprocessed Mappings
    end if
  else
    if it can be interpreted using Mappings then
      if its subjects are in the focus list then
        answer the question
        add any new mappings to Mappings
        add its subjects to the focus list
      else
        if a random number is greater than a given threshold then {this is where a
        random element is introduced}
          answer the question
          add any new mappings to Mappings
          add the subjects to the focus list
        else
          pass the communication
        end if
      end if
    else
      pass the communication
    end if
  end if
end while
```

Table 5: Automated users' algorithm

ever, this may still be difficult to achieve if for example the users need a lot of common-sense knowledge.

Not all new dialogue management strategies can be tested in this way. The use of a metric puts constraints on the type of dialogues that can be evaluated. For example, the length of the dialogue was shown not to be a good metric in our case. However this metric could be good for other dialogues, e.g., if the dialogue manager's main role was to remove communications rather than ordering them. If the type of dialogues evaluated changes, a new metric should be used. The same pertains to automated users. These users have been written to have a dialogue with a particular type of dialogue manager. If the latter changes too much, automated users may not be able to deal adequately with it. It is therefore important to compare dialogue managers that are not "radically" different.

Using the metric and the automated users, we were able to test if other dialogue management strategies that were not tested during the first experiment could also provide improvements in dialogue quality. One such strategy was to simply follow the elicitation module order (in effect by-passing the dialogue manager and considering the communication pool as a FIFO store). This strategy turns out to perform slightly better than focus theories. This surprising result comes from the way our elicitation module works. By always outputting all the questions it can, it effectively implements a strategy that has some resemblance with the global and local theories mixed together (although the jump between global focus space is more random and local coherence is not strictly enforced). This would not be true of elicitation system in general, which by default do not manage dialogues. We also tested the theories presented in section 3 separately. The evaluations are shown in figure 7. (The theory presented in section 3.1 is called "global" in the figure and the theory presented in section 3.2 is called "local".) The local focus theory, although better than the no management policy, performs less well than the global or elicitation order strategies. This could be due to the fact that a local dialogue management leads to "spaghetti" dialogues (Sibun 1992) which are not optimal in our case. Moreover combining the local strategy with the elicitation order does not provide any improvements. Since the elicitation module order is not formalised, it is difficult to predict how it will interact with other focus theories. In this case, the "spaghetti" nature of the local theory seems more important than the structuring effect of the elicitation module.

Based on these results, three strategies were selected for further evaluation: elicitation module order, global focus theory and the mix of global and local focus theories. The strategies were selected because they form the best cluster in the evaluation and we have found reasons to reject the other strategies.

5.1.3 Second experiment

In the following evaluation, we evaluated the three selected theories without developing new ones. (The local theory has been dismissed by the previous experiment but is added as a base case.) We did not search for a correlated metric (although we recorded metric values in case of further development) and we did not develop automated users.

We tested if the strategies were ensuring good global coherence. We tested this hypothesis by asking people to read transcripts of dialogues generated with and without dialogue strategies and asking them to compare the dialogues pair-wise based on their coherence.

The experiment involved six persons for 90 minutes. (The experiment sheet is presented in appendix E.) Six dialogues were evaluated, resulting in 36 pair-wise comparisons. (Each participant compared four dialogues.)

Pair-wise comparisons can be transformed in an overall ranking of the alternative dialogues (Saaty 1990). This is done by computing the principal eigen vector of the comparison grid. Its coordinates give the relative importance of the different strategies.

Local and global theories order	33
Global theory order	47
Elicitation system order	12
Local theory order	*8*
	100

Table 6: Theory ranking by global coherence

Local and global theories order	50
Global theory order	37
Elicitation system order	*13*
	100

Table 7: Theory ranking by local coherence

For example, a theory with a value of 60 is considered three times more globally coherent than a theory with a value of 20.

The results are presented in table 6 (consistency ratio = 0.0818[5]). They clearly shows that following the elicitation system orders performs poorly compared to using a focus theory (3 to 4 times worse than a focus theory). This strategy is therefore abandoned. The results also confirms that local theory is performing badly as was predicted in the previous stage. The local and global theories are performing less well than the global theory alone because the local theory smooths the transitions between focus spaces. Therefore the dialogue seems less well divided than with the global theory only.

5.1.4 Third experiment

Finally we tested the two remaining theories' capability to improve local coherence. We tested this hypothesis with the same experiment as in the previous evaluation. The elicitation system order which has been dismissed in the previous experiment is added as a base case.

The results are presented in table 7 (the elicitation system order was added as a base case, consistency ratio = 0.1664). The theory mix consisting of a local and a global theories is performing better than the global theory alone. However the difference is not big. This can be explained by the small size of the focus spaces. Since just a few communications are output in each space (around 4 on average), a suboptimal coherence strategy does not decrease the perceived local coherence too much.

5.1.5 Summary

From this analysis we can conclude that, in our domain, the local and global theory mix and the global theory on its own provide definitive improvements in dialogue quality and especially in dialogue coherence. The global theory performs better on global coherence as can be expected and the theory mix performs well on local aspects. It should be noted that the local theory which seemed a potential strategy was eliminated quickly in the development process. The elicitation order which was not at first considered has

[5] A consistency ratio less than 0.1 indicates reliable results.

been introduced at a very low cost in the evaluation process but was later dismissed for its poor performance in global and local coherence. All the experiments needed to develop and evaluate the system took only 25 person-hours on aggregate.

This analysis was performed on a limited domain. However, the evaluation was not dependent on the domain. We therefore believe that the results have a wider application than the requirements engineering domain.

5.2 Development

In this section we show how we integrated the automation of the evaluation we used to get the results presented in the previous section into a general development method. We used this development method to produce the theories and results presented above. This method could be the basis for a new and generally useful development approach for interactive research systems.

Because there is not yet a clear method for the development of dialogue managers, an iterative approach using evolutionary prototypes seems adequate (Partridge 1992; Sommerville 1995). Each iteration is a step towards the final program. However, not going astray is a major difficulty. Evaluating whether a dialogue management strategy is beneficial or not, in terms of performance but also in terms of cost-effectiveness, is therefore important.

Evaluation of dialogue management strategies is a difficult task. Several sources can influence the dialogue beside the management techniques, and dialogue can be evaluated along several dimensions. A lot of current approaches to evaluation require heavy involvement from users (Sikorski and Allen 1996; Walker et al. 1997a). This makes them difficult to use during development. Some are based on automated users:

- Walker (1996) uses automated users whose dialogue strategies and memory and reasoning capabilities can be adjusted. Communication and reasoning have an associated cost. Dialogue performance is then calculated by considering the overall cost of the dialogue and how successful it was. This framework is useful for studying the theoretical implications of dialogue strategies depending on cognitive constraints (Hanks et al. 1993). However, it is unclear what the relation is between the dialogue performance computed and a real dialogue evaluation. (For example, counting the numbers of utterances may be less representative of the dialogue quality than of the elicitation system's capability to guess some of the specification.) Therefore the application of this framework to dialogue quality evaluation is limited.

- Eckert et al. (1997, 1998) also uses automated users whose dialogue strategies can be adjusted. The adjustments are made by studying a corpus of existing dialogues. Dialogue quality is then calculating by using a metric which is chosen to be well correlated with real users' evaluation (Walker et al. 1997b). This framework is useful to test different dialogue strategies on a particular aspect of the dialogue and for specific users. However, since only one metric is used, it is not clear how to test the dialogues on different aspects, such as global and local coherence. It is also unclear if investing a lot of effort in accurate automated users is worthwhile when the dialogue manager is undergoing modification which is the case during development. New automated users may be needed if the dialogue manager evolves too much. It may therefore be better to write simpler automated users adapted to the metric used to test a particular aspect of the dialogues. The application of this framework during the development phase where

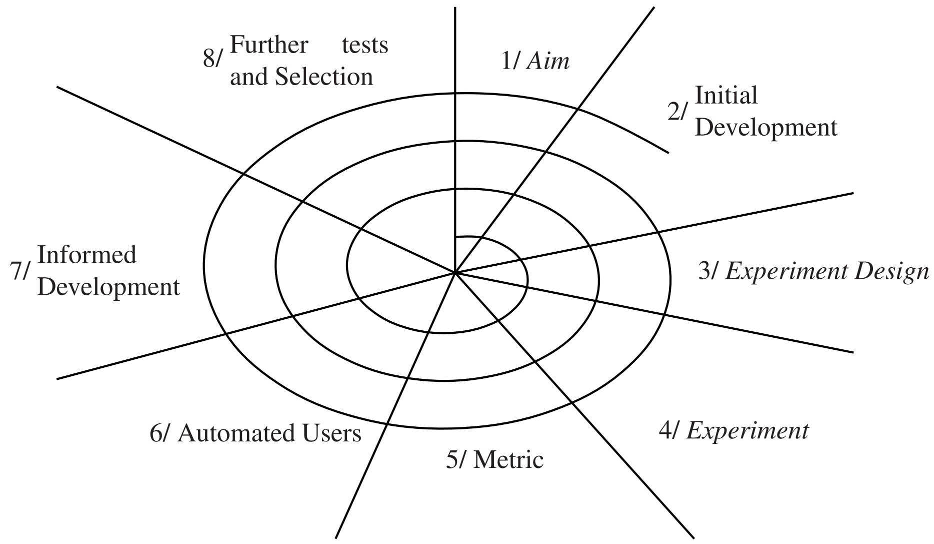

Figure 8: Development and Evaluation Cycle

the dialogue manager evolves and several aspects of the dialogues needs to be evaluated seems therefore limited.

- Ishizaki (1997) also makes use of automated users to evaluate the conditions under which mixed-initiative dialogues are more effective than non mixed-initiative dialogues. In this case, a simple metric, consisting of the number of characters of the dialogue, is used. The automated users are also simple and only roughly based on existing dialogues. However a mathematical model gives a sound theoretical basis to the study and the simulation is used to confirm the theoretical findings. Unfortunately such an approach is difficult to apply in our case since the notion of coherence is difficult to formalise and to study mathematically.

The main objective in our approach is to provide quick but faithful evaluations on particular aspects of the system. This is done by a light-weight evaluation process. Compared to the last two approaches presented above, we make less assumptions about the possibility of studying the dialogues statistically or analytically. This enables us to deal with more types of dialogues. On the other hand, we are not able to benefit from proven mathematical results which means we have to ensure the correctness of our evaluation by repeating experiments. The process can be used while developing the system and can provide guidance on which aspect of the system should be worked on. The process is based on a spiral model of software development (Boehm 1987) which is well suited to our case. It is composed of a cycle shown in figure 8 and described below. Each turn of the cycle represents the development and evaluation of one aspect of the system. The radial movement therefore represents the amount of testing carried out on the system.

1. The first step consists in defining the aim of the current cycle turn. This step has one of two results:

- The cycle is stopped. The system has reached its objective or there is no cost-effective way of making progress.
- A possibility of improvement is identified. Knowing which improvements should be attempted and tested first depends on the final requirements for the system and on the perceived risk in achieving some progress. More important and more risky aspects of the system should be considered first. The constraints under which the improvement should be obtained and a testable criterion are defined. The improvement to be achieved is expressed in terms of that criterion. The aim of this criterion is to allow the partition of the possible additions to the system into those that improve the performance of the system in the way defined in the preceding step and those that do not. (We did this by defining a criterion for dialogue quality.) This ensures that the improvement can be measured and verified (Sommerville and Sawyer 1997, good practice 8.7).

  Application: In our case, in the first turn of the cycle we considered whether dialogue management was providing improvements in dialogue quality. In the second turn, we studied the effect of management on global coherence and in the third we studied the the effect of management on local coherence. Finally we stopped in the fourth turn.

2. Next, development is carried out to provide the improvement wanted. Several competing strategies can be used to achieve this goal.

   Application: In our case, we developed several strategies, such as those based on local and global focus theories.

3. During this step, we design the evaluation that will be carried out in this turn of the cycle. Users' evaluation is necessary since there is usually no other way of knowing if a system is performing correctly: " in general, the straightforward notion of decidably correct or incorrect system behaviour even for specific inputs is difficult to pin down for many AI problems" (Partridge 1992, p30–31). We found that two kinds of experiments are especially useful for evaluation:

   - Interactive experiments where users play with the system and then answer a questionnaire. In general, this is a cost effective way of obtaining subjective data (Macleod 1992, p.26-27). This is all the more the case here since we test only a small aspect of the system defined in step 1.
   - Static experiments where users do not directly interact with the system but judge some of its outputs. Comparison techniques (e.g., pair-wise comparison (Karlsson 1996)) can be used, usually reducing the amount of evaluation to be done.

     Application: In our case, we chose an interactive experiment in the first turn of the cycle. We then chose static ones for the second and third turns.

4. The next step consists in the actual evaluation of the system by users. This is done by asking users to complete the experiment designed in step 3. At the same time, dialogues are recorded and metrics are measured. This step provides two results:

   - It enables us to know if there is a need for further evaluation. If no improvements can be gained then there is no need to go further.

- Assuming that there are improvements, we can automate the evaluation. This is the role of steps 5 and 6.

Application: In our case, we found that some management strategies had an effect on the aspect we were studying.

5. A major problem in obtaining a cost-effective development method is the need to automate the evaluation step because evaluation by users is time-consuming. As explained in section 5.1.2, if we can find a metric correlated with the results found in step 4, then we can use this metric in place of the questionnaire given to the users. The idea is that the metric should allow an easy and clear separation between the strategies which provide improvements and those that do not. Finding this metric is a difficult part of the process. Metrics should be kept as simple as possible especially during the first cycles of the approach. Metrics should be based on dialogue properties and not on properties of the particular strategies used to generate the dialogues since the comparison of the internal working of different theories could be difficult or even impossible. Note that the metrics need not explain the results but just be correlated with them. (In our case, the search for a metric was based on proposals made in the literature.) The difference between this metric and the criterion defined in step 1 is that the metric is measurable, i.e., is objective, whereas the criterion is testable, i.e., depends on users' evaluation.

   Application: In our case, we found a metric during the first turn of the cycle that accounted for 65% of the users' evaluation. We did not search for a metric during the next turns as we were not developing new strategies.

6. The optional step consists in writing automated users. It should be performed when users were directly experimenting with the system. Otherwise, it can be ignored. The aim of these programs is to approximate the dialogues of real users for the metric of step 5. Knowing which characteristics of the dialogue should be replicated by the automated users makes it easier to program them without having big corpora for evaluation. Once such automated users have been written, we can use them in place of real users. This is a big advantage as real users do often not have time to test a system under development. In particular this enables us to test strategies that were not evaluated by real users (for example because they were developed later). This is the aim of the next step.

   Application: In our case, we developed, during the first turn of the cycle, simple automated users answering or passing the system's communication.

7. In the next step, new strategies are developed by taking into account the insights provided by the experiment. This informed phase of development is better focused on the problem and should result in better strategies.

   Application: In our case, this step was performed during the first turn of the cycle. It mainly consisted in debugging the dialogue management strategies.

8. In this last step, we use the automated users of step 6 to produce dialogues and we evaluate them using the metric of step 5. We can therefore test different dialogue management strategies developed in the previous step at a minimum cost. If we want to test a new strategy, we can do it without requiring real users evaluation. Care should be taken however not to tailor the strategies to the particular

		Cycles			
		1	2	3	4
Steps	1	Issue: quality improvements	Issue: global coherence	Issue: local coherence	Stop
	2	Dialogue managers development (e.g., with local and global focus rules)			
	3	Interactive experiment design	Static experiment design	Static experiment design	
	4	Experiment	Experiment	Experiment	
	5	Metric			
	6	Automated users development			
	7	Further development and debugging			
	8	Further evaluations and selection	Selection	Selection	

Table 8: Applying development and test cycle in our approach

tests carried out. Separating training and evaluation data is a good way of ensuring that the new developments are well-tuned but not over-specified for the tests used.

Once strategies have been selected for further development, they are fed back to step 1 and the cycle recurs.

Application: In our case, we tested several new strategies (or strategy mixes) during the first turn of the cycle. During all turns we also dismissed poorly performing dialogue strategies.

Not all steps in the cycle are performed at each turn. A full turn is carried out during development. However during evaluation only, i.e., when no new strategies are created, the cycle is reduced to steps 1, 3 and 4 (typeset in italic in figure 8). This is particularly the case towards the end of development when the emphasis change from development to testing. It is a good idea to terminate the process by some pure evaluation turns (as was done in our work as shown in table 8).

It should be noted that each new cycle is not necessarily concerned with a subset of the previous turns' concerns. An orthogonal issue can be dealt with. However because dialogue management strategies are dismissed as soon as possible, more important issues should be dealt with first. Then only the most important strategies will be ranked during the next turns. If a minor issue is dealt with first, an important strategy could wrongly be eliminated. This is why we tested our strategies on issues of decreasing importance from first to third experiment.

The evaluation approach presented here can be used during development. The low cost of this method enables us to screen many different dialogue management strategies to pick out the best. These strategies can then be subjected to more in-depth evaluation. However it can also be used to test fully-blown systems. In this case, each cycle test a particularly important point of the systems so that their strength and weaknesses are evaluated. The evaluation stops as soon as one of the systems can be selected.

6 Further Work

In this section we review two areas where further work could prove beneficial.

The first area is user modelling. The current focus theories do not take into account the users' attentional preferences. If a topic can be spoken about, it will be mentioned. However tailoring the dialogues so that topics are ignored or presented in a satisfactory manner is needed to improve their perceived quality. This may involve taking a particular perspective on topics (McCoy 1988) or using cue words for marking unexpected topics (see section 3.1.3).

The second area is planning. A lot of work has been done on planning utterances to achieve some effect such as persuasion (Hovy 1993; Moore and Paris 1993). For the moment our system is using a very simple natural language interface. It could benefit from a better output module. This is especially the case if we want the system being able to provide motivations for its decisions. However, it is not clear how this can be done since the current elicitation modules do not provide much information on their reasoning which would make persuasion or argumentation difficult.

7 Conclusion

In this paper, we have shown why focus theories are important for elicitation systems. We then presented a formalisation of two focus theories. We have shown that these theories can be used in a practical application. We have then evaluated the theories and found that they provide definitive improvements in dialogue coherence, except for the local theory on its own. In order to perform the evaluation we have developed an innovative way of automating its process. Finally we presented a generalisation of our evaluation strategy suited to incremental development.

References

Barry W. Boehm. A spiral model of software development and enhancement. In Richard H. Thayer, editor, *Software Engineering Project Management*, pages 128–142. IEEE Computer Society Press, 1987. Reprinted in Thayer and Dorfman (1997, p. 416–430).

Jonathan P. Bowen. Seven more myths of formal methods. *IEEE Software*, 12(4): 34–41, 1995.

Susan E. Brennan, Marilyn Walker Friedman, and Carl J. Pollard. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics, Stanford University, Stanford, California, USA*, pages 155–162, 1987.

Peter Pin-Shan Chen. The entity-relationship model–toward a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36, 1976.

Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors. *Intentions in Communication*. The MIT Press, 1990.

Robin Cohen. A computational theory of the function of clue words in argument understanding. In *Proceedings of the 22nd annual meeting of the Association for Computational Linguistics, Stanford University, California, USA*, pages 251–258, 1984.

Robin Cohen. Analyzing the structure of argumentative discourse. *Computational Linguistics*, 13(1–2):11–24, 1987.

Robert Darimont. *Process Support for Requirements Elicitation*. PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 1995.

Alan Davis. A comparison of techniques for the specification of external system behavior. *Communications of the ACM*, 31(9):1098–1115, 1988.

Alan Davis, Scott Overmyer, Kathleen Jordan, Joseph Caruso, Fatma Dandashi, Anhtuan Dinh, Gary Kincaid, Glen Ledeboer, Patricia Reynolds, Pradid Sitaram, Anh Ta, and Mary Theofanos. Identifying and measuring quality in a software requirements specification. In *Proceedings of the* 1^{st} *International Software Metrics Symposium*, pages 141–152, 1993. Reprinted in Thayer and Dorfman (1997, p. 164–175).

Alan M. Davis. *Software Requirements: Objects, Functions, and States*. Prentice Hall, revised edition, 1993.

Norman Delisle and David Garlan. A formal specification of an oscilloscope. *IEEE Software*, 7(5):29–36, 1990.

Wieland Eckert, Esther Levin, and Roberto Pieraccini. User modeling for spoken dialogue system evaluation. Technical Report 97.33.1, AT&T Labs Research, 1997.

Wieland Eckert, Esther Levin, and Roberto Pieraccini. Automatic evaluation of spoken dialogue systems. Technical Report 98.9.1, AT&T Labs Research, 1998.

Norbert E. Fuchs and Rolf Schwitter. Attempto controlled English (ACE). In *Proceedings of the first international Workshop on Controlled Language Applications, Leuven, Belgium*, 1996.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2): 203–226, 1995.

Barbara J. Grosz and Candace Lee Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

Udo Hahn and Michael Strube. Centering-in-the-large: computing referential discourse segments. In *Proceedings of the* 35^{th} *Annual Meeting of the Assocation for Computational Linguistics and the* 8^{th} *Conference of the European Chapter of the Association for Computational Linguistics, Madrid, Spain*, pages 104–111. Association for Computational Linguistics, 1997.

Anthony Hall. Seven myths of formal methods. *IEEE Software*, 7(5):11–19, 1990.

Steve Hanks, Martha E. pollack, and Paul R. Cohen. Benchmarks, testbeds, controlled experimentation, and the design of agent architectures. *AI Magazine*, 14(4):17–42, 1993.

Janet Hitzeman and Massimo Poesio. Long distance pronominalisation and global focus. In *Proceedings of the* 17^{th} *International Conference on Computational Linguistics and* 36^{th} *annual conference of the Association for Computational Linguistics, Montréal, Canada*, pages 550–556, 1998.

Eduard H. Hovy. Automated discourse generation using discourse structure relations. *Artificial Intelligence*, 63:341–385, 1993.

Xiaorong Huang. Planning reference choices for argumentative texts. In Robin Fawcett, Eduard Hovy, David McDonald, Marie Meteer, Donia Scott, and Koenraad deSmedt, editors, *Proceedings of the 7^{th} International Workshop on Natural Language Generation, Kennebunkport, Maine, USA*, pages 145–152, 1994.

Masato Ishizaki. *Mixed-Initiative Natural Language Dialogue with Variable Communicative Modes*. PhD thesis, University of Edinburgh, Scotland, 1997.

Joachim Karlsson. Software requirements prioritizing. In *Proceedings of the second IEEE International Conference on Requirements Engineering, Colorado-Springs, CO, USA*, pages 110–118, 1996.

Richard Kittredge, Tanya Korelsky, and Owen Rambow. On the need for domain communication knowledge. *Computational Intelligence*, 7(4):305–314, 1991.

Renaud Lecœuche, Chris Mellish, and Dave Robertson. A framework for requirements elicitation through mixed-initiative dialogue. In *Proceedings of the third International Conference on Requirements Engineering, Colorado-Springs, CO, USA*, pages 190–196. IEEE Computer Press, 1998.

Miles Macleod. An introduction to usability evaluation. Technical Report DITC 202/92, National Physical Laboratory, 1992.

Kathleen F. McCoy. Reasoning on a highlighted user model to respond to misconceptions. *Computational Linguistics*, 14(3):52–63, 1988.

Kathleen R. McKeown. Discourse strategies for generating natural-language text. *Artificial Intelligence*, 27(1):1–41, 1985.

Bertrand Meyer. On formalism in specifications. *IEEE Software*, 2(1):6–26, 1985.

L. Mich. NL-OOPS: from natural language to object oriented requirements using the natural language processing system LOLITA. *Natural Language Engineering*, 2(2): 161–187, 1996.

George A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63:81–97, 1956.

Vibhu O. Mittal, Johanna D. Moore, Giuseppe Carenini, and Steven Roth. Describing complex charts in natural language: A caption generation system. *Computational Linguistics*, 24(3):431–467, 1998.

Johanna D. Moore and Cécile L. Paris. Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Computational Linguistics*, 19(4):651–694, 1993.

Sastry Nanduri and Spencer Rugaber. Requirements validation via automated language parsing. *Journal of Management Information Systems*, 12(2):9–19, 1996.

Derek Partridge. *Engineering Artificial Intelligence Software*. Intellect Books, Oxford, United Kingdom, 1992.

Livia Polanyi. A theory of discourse structure and discourse coherence. In W. H. Eilfort, P. D. Kroeber, and K.L Peterson, editors, *Papers from the general session of the* 21^{st} *regional meeting of the Chicago linguistics society, Chicago, Illinois, USA*, pages 306–322, 1985.

Chris A. Reed and Derek P. Long. Content ordering in the generation of persuasive discourse. In *Proceedings of the* 15^{th} *International Joint Conference on Artificial Intelligence, Nagoya, Japan*, 1997a.

Chris A. Reed and Derek P. Long. Ordering and focusing in an architecture for persuasive discourse planning. In W. Hoeppner, editor, *Proceedings of the* 6^{th} *European Workshop on Natural Language Generation, Duisburg, Germany*, 1997b.

Rachel Reichman. Conversational coherency. *Cognitive Science*, 2:283–327, 1978.

Rachel Reichman. Plain speaking: A theory and grammar of spontaneous discourse. Technical Report 4681, Bolt, Beranek and Newman Inc., Cambridge, MA, USA, 1981.

Rachel Reichman. Extended person-machine interface. *Artificial Intelligence*, 22:157–218, 1984.

Rachel Reichman. *Getting Computers to Talk Like You and Me*. The MIT Press, Cambridge, MA, USA, 1985.

Ehud Reiter. NLG vs. templates. In *Proceedings of the* 5^{th} *European workshop on natural language generation, Leiden, The Netherlands*, 1995.

Colette Rolland. A contextual approach for the requirements engineering process. In *Proceedings of the sixth international conference on Software Engineering and Knowledge Engineering, Vilnius, Lithuenia*, 1994.

Colette Rolland and Camille Ben Achour. Guiding the construction of textual use case specifications. *Data and knowledge engineering journal*, 25(1–2):125–160, 1998.

Thomas L. Saaty. How to make a decision: The Analytic Hierarchy Process. *European Journal of Operational Research*, 48:9–26, 1990.

Roger C. Schank. Rules and topics in conversation. *Cognitive Science*, 1:421–441, 1977.

Samira Si-Saïd and Colette Rolland. Formalising guidance for the CREWS goal-scenario approach to requirements engineering. In *Proceedings of the* 8^{th} *European-Japanese Conference on Information Modelling and Knowledge Bases, Vamala, Finland*, 1998.

Penelope Sibun. Generating text without trees. *Computational Intelligence*, 8(1):102–122, 1992.

Teresa Sikorski and James F. Allen. A task-based evaluation of the TRAINS-95 dialogue system. In Elisabeth Maier, Marion Mast, and Susann LuperFoy, editors, *Proceedings of the ECAI Workshop on Dialogue Processing in Spoken Languages Systems, Budapest, Hungary*, volume 1236 of *LNAI*, pages 207–220, Berlin, 1996. Springer.

Ian Sommerville. *Software Engineering*. Addison-Wesley, 5^{th} edition, 1995.

Ian Sommerville and Pete Sawyer. *Requirements Engineering: A good practice guide*. John Wiley & Sons, 1997.

Jeanine Souquières. *Aides au Développement de Spécifications*. PhD thesis, Université de Nancy I, Nancy, France, 1993.

Jeanine Souquières and Nicole Lévy. Proplane: A specification development environment. In *Proceedings of AMAST'96, Munich, Germany*, 1996.

Karen Sparck Jones and Julia R. Galliers. *Evaluating Natural Language Processing Systems*, volume 1083 of *Lecture Notes in Artificial Intelligence*. Springer, 1996.

J. Michael Spivey. Specifying a real-time kernel. *IEEE Software*, 7(5):21–28, 1990.

Richard Thayer and Martin Dorfman, editors. *Software Requirements Engineering*. IEEE Computer Society Press, 2^{nd} edition, 1997.

Marilyn A. Walker. The effect of resource limits and task complexity on collaborative planning in dialogue. *Artificial Intelligence*, 85(1–2):181–243, 1996.

Marilyn A. Walker, Don Hindle, Jeanne Fromer, Giuseppe Di Fabbrizio, and Craig Mestel. Evaluating competing agent strategies for a voice email agent. In *Proceedings of the* 5^{th} *European Conference on Speech Technology and Communication , Rhodes, Greece*, 1997a.

Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella. PARADISE: A framework for evaluating spoken dialogue agents. In *Proceedings of the* 35^{th} *annual meeting of the Association for Computational Linguistics, Madrid, Spain*, 1997b.

Paul T. Ward. The transformation schema: An extension of the data flow diagram to represent control and timing. *IEEE Transactions on Software Engineering*, 12(2): 198–210, 1986.

Roel J. Wieringa. *Requirements Engineering – Frameworks for understanding*. John Wiley & Sons, 1996.

A Entity-Relationship model

entityset(research group)	entityset(publication set)
entityset(researcher set)	entityset(site)
entityset(home page)	entityset(page)
relationset(present)	relationset(involve)
relationset(describe)	relationset(link)
cardmin(research group, 1)	cardmax(research group, 1)
cardmin(publication set, 1)	cardmax(publication set, 1)
cardmin(researcher set, 1)	cardmax(researcher set, 1)
cardmin(site, 0)	cardmax(site, ∞)
cardmin(home page, 0)	cardmax(home page, ∞)
cardmin(page, 0)	cardmax(page, ∞)
role(present, presented by)	role(present, presenting)
role(involve, involved in)	role(involve, involving)
role(describe, described by)	role(describe, describing)
role(link, linked from)	role(link, linked to)
rolefiller(present, presented by, site)	
rolefiller(present, presented by, page)	
rolefiller(present, presenting, research group)	
rolefiller(present, presenting, researcher set)	
rolefiller(present, presenting, publication set)	
rolefiller(involve, involved in, research group)	
rolefiller(involve, involving, publication set)	
rolefiller(involve, involving, researcher set)	
rolefiller(describe, described by, home page)	
rolefiller(describe, describing, site)	
rolefiller(link, linked from, home page)	
rolefiller(link, linked from, page)	
rolefiller(link, linked to, home page)	
rolefiller(link, linked to, page)	
cardmin(present, presented by, 0)	cardmax(present, presented by, 1)
cardmin(present, presenting, 0)	cardmax(present, presenting, 1)
cardmin(involve, involved in, 1)	cardmax(involve, involved in, 1)
cardmin(involve, involving, 1)	cardmax(involve, involving, 1)
cardmin(describe, described by, 0)	cardmax(describe, described by, 1)
cardmin(describe, describing, 1)	cardmax(describe, describing, 1)
cardmin(link, linked from, 0)	cardmax(link, linked from, ∞)
cardmin(link, linked to, 0)	cardmax(link, linked to, ∞)
rolecardmin(present,presenting, 1)	
attribute(home page, title, text)	
attribute(page, title, text)	

Table 9: Example of ER model – textual notation

B Translation Rules

The relations presented below hold for any time t.

Rule	ER Model	Focus Rule Model	Explanation
D1	$entity(E,T_e) \wedge attribute(T_e,A,T_a)$	$dir_t(D_i,E,T_a)$	If an entity has an attribute, this attribute can be mentioned immediately.
D2	$rolefillervalue(R,Role_1,E_1) \wedge rolefillervalue(R,Role_2,E_2) \wedge Role_1 \neq Role_2 \wedge E_1 \in \mathcal{F}_{i,t} \wedge E_2 \in \mathcal{F}_{i,t}$	$dir_t(D_i,E_1,R) \wedge dir_t(D_j,E_2,R)$	If two entities are in the same space, the relation linking them can be mentioned immediately.
D3	$relation(R,T_r) \wedge role(T_r,Role_1) \wedge role(T_r,Role_2) \wedge Role_1 \neq Role_2 \wedge rolecardmin(Role_1,1) \wedge cardmax(T_r,Role_1,1) \wedge cardmin(T_r,Role_2,0) \wedge cardmax(T_r,Role_2,1) \wedge rolefillervalue(R,Role_1,E_1)$	$dir_t(D_i,E_1,R)$	If an entity is optional and dependent on another one, the relation linking the two can be mentioned immediately.
D4	$rolefillervalue(R,Role,E)$	$dir_t(D_i,R,E)$	If a relation refers to an entity, this entity can be mentioned immediately.
S1	$relation(R,T_r) \wedge role(T_r,Role_1) \wedge role(T_r,Role_2) \wedge Role_1 \neq Role_2 \wedge rolecardmax(Role_2,C) \wedge C > 2 \wedge rolefillervalue(R,Role_1,E_1)$	$spe_t(S_i,E_1,R)$	If an entity can potentially be linked to several others filling the same role, the corresponding relations can be pushed on new (parallel) spaces.
S2	$relation(R,T_r) \wedge role(T_r,Role_1) \wedge role(T_r,Role_2) \wedge Role_1 \neq Role_2 \wedge cardmin(T_r,Role_1,0) \wedge cardmax(T_r,Role_1,1) \wedge rolecardmin(T_r,Role_2,0) \wedge cardmax(T_r,Role_2,1) \wedge rolefillervalue(R,Role_1,E_1)$	$spe_t(S_i,E_1,R)$	If an entity is optional but not dependent on another one, the relation linking the two can be pushed on a new space.
S3	$rolefillervalue(R,Role_1,E_1) \wedge rolefillervalue(R,Role_2,E_2) \wedge Role_1 \neq Role_2 \wedge spe_t(S_i,E_1,R)$	$spe_t(S_i,E_1,E_2)$	Two entities are in a specialisation relation if the relation linking them is specialisation relation with one of them.
G1	$rolefillervalue(R,Role_1,E_1) \wedge rolefillervalue(R,Role_2,E_2) \wedge Role_1 \neq Role_2 \wedge spe_t(S_i,E_2,R)$	$gen_t(G_i,E_1,R)$	An entity is in a generalisation relation with a relation if the other role filler is in a specialisation relation with this relation.
R1	$rolefillervalue(R,Role,E)$	$rel_t(R_i,E,R) \wedge rel_t(R_j,R,E)$	An entity and a relation are in relation if the entity is a role filler of the relation.
R2	$rolefillervalue(R,Role_1,E_1) \wedge rolefillervalue(R,Role_2,E_2) \wedge Role_1 \neq Role_2$	$rel_t(R_i,E_1,E_2)$	Two entities are in relation if they are linked by a relation.

Table 10: Translation Rules

C Dialogue example

S: research group 1 is a research group.
Do you want a site presenting it?
U: Yes.
S: A site presents research group 1.
Which site presents research group 1?
U: A new site.
S: Site 22 presents research group 1.
Do you want a home page describing site 22?
U: Yes.
S: A home page describes site 22.
Which home page describes site 22?
U: A new home page.
S: Home page 24 describes site 22.
What is the title of home page 24?
U: research group.
S: Do you want it a page linked from it?
U: Yes.
S: A page is linked from research group home page 24.
Which page is linked from research group home page 24?
U: A new page.
S: Page 31 is linked from research group home page 24.
What is the title of page 31?
U: researcher set.
S: Do you want another home page linked to it?
U: No.
S: Do you want another page linked from it?
U: No.
S: Do you want the researcher set presented by it?
U: Yes.
S: researcher set 7 is presented by researcher set page 31.
researcher set 7 is a researcher set.
it is involved by research group 1.
Do you want a home page linked from researcher set page 31?
U: I'll answer this question later.
I pass this question.
S: Do you want a page linked to researcher set page 31?
U: I'll answer this question later.
I pass this question.
S: Do you want a navigator pointing to it?
U: I'll answer this question later.
I pass this question.
S: Do you want a navigator displayed by it?
U: I'll answer this question later.
I pass this question.
S: Let's come back to our previous topic.
project set 15 is involved by research group 1.
project set 15 is a project set.

Do you want a site presenting it?
U: No.
S: Do you want a page presenting it?
U: Yes.
S: A page presents project set 15.
Which page presents project set 15?
U: A new page.
S: Page 43 presents project set 15.
What is the title of page 43?
U: project set.
S: Do you want a page linked from it?
U: No.
S: Do you want a page linked to it?
U: No.
S: Do you want a home page linked from it?
U: No.
S: Do you want a home page linked to it?
U: Yes.
S: A home page is linked to project set page 43.
Which home page is linked to project set page 43?
U: home page 24.
S: Do you want project set page 43 linked from another home page?
U: I'll answer this question later.
I pass this question.
S: Do you want project set page 43 pointed by a navigator?
U: I'll answer this question later.
I pass this question.
S: Do you want it displaying a navigator?
U: I'll answer this question later.
I pass this question.
S: Let's come back to our previous topic.
publication set 11 is a publication set.
it is involved by research group 1.
Do you want a site presenting it?
U: No.
S: Do you want a page presenting it?
U: Yes.
S: A page presents publication set 11.
Which page presents publication set 11?
U: A new page.
S: Page 51 presents publication set 11.
What is the title of page 51?
U: publication set.
S: Do you want a home page linked to it?
U: Yes.
S: A home page is linked to publication set page 51.
Which home page is linked to publication set page 51 ?
U: home page 24.
S: Do you want another page linked from publication set page 51?
U: No.

S: Do you want publication set page 51 linked from another home page ?
U: No.
S: Do you want a page linked to it?
U: No.
S: Do you want a home page linked from it?
U: No.
S: Do you want a navigator displayed by it?
U: Yes.
S: A navigator is displayed by publication set page 51.
Which navigator is displayed by publication set page 51 ?
U: A new navigator.
S: Navigator 59 is displayed by publication set page 51.
Do you want navigator 59 displayed by another page ?
U: I'll answer this question later.
I pass this question.
S: Do you want a home page displaying it?
U: I'll answer this question later.
I pass this question.
S: Do you want a page pointed by it?
U: I'll answer this question later.
I pass this question.
S: Do you want a home page pointed by it?
U: I'll answer this question later.
I pass this question.
S: Do you want publication set page 51 pointed by the navigator ?
U: I'll answer this question later.
I pass this question.
S: Do you want research group home page 24 linked to another page ?
U: I'll answer this question later.
I pass this question.
S: Do you want a page linked to research group home page 24?
U: No.
S: Do you want a home page linked from it?
U: No.
S: Do you want a home page linked to it?
U: No.
S: Do you want the navigator pointing to it?
U: Yes.
S: Navigator 59 is pointing to research group home page 24.
Do you want research group home page 24 displaying the navigator ?
U: No.
S: Let's come back to the questions that are still unanswered.
Do you want another page linked from it?
U: No.
S: Do you want navigator 59 pointing to a page ?
U: No.
S: Do you want navigator 59 displayed by another page ?
U: Yes.
S: A page displays navigator 59.
Which page displays navigator 59?

U: page 31.
S: Do you want navigator 59 pointing to another home page ?
U: No.
S: Do you want navigator 59 displayed by another page ?
U: Yes.
S: A page displays navigator 59.
Which page displays navigator 59 ?
U: page 43.
S: Do you want a home page displaying navigator 59?
U: No.
S: Do you want navigator 59 displayed by another page?
U: No.
S: Do you want project set page 43 pointed by the navigator?
U: No.
S: Do you want project set page 43 linked from another home page?
U: No.
S: Do you want researcher set page 31 linked to a home page?
U: I don't remember the identifier.
I pass this question.
S: Do you want researcher set page 31 pointed by the navigator?
U: No.
S: Do you want a page linked to it?
U: No.
S: Do you want publication set page 51 pointed by the navigator?
U: No.
S: Do you want researcher set page 31 linked to a home page ?
U: No.

D Questionnaire for Evaluation 1

Please answer these questions based on your latest dialogue with the system. For specific comments on the dialogue, refer to the sentence numbers given by the replay_dialogue function.

- For how long have you been using this system before this dialogue ?
  never / less than 5 min. / between 5 and 20 min. / more than 20 min.
- Was it easy to state the requirements ?
  not at all / no / somewhat / yes / yes very much
- Was the dialogue related to the task you were performing ?
  rarely / sometimes / usually / often / almost always
- Was the dialogue easy to understand ?
  not at all / no / somewhat / yes / yes very much
- Was the order of the communications (presentations and questions) correct ?
  rarely / sometimes / usually / often / almost always
- Additional Comments:

E Comparison Grids for Evaluations 2 and 3

E.1 Filling the grids

	Dialogue 1	Dialogue 2	Dialogue 3	Dialogue 4	Dialogue 5	Dialogue 6
Dialogue1	1					
Dialogue2	X	1				
Dialogue3	X	X	1			
Dialogue4	X	X	X	1		
Dialogue5	X	X	X	X	1	
Dialogue6	X	X	X	X	X	1

Table 12: Grid

Fill the cell at line i and row j with[6]:

1 If Dialogue i and Dialogue j are equivalent (for the criterion under consideration)

3 If Dialogue i is somewhat better than Dialogue j

5 If Dialogue i is better than Dialogue j

7 If Dialogue i is much better than Dialogue j

9 If Dialogue i is outstanding compared to Dialogue j

1/3 If Dialogue j is somewhat better than Dialogue i

1/5 If Dialogue j is better than Dialogue i

1/7 If Dialogue j is much better than Dialogue i

1/9 If Dialogue j is outstanding compared to Dialogue i

E.2 Global Coherence

	Dialogue 1	Dialogue 2	Dialogue 3	Dialogue 4	Dialogue 5	Dialogue 6
Dialogue1	1					
Dialogue2	X	1				
Dialogue3	X	X	1			
Dialogue4	X	X	X	1		
Dialogue5	X	X	X	X	1	
Dialogue6	X	X	X	X	X	1

Table 13: Global Coherence Grid

Which dialogue is *globally* more coherent, i.e., is divided into well defined chunks that are correctly related together ?
Globally coherent dialogues should focus on something for a while and then move on a related topic. Here are examples of global coherence and non-coherence.
Coherent:

[6]These numbers are given in Saaty (1990)

> B: How do you teach your students to use a calculator ?
> A: I think students should use a calculator for a while. I give them problems to solve with it, and when they have trouble, I answer their questions about the problems.
> B: That's all well and good, but I think they need more instruction on the device to reduce the number of questions. Instead I give them instructions, and they use these to solve problems. They don't have much trouble learning to use the machine.

Not Coherent:

> B: How do you teach your students to use a calculator ?
> A: I think students should use a calculator for a while. I give them problems to solve with it, and when they have trouble, I answer their questions about the problems.
> B: Well, I think you are wrong. Here's why. I'm going on a vacation to Tahiti tomorrow. I'm going by plane, and I'll be there about a week. It is going to cost me a bundle of money.

More paticularly, if we compare the beginning of dialogue 1 wih the beginnings of dialogues 5 or 6 (see table 14), we may find that the latter are less globally coherent because they move around too much. Other problems may be, for example, the fact of never jumping back to topics that should be dealt with or jumping without warning. The comparison value depends on your perception of the overall difference between the two dialogues on this aspect. The more you perceive the difference, the bigger the value should be.

Dialogue 1	Dialogue6
...	...
research group 1 is a research group	project set 15 is involved by research group 1
Do you want it presented by a page ?	research group 1 is a research group
No	researcher set 7 is involved by it
Do you want it presented by a site ?	researcher set 7 is a researcher set
Yes	Do you want it presented by a site ?
research group 1 is presented by a site	No
Which site is presenting research group 1 ?	Do you want it presented by a page ?
A new site	Yes
research group 1 is presented by site 22	researcher set 7 is presented by a page
Do you want site 22 described by a homepage ?	Which page is presenting researcher set 7 ?
Yes	A new page
...	...

Table 14: Global Coherence Example

(The answers of the user should not be considered. In particular, the fact that some communications are passed, i.e., not answered immediately after they are asked, should not be taken into account.)

E.3 Local Coherence

	Dialogue 1	Dialogue 2	Dialogue 3	Dialogue 4	Dialogue 5	Dialogue 6
Dialogue1	1					
Dialogue2	X	1				
Dialogue3	X	X	1			
Dialogue4	X	X	X	1		
Dialogue5	X	X	X	X	1	
Dialogue6	X	X	X	X	X	1

Table 15: Local Coherence Grid

Which dialogue is *locally* more coherent, i.e., has well related successive sentences ? A dialogue is locally coherent if when sliding a window spanning two sentences over the text, the visible part is still related. Here are examples of local coherence and non-coherence.
Coherent:

A: I just bought a new hat.
B: It's nice. Where did you buy it ?
A: In the shop at the corner.

Not Coherent:

A: I just bought a new hat.
B: Fred eats hamburgers.
A: My car is fast.

More paticularly, if we compare dialogue 1 with dialogue 5 (see table 16), we can see that questions in dialogue 5 shift from, for example, page to homepage whereas they do not in dialogue 1. Therefore dialogue 1 is more locally coherent than dialogue 5 at that point. (The same may of course happen with other entities then page and homepage.)

Dialogue 1	Dialogue5
...	...
Do you want it linked to a page ?	Do you want project set page 38 linked from a page ?
No	No
Do you want it linked from a page ?	Do you want project set page 38 linked from a homepage ?
No	Yes
Do you want it linked to a homepage ?	Do you want project set page 38 linked to another page ?
No	No
Do you want it linked from a homepage ?	Do you want project set page 38 linked to a homepage ?
Yes	No
...	...

Table 16: Local Coherence Example