User-System Dialogues and the Notion of Focus

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

Chris Mellish[†]
chrism@dai.ed.ac.uk

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

Dave Robertson[†]
dr@dai.ed.ac.uk

Abstract

In recent years, the capabilities of knowledge-based systems to communicate with their users have evolved from simple interactions to complex dialogues. With this evolution comes a need to understand what makes a good dialogue. In this paper, we are concerned with dialogue coherence. We review the notion of focus, which partly explains this property, and its use for user-system communication. First, we examine the major theories dealing with this notion. We describe what their contribution is and how they differ. Then, we illustrate the benefits of using the notion of focus and especially the improvement in text coherence. We pay particular attention to how the notion can concretely be implemented. Its integration with other techniques and theories is described. We conclude the paper by pointing out remaining issues in the understanding of the notion of focus. The contribution of this paper is to provide a classification of the theories of focus and to show the improvements they offer in elaborate user-system dialogues.

1 Introduction

Since the early days of expert systems, the capabilities of knowledge-based systems to communicate have greatly improved. For example, explanation has been recognised has a complex task that requires specific knowledge (Paris et al. 1988; Paris 1991; Southwick 1989, 1991; Wick 1994). As a result, systems are now able to engage in a dialogue with their users. Our research leads us also to consider how these capabilities can be used for requirements elicitation systems (Lecœuche et al. 1998). Communication is indeed beginning to play a more and more central role in these systems. The emphasis is on guiding the users through the whole elicitation process (Rolland 1993, 1994). This involves participating in a coherent dialogue with the users. This dialogue should enable them to readily understand what the system wants to elicit. In example 1, a system which is eliciting information about a World Wide Web site directs the users' attention to the navigation between pages before asking its question. This helps the user in following the system reasoning.

[*]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

Example 1

> Let's see how pages are linked together. Do you want to link all the pages to the site home page ?
> ...

A set of rules accounting for easily intelligible dialogues is needed to provide such a dialogue. Thus, constructing full discourses requires more knowledge than the sentence long utterances of early systems because relations between sentences must be considered (Mann et al. 1981; Grice 1975, p. 46). Not taking this point into account leads to disconnected discourses that are difficult to understand. Users usually presume that the computer will produce coherent text. If this is not the case, they will waste their time in trying to understand why sentences are put together as they are.

In this paper, we review one set of discourse rules. Focus rules define dialogue coherence in terms of focus and focus shifts[1]: the focus of a dialogue can only change in some well defined ways. The rules operate at different levels, from word selection up to global dialogue structure, and are based on different concepts, such as dialogue purpose and "real-world" organisation. Although they are necessary for producing coherent text, focus rules have often been ignored in the generation of dialogues.

We also present applications making use of these rules. These examples provide evidence that the notion of focus is mature enough to provide real improvements in the production of dialogues and their understanding by users.

The paper is organised as follows. In section 2, we review the major theories that explain the notion of focus. We first define the notion of focus and distinguish between global focus and local focus, reflecting the level of focus considered by the theories. Then we describe the theories in more detail and we classify them by the kind of information they use to keep track of the focus. In section 3, we illustrate the notion of focus with applications demonstrating its use and its usefulness. We pay particular attention to how the notion of focus can concretely be implemented. Its integration with other techniques and theories, such as explanation strategy selection and rhetorical structure theory, is described. In section 4, we point out open issues in the understanding of the notion of focus. These are important for the actual use of the notion but have not yet been fully researched. Finally we conclude in section 5 with some thoughts on the future use of the notion of focus in user-system dialogues.

2 Theory

A great deal of work is related to the notion of focus. In this paper we present theories that (1) are general enough to explain a wide range of focus-related linguistic phenomena (2) have direct implications for knowledge-based system communication. The presentation is organised along two criteria. The first

[1] These terms have been used extensively in the literature with varying meanings. We use them here informally. Focus should be understood as being the things to which the user is attending. A precise definition will be given in the next section.

Focus Theories				*section 2*
	Global focus theories			*section 2.b*
		“Real world” based theories		
			Task based theories	*section 2.b.1*
			Domain based theories	*section 2.b.2*
		“Participant” based theories		
			Intention based theories	*section 2.b.3*
			Memory based theories	*section 2.b.4*
		Dialogue based theories		
			Dialogue move based theories	*section 2.b.5*
	Local focus theories			*section 2.c*
		“Real world” based theories		
			Domain based theories	*section 2.c.4*
		“Participant” based theories		
			Semantics based theories	*section 2.c.3*
		Dialogue based theories		
			Sentence form based theories	*section 2.c.1*
			Dialogue move based theories	*section 2.c.2*

Table 1: Classification of focus theories and corresponding sections

distinguishes between global and local focus theories. (The definitions of global focus and local focus are given in the next section.) The second differentiates the theories according to the knowledge they use to keep track of the focus. We distinguish between “real world” based theories, “participant” based theories and dialogue based theories. In “real world” based theories, focus evolution is mainly constrained by some property of the external environment (e.g., task to be achieved, structure of the objects spoken about). In “participant” based theories, focus evolution is mainly constrained by some property of the dialogue participants (e.g., their intentions or beliefs). In dialogue based theories, focus evolution is mainly constrained by the previous dialogue.

This classification is summarised in table 1.

2.a Definitions

We define the focus as being *the set of all the things to which participants in a dialogue are attending at a certain point in a dialogue and the point of view they have on these things.* The things represent *what* is focused on. They can be concrete or abstract and they can either be explicitly mentioned in the dialogue or they can be implicit. The point of view represent *how* the things are focused on. It depends on the dialogue history.

For example, imagine a conversation where the US military strength in general is discussed. The conversation then evolves to the Star Wars project and its funding. The focus of this conversation could be represented as in figure 1. The things in focus are what has been discussed while the point of view de-

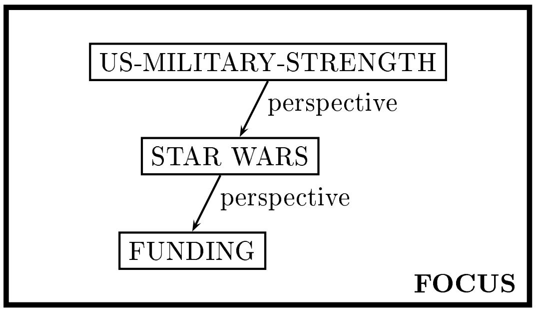

Figure 1: Focus definition (adapted from McCoy and Cheng (1991), p. 109)

pends on the dialogue history. (The funding is seen from the Star Wars project perspective and is related to the overall issue of the US military strength.)

Focus represents the conversational context in which future utterances are interpreted. If the new sentences cannot be interpreted with respect to the focus, the discourse will seem incoherent.

A dialogue can be coherent or incoherent at several levels. For example, the dialogue in example 2 is incoherent at the sentence level: the two sentences do not seem to focus on the same topic. In example 3, the dialogue is incoherent

---

Example 2 (from Schank (1977), p. 421)

A: I just bought a new hat.
B: Fred eats hamburgers.

---

at a higher level: the last paragraph, although internally coherent, does not fit with the preceding conversation.

Example 3 (from Sidner (1983), p. 111–112 and 127–128)

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.
C: Well, I think you are both 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.

These different levels of coherence are reflected in the distinction between global focus and local focus (McKeown 1985a, p. 56). On the one hand, *local* *focus*, also called immediate focus, deals with sentence level coherence (Sidner 1983, p. 107). It explains how two consecutive sentences are related. *Global focus*, on the other hand, deals with higher-level coherence (Grosz 1977, p. 5). It explains how parts of the dialogue are related. As we shall see, distinguishing between these two kinds of focus is necessary because they obey different sets of rules.

We review now these two notions in turn.

2.b Global focus

In this section, the global focus theories are classified according to the kind of knowledge they use to keep track of the focus. We distinguish between theories based on "real world" properties such a task structure and domain structure, theories based on participants' properties such as intentions or memory limitations, and theories based on dialogue properties such as dialogue moves.

The theories are presented in order of complexity; later theories usually extending previous ones. For each category, we have selected a main piece of work representing it, preferring work that has been used for implementing dialogue management systems.

2.b.1 Based on task structure: Grosz (Grosz 1977)

In this approach, the focus is divided into focus spaces organised as a *stack*. A focus space represents the information in focus during part of the dialogue. It is divided into two components. The first component is the *explicit focus*. It contains the objects and events that have been mentioned during that discourse part. The second component is the *implicit focus*. It contains objects and events related to the elements in explicit focus but not directly mentioned in the discourse. Recording both types of elements is important as the dialogue can refer to objects or events in explicit or implicit focus as shown in example 4. In this example, the explicit focus only contains references to the lid, the container and the action of attaching them together. However, the bolts are put in implicit focus because they are related to the action of attaching the lid to the container. Therefore the reference to them is easily understood.

Example 4 (from Grosz (1977), p. 68)

> S: Attach the lid to the container.
> R: Where are the bolts ?

The top-most space in the stack is the *active* space. It contains the elements that are in focus at the current point in the dialogue. When a new topic is discussed, a new space is pushed on the stack and therefore becomes active. The other spaces are *open* spaces. They contain information about previous topics that have not yet been completed. This information is still available for reference in the dialogue but is less accessible than the information in the active space. The open spaces can become active again as the dialogue evolves and shifts back to uncompleted issues. This is shown in figure 2. In this example, Space 0 is pushed on top of the existing spaces. It contains elements that are

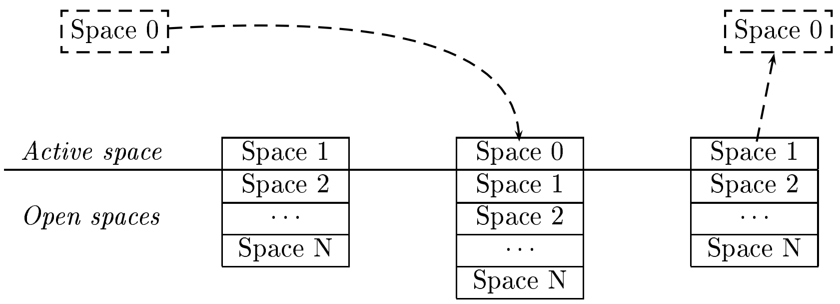

Figure 2: Examples of stack movements. Space 0 becomes the new active space. Then it is popped and Space 1 becomes active again.

then more salient in the discourse than the elements of the open spaces. Then it is popped and Space 1 becomes active again.

Grosz used this framework to study the global focus in conversations between an expert and an apprentice assembling part of a compressor. She found that the movements of the focus, i.e., the stacking and un-stacking of focus spaces, mirror the structure of the task discussed (Grosz 1977, p. 104–105). A new focus space is pushed on the stack whenever a new task begins. The focus space is then used for reference during the realization of the task. It is finally popped from the stack when the task is completed. For example, if completion of Task 1 requires completion of Task 0, a space for Task 1 will first be pushed on the stack. Then a space for Task 0 will be pushed as this subtask is dealt with. Then it will be popped and Task 1 space will become active again. This corresponds to the stack movements in figure 2.

The stack movements are reflected in the dialogue by different means. Pushing a new space is usually indicated by a sentence that refers to the goal or the objects involved in the new task. Popping a space is usually marked by references to elements in an open space of the stack. The stack is then popped so that the open space referred to becomes the active focus space again. Recognition of topic shifts is also helped by the task structure *shared* by the participants: the hearer knows when to expect a shift.

As new elements are included into a focus space, an important question is to know what implicit information should be added, since this information can influence focus shifts. Grosz proposes two principles (Grosz 1977, p. 67–68):

- for physical objects, add the subparts of the objects;
- for events, add the sub-events as well as their participants.

Although similar choices are also made in other theories, no explanation is given to back it up.

A number of problems remain open in this theory (Grosz 1981). Notably, the relation between focus and belief is not clear. Speaker and hearer may have different knowledge of the concepts spoken about in the dialogue. Therefore they may focus differently on them (Levelt 1989, p. 119–120). This is not reflected in the theory where a unique focus space is used. This is because focus is seen as "a property of the discourse itself, not of the discourse participants"

(Grosz and Sidner 1986, p. 179). Using two spaces, one for the speaker and one for the hearer, would require one to study issues such as how the two are synchronised and how the speaker's belief about the hearer's focus influences the dialogue.

2.b.2 Based on domain structure: Linde (Linde 1979)

Linde's approach is very similar to that of Grosz. The focus representation is identical but the shifts depend on the domain structure rather than on the tasks to be carried out (although there is an overlap between the two).

Linde used her framework to study the global focus in descriptions of apartments. She found that the movements of the focus are based on the apartment lay-out. The descriptions begin with the entrance hall and then describe each room in turn as if the speaker was walking through the apartment.

This approach first shows that the task structure is not the only organisation influencing the focus moves. Domain structure is another possibility. This is confirmed by Levelt (Levelt 1989, p. 142–143). Levelt observed that people describing a simple spatial configuration consisting of linked nodes use it to structure their discourse. In particular, they follow the "principle of connectivity" choosing as the next node to describe one that is linked to the latest node presented. When people have to come back to a previous node because they reach a dead-end, they do it by returning to the nearest choice point not explored. Levelt calls this the "stack principle" which also confirms Linde's theory.

A second point made by Linde is that the task or domain structures are not sufficient to determine the focus moves done in the dialogue. For example, if a room R1 leads to two others, R2 and R3, two possible focus moves are possible: R1 to R2 or R1 to R3. Choosing between these choices depends on criteria independent of the domain structure, such as the intention of the speaker in describing the apartments. This means that the focus shifts reflect the *dialogue organisation*. Focus shifts reflect the domain structure (or task structure) in so far as it constrains the dialogue organisation (by the principle of connectivity for example). In his experiment with a simple lay-out, Levelt found that people tend to minimise their cognitive load by describing simpler parts first, i.e., the ones that require less backtracking. However, it not clear how this applies to more complex lay-outs.

2.b.3 Based on dialogue intentions: Grosz & Sidner (Grosz and Sidner 1986)

Grosz and Sidner propose to divide up the dialogue according to the aims of its parts and to base the shifts on this division, rather than on a task or domain structure.

The basic focus mechanism is again similar to that of Grosz's original work. The focus space however not only contains the concepts spoken about but also the "discourse segment purpose" (DSP) (Grosz and Sidner 1986, p. 179). The DSP represents the intention of a particular part of the discourse. DSPs form a structure organised by two relations: dominance and satisfaction-precedes. Dominance means that the completion of one discourse segment contributes to the completion of the dominant discourse segment. Satisfaction-precedes means

Focus shift	Cue words
dominance push	for example, to wit, first, second, and, moreover, furthermore, therefore, finally
satisfaction-precedes push	in the first place, first, second, finally, moreover, furthermore
pop	anyway, but anyway, in any case, now back to, the end, ok, fine

Table 2: Cue words (from Grosz and Sidner (1986), p. 198)

that the first discourse purpose must be completed before the second can be satisfied. For example, if you intend to open a door, you may first intend me to tell you where the keys are and then you may intend to unlock the door. There is a dominance relation between the intention of opening the door and the intentions of finding the keys and opening the door. There is a satisfaction-precedance between the intention of finding the keys and the intention of opening the door.

The focus shifts are then dependent on the intentional structure of the dialogue: "a push [on the focus stack] occurs when the DSP of a new segment contributes to the DSP of the immediately preceding segment. When the DSP contributes to some intention higher in the dominance hierarchy, several focus spaces are popped from the stack before the new one is inserted" (Grosz and Sidner 1986, p. 180).

An advantage of this approach is that it bases the focus shifts on *dialogue* intentions and not on structures that are only indirectly related to the dialogue organisation (see section 2.b.2).

A major difficulty with this approach is the need to recognise a speaker's intention to determine the appropriate shift of focus (dominance or satisfaction-precedes). This implies that the speaker should provide enough information for the hearer to find out the correct focus shift *at the start* of a discourse segment. Three main kinds of information can be used to determine how the DSP of a new segment relates to the previous one: cue words, i.e., special phrases indicating a particular shift, utterance-level intentions and general world knowledge (Grosz and Sidner 1986, p. 188). Since cue words do not require inferences and can be put at the beginning of a new segment, they often indicate focus shifts. Grosz and Sidner propose different cue words for dominance push, satisfaction-precedes push, and pop[2]. These cue words are summarised in table 2. Recognition of focus shifts thanks to utterance-level intentions or general world knowledge is more complex and not yet fully understood.

2.b.4 Based on memory limitations: Walker (Walker 1996b)

Walker points out that stack based theories (see sections 2.b.1, 2.b.2 and 2.b.3) do not include "constraints related to the length, depth, or the amount of processing required for an embedded segment" (Walker 1996b, p.256). In other words, when a segment is popped from the stack, the previous segment infor-

[2]Grosz and Sidner take into account some other relations between discourse segments such as digression or feedback. We do not consider them here.

mation becomes immediately available again regardless of the popped segment length. This seems unlikely given people's memory limitations (Miller 1956). The longer the intervening segment is, the less likely previous segment information will still be available[3]. Therefore, Walker models focus by a cache rather than by a stack. The main characteristics of the cache are (Walker 1993, p. 33+):

1. The elements pushed more recently in the cache are more likely to be retrieved.

2. The elements pushed more frequently in the cache are more likely to be retrieved.

3. Elements are never explicitly popped from the cache but may be displaced. Displaced elements go to main memory. They can later be recalled to the cache if need be.

Because dialogue elements have different importance, priority levels are assigned to them. For example, elements related to the current topic or to a previous point that is not closed are assigned a higher priority level since we expect to continue speaking about them. Elements are then displaced from the cache by priority level order. This explains why we can "keep in mind" information that will be useful when returning to a previous topic (Guindon 1985).

The advantages of Walker's theory over stack based theories are:

1. a better account of how focused elements are kept salient for some time after their related dialogue segment is closed (Walker 1997). In the stack model, elements are immediately popped.

2. an explanation for "information redundant units" (IRU) (Walker 1996a). These units are so called because they reintroduce information already expressed in the dialogue. Walker argues that one of the roles of these units is to indicate that some elements should be put back into focus, i.e., transfered from main memory to cache. Stack based theories do not offer any explanations for IRU.

However, it is not clear how elements in the cache are prioritised. Recognition of the speaker's intention seems necessary (Walker 1996b). This encounters the same problems as for task based theories (see section 2.b.3).

2.b.5 Based on dialogue moves: Reichman (Reichman 1985)

This theory has three interesting features: (1) different kinds of focus spaces are distinguished, (2) various dialogue moves are used to navigate between spaces and these moves are constrained by the previous dialogue, and (3) focus spaces have activation levels. We review now these three points in more detail:

1. A first feature of Reichman's work is the use of different focus spaces depending on the nature of the discourse. For example, during debative

[3] This problem also arises in Reichman's theory (see section 2.b.5). For example, the length of an interruption should be taken into account when returning to the initial space. However, because Reichman's theory includes local focus rules changing the return space information focus level (see below section 2.c.2) the problem has less impact than for stack based theories.

Conversational move	Clue word(s)
Support	Because. . . ; Like. . . ; Like when. . .
Restatement and/or conclusion of point being supported	So. . .
Interruption	Incidentally. . . ; By the way. . .
Return to previously interrupted context space	Anyway. . . ; In any case. . .
Indirect challenge	Yes/Right but. . . ; Except, however. . .
Direct challenge	(No) But. . .
Subargument concession	All right/Okay but. . .
Prior logical abstraction	But look/listen/you see. . .
Further development	Now. . .

Table 3: Conversational moves (from Reichman (1984), p. 180)

discourses, `issue` spaces and `support` spaces are used. Differentiating various spaces allows one to record not only the elements in focus and the intention of a discourse part, as in Grosz and Sidner's work, but also specific information related to the particular type of discourse taking place. For instance, an `issue` space lists, among other, what kind of issue (epistemic, evaluative, deontic) is under discussion and its topic. A `support` space on the other hand contains information such as the source of the support, or its credentials. Using a richer representation framework to record specific information about each space allows a finer evaluation of the moves that can take place in the dialogue.

2. A second feature of Reichman's work is the differentiation of several conversational moves to explain shifts in focus. A conversational move is defined as a "communicative act serving a new discourse role" (Reichman 1984, p. 163). In other words, a conversational move provokes a focus shift. Reichman proposes several conversational moves such as interruption and return among others (Reichman 1978, p. 292–298). Each move is defined by a set of preconditions determining when it can apply, and a set of actions defining the result of its application. For example, if a participant defends a claim by making an analogy, a possible move is to challenge this support by expressing doubts on the validity of the analogy mapping. The precondition of this move is that indeed an analogy has been made and its result is to undermine the claim's support (leading to a counter-challenge or to the concession of the inadequacy of the analogy).

   A set of clue words indicates the moves taken in the dialogue. Some of the conversational moves and their associated cue words are summarised in table 3.

3. A final feature of Reichman's work is the recording of a focus space's level of activation. In stack-based theories, the spaces nearer the top of the stack are in higher focus. Spaces not in focus are popped from the stack. Reichman's framework is more complex. Focus spaces are not placed into a stack but assigned a focus level. The focus level has one of the following values (Reichman 1985, p. 54):

Active indicates the unique space where the current utterances are placed. It has the same meaning as the active space in Grosz's stack framework.

Controlling indicates the unique space to which the active space is related. Only the elements of the controlling and active spaces can be spoken about in the dialogue.

Open indicates a previously active space that has not been completed. Open spaces are candidates for focus shifts as in the Grosz stack framework.

Generating indicates a space in an indirect relation to the active space. For example, an issue space created when a support space is active would become active while the support space would be reassigned a generating focus level.

Closed indicates a space whose conversational move has been completed. Closed focuses are not deleted and may be reopened if necessary (although this is not expected).

Superseded indicates a previously active space that has not been completed but that is not expected to become active again.

Conversational moves modify the focus level of the focus spaces. When a space becomes active, the focus levels of all the focus spaces related to it are reassigned new values depending on their relations with the new active space. For example, an interruption move creates a new focus space, assigns an active level to it and assigns an open level to the previously active space since this space is not related, directly or indirectly, to the new active space (Reichman 1981, p. 52 and p. 100). At the same time a return move is put into an expectation list. It then has a higher priority to apply than other moves since an interruption should come back to the interrupted space[4]. At the end of the interruption, the expected return move closes the interruption space and re-assigns an active level to the suspended space. The use of the various focus levels as well as the expectation list makes the framework richer than the stack approach.

A possible problem with this approach is the ability to re-open closed focus spaces. As a result, clauses that are far apart in the dialogue can be put into the same space although the context of the dialogue may have changed. Polanyi proposes that closed spaces should remain closed. Related material should be put in a new space which is then explicitly linked with the closed space if need be (Polanyi 1985). The focus evolution is then closer to the dialogue structure (which is also the case for stack-based approaches).

2.b.6 Summary

Compared with the earlier theories of focus, we can notice in more recent theories an evolution towards richer representations. The representation has been

[4]This rule could be considered as modelling more than just how focus evolves but also how the dialogue is structured. However, since this structure is based on focus expectation, we consider it as being part of the focus theory. This is different from RST relations, discussed later in the paper, because these relations are not based on focus. Although they propose an account of text coherence, these relations are therefore not part of focus theories (Some attempts have been made to link RST and Grosz and Sidner's theory (Moser and Moore 1995). However they mainly concern the intentional aspect of the theories and not the attentional one.)

extended along three lines:

- It provides a better account of the role of intention in focusing. This is exemplified in Grosz and Sidner's theory (see section 2.b.3) where the focus shifts are dependent on the discourse participants' intentions.

- It provides a finer description of the dialogue structure. This is exemplified in Reichman's theory (see section 2.b.5) where numerous dialogue moves, with their preconditions and effects, are identified.

- It provides a better model of human information processing. This is exemplified by Walker's theory (see section 2.b.4).

These extensions take into account dialogue properties as well as "real world" properties such as task or domain structures (see table 1). They are also more realistic by recognising human limitations.

This evolution is accompanied with difficulties. For example, taking intention into account is a difficult task because it implies the capability to recognise speaker's intentions which is a complex problem. To achieve this task, Grosz and Sidner proposes the notion of shared plan between the dialogue participants (Grosz and Sidner 1990; Grosz and Lochbaum 1993). Knowing what has to be done enables them to decide when a shift occurs. The second line of improvement does not encounter such problems. Recognising a dialogue move is usually easier than recognising an intention because dialogue moves are more constrained. In other words, only a few moves can possibly apply at any point in a dialogue. However, while recognition of focus shifts is made easier, the generation of dialogue is not fully specified by these theories. Speaker's intentions should, at some point in the dialogue production process, play a role. Further progress in understanding this problem would require the study of the relation between intentions and dialogue moves in more detail. This relation may be a complex one since intentions and discourse relations are not mapped one-to-one in general (Moore and Paris 1993, p. 667).

A number of open issues remain even in the newer theories (Grosz 1981):

- What is the relation between focus and perspective ? People usually focus on concepts from a certain point of view (McCoy 1988). This is not reflected in the focus space where entities are in focus or not. A finer representation is required to reflect this phenomenon. Accordingly, finer shifting rules are required to allow only shifts that are compatible with the perspective taken.

- When is a reference in focus ? The focus mechanism is an on-off one. It is not clear when to consider an inexact referring phrase to still refer to a focused entity or to consider it as a new entity and maybe shift the focus based on that assumption.

2.c Local focus

In this subsection we describe theories of local focus, i.e., theories explaining fine-grained coherence. The theories are classified according to the kind of knowledge they use to keep track of the focus. We again distinguish between theories based on "real world" properties such as domain structure, theories based on participants' properties such as their cultural background, and theories based on dialogue properties such as sentence form and dialogue moves.

2.c.1 Based on sentence form: Sidner (Sidner 1979) and Centering (Grosz et al. 1995)

These two major theories are very similar and based on the same ideas. We first present their commonalities before describing them in turn.

Common ideas

These two theories of local focus explain the coherence between consecutive sentences based on the syntactic, grammatical and thematic functions of the concepts in the sentences. Each concept is assigned a focus level depending on its function. These levels then constrain the construction of the next sentence in the dialogue. This sentence sets, in turn, new focus levels that are used to constrain the following one, and the process recurs.

More formally, each sentence, except the first sentence of a dialogue, has a main element with respect to the previous sentence. This element is called the sentence focus. The first sentence of a dialogue does not have a focus because a sentence *on its own* does not have one. For example, if I say: "I bought a new hat at the shop round the corner yesterday", the sentence could be about me ("I am going to a wedding"), about the hat ("It is nice, isn't it") or about the shop ("It is amazing what they have in the store"). The focus is defined by the *relation between two sentences*, i.e., the focus of a sentence is a concept which is echoed from a previous one. Each sentence also has a list of potential focuses composed of the other concepts present in the sentence. The list is ordered by decreasing focus level.

The focus and the potential focus list are used to determine the focus of the next sentence. In other words, if we slide a window spanning two sentences over the dialogue, the first sentence in the window sets expectations on what the focus will be, based on the current focus and the new concepts introduced, and these expectations are confirmed, or not, by the other sentence.

The two theories differ in the way focus is represented, in the way focus levels are assigned to the sentence components and in the way these levels constrain the construction of the next sentence.

We present them now in turn.

Sidner's Theory

In this approach, two focuses are maintained. These two focuses are (1) the discourse focus, which has been discussed above, and (2) the actor focus which is the subject of the sentence. This distinction helps finding the focuses in discourses where the actor focus and the discourse focus are distinct but are both pronominalised as in example 5[5]. In this example, the actor focus is Jerome and the discourse focus is his pigeon.

Example 5 (from Sidner (1983), p. 282)

> Jerome took his pigeon out on a leash.
> Since he was trying to train it,
> he hollered "heel" and "run" at it,
> as they sauntered along.

When these focuses change in the course of the dialogue, they are stacked in two separate stacks. Therefore, it is possible to come back to a previous focus by un-stacking either stack.

As explained in the preceding introduction, the most important points in the theory are the assignments of a focus level to the concepts in a sentence, the computation of the expected focus and its confirmation or rejection. We now describe these three points.

The focus level assignments are done in a first step. During this step, the potential focus list, which contains the sentence elements ordered by decreasing focus level, is computed by the following rules (For the first sentence of the dialogue, special rules apply since there is no existing focus (Sidner 1983, p.287 and p. 297)):

1. If a cleft[6] or pseudo-cleft[7] sentence is used, the potential focus list is the cleft item if and only if the element in non-clifting position co-specifies the focus. When it does not the sentence is incoherent.

2. The potential focus list is composed of the concepts filling a thematic relation in the sentence, excluding the agent position[8] and the noun phrase which co-specifies the focus if one exists. The last member of the list is the sentence verb phrase.

The determination of the expected focus is then based on the current focus, the potential focus list and the next sentence form, since the focus is defined by the relation between two sentences. It is performed by the following algorithm (adapted from Sidner (1983), p. 292–294):

[5] A major use of this theory is indeed the resolution of pronoun references.

[6] A "cleft" sentence has the following form : It is/was NP that VP. For example, "It is John that ate the strawberries" is a cleft sentence.

[7] A "pseudo-cleft" sentence has the following form : What VP is/was NP. For example, "What John ate was the strawberries" is a pseudo-cleft sentence.

[8] The "agent" of a sentence is the thing that is described as being the cause or having the control of the event presented. For example, in "I have joined the KBS group", I am the agent of the sentence.

1. If the next sentence contains a do-anaphor[9] choose the last element of the potential focus list, i.e., the verb phrase, as the focus. Stack the current focus in the focus stack. Halt.
2. If there is an anaphor co-specifying the current focus, i.e., a word such as a pronoun referring to the current focus, and another co-specifying some member of the potential focus list, retain the current focus as focus if the anaphor is not in agent position. If it is, take the potential focus list member as the new focus and stack the current focus in the focus stack. Halt.
3. If there are anaphors which co-specify only the current focus, retain the current focus as focus. Halt.
4. If there are anaphors which co-specify only a member of the potential focus list, move the focus to it. If several members are co-specified, choose the one with the highest focus level. Stack the current focus in the focus stack. Halt.
5. If there are anaphors which co-specify only a member of the focus stack, move the focus to the stack member by popping the stack. Halt.
6. If there are no anaphors co-specifying any of the current focus, potential focus list member or focus stack member but the current focus can fill a non-obligatory case in the sentence or if the verb phrase is related to the current focus by pronominalisation, retain the current focus as focus. Halt.
7. If there are no foci mentioned, retain the current focus as focus.

Finally, the expected focus can be rejected if it is grammatically or semantically incorrect. For example, a pronoun referring to the focus and the focus should agree in gender and number. If no such incoherence exists, the expected focus becomes the current focus; otherwise the preceding algorithm is run again to find the second best expected focus.

A final point to consider in this approach is the use of the actor focus and its interaction with the discourse focus.

Tracking actor focus is much simpler than tracking the discourse focus. The actor focus is indicated by the subject of the sentence. If the subject is a pronoun, it refers to the current actor focus except if the discourse focus is animate and is established earlier in the discourse than the actor focus. In that case the discourse focus takes precedence (Sidner 1983, p. 307). More complex interactions are explained in Sidner (1979) but we do not consider them here.

We show now on a small example how this algorithm works (see example 6).

Centering Theory

[9]An anaphor is an expression which refers to or stands for another word or words. In "I like ice-cream. It is good.", "it" is an anaphor standing for "ice-cream". A "do-anaphor" is special kind of anaphor which has the following form : To do so/it is/was NP. For example "I went hill climbing yesterday. To do it is a lot of fun." is a do-anaphora.

[10]Actor Focus

[11]Potential Focus List

[12]For the first sentence the focus is by default the first element of the PFL

Example 6 (adapted from Grosz et al. (1995), p.214)

a. John has been acting quite odd.
AF[10]*= John, PFL*[11]*= {John, acting odd}, Focus = John*[12]
b. He called up Mike yesterday.
AF = John, PFL = {Mike, calling up}, Focus = John (rule 7 and actor focus rule)
c. Mike was studying for his driver's test.
AF = Mike, PFL = {driver's test, Mike, studying}, Focus = Mike (rule 4 because "his" cannot specify John for semantic reasons)
d. He was annoyed by John's call.
AF = Mike, PFL ={Mike, call, being annoyed}, Focus = Mike (rule 7 and actor focus rule)

In this approach the actor focus is no longer distinguished from the discourse focus and the algorithms used to order the potential focus list and select the expected focus are modified to take this difference into account[13]. The potential focus list is now ordered by the following rules:

- A sentence subject is ranked higher than its objects.
- Objects are ranked higher than the other elements.

Several focus transitions between consecutive sentences are also defined (Grosz et al. 1995, p. 210):

Continuation the focus of sentence n+1 is the same as the focus of sentence n ($Cb(S_{n+1}) = Cb(S_n)$) and is also the most highly ranked element of sentence n+1 potential focus list ($Cb(S_{n+1}) = Cp(S_{n+1})$). In that case, it is likely to be the focus of sentence n+2.

Retaining the focus of sentence n+1 is the same as the focus of sentence n ($Cb(S_{n+1}) = Cb(S_n)$) but is not the most highly ranked element of sentence n+1 potential focus list ($Cb(S_{n+1}) \neq Cp(S_{n+1})$). In that case it is unlikely to be the focus of sentence n+2.

Shifting the focus of sentence n is not the same as that of sentence n+1 ($Cb(S_{n+1}) \neq Cb(S_n)$).

These transitions are summarised in table 4.

These transitions and the potential focus list (Cf) are used to select the expected focus by applying two new rules (Grosz et al. 1995, p. 214):

Rule 1 if any element of the potential focus list of sentence n is realised by a pronoun in sentence n+1, then the focus of sentence n+1 must also be realised by a pronoun.

[13]The terminology used in the centering theory differs from Sidner's theory. The focus of a sentence is called its backward-looking center (Cb). The PFL is called the set of forward-looking centers (Cf). The most highly ranked element of Cf is called the preferred forward-looking center (Cp). We will use Sidner's terminology when pointing out the similarities and differences between the theories.

	$Cb(S_{n+1}) = Cb(S_n)$	$Cb(S_{n+1}) \neq Cb(S_n)$
$Cb(S_{n+1}) = Cp(S_{n+1})$	Continuing	Shifting
$Cb(S_{n+1}) \neq Cp(S_{n+1})$	Retaining	

Table 4: Transitions in Centering Theory

Rule 2 Sequences of continuation transitions should be preferred over sequences of retaining transitions. These sequences should themselves be preferred over sequences of shifting transitions.

These rules constrain what the focus can be in several ways. In text understanding for example, rule 1 restricts the focus to the elements realised by pronouns (if any). In text generation, rule 2 provides constraints on what can be said next depending on the current focus (see section 3).

This approach is now applied on a small example (see example 7). Although the PFL is different from the one computed in example 6, this theory finds similar results to that of Sidner.

Example 7 (adapted from Grosz et al. (1995), p. 214)

a. John has been acting quite odd.
Cf = {John}, Cp = John
b. He called up Mike yesterday.
Cb = John, Cf = {John,Mike}, Cp = John
c. Mike was studying for his driver's test.
Cb = Mike, Cf = {Mike, driver's test}, Cp = Mike, Shifting
d. He was annoyed by John's call.
Cb = Mike, Cf ={Mike, call, John}, Cp = Mike, Continuation

Extension of the Centering Theory An extension of the centering theory has been proposed in Brennan et al. (1987) by remarking that the shifting transition can be refined into two transitions:

- A smooth-shift transition holds when the focus of sentence n+1 is not the same as the focus of sentence n ($Cb(S_{n+1}) \neq Cb(S_n)$) but is the most highly ranked element of sentence n+1 potential focus list ($Cb(S_{n+1}) = Cp(S_{n+1})$).
- A rough-shift transition holds when the focus of sentence n+1 is not the same as the focus of sentence n ($Cb(S_{n+1}) \neq Cb(S_n)$) and is not the

	$Cb(S_{n+1}) = Cb(S_n)$	$Cb(S_{n+1}) \neq Cb(S_n)$
$Cb(S_{n+1}) = Cp(S_{n+1})$	Continuing	Smooth-Shift
$Cb(S_{n+1}) \neq Cp(S_{n+1})$	Retaining	Rough-Shift

Table 5: Transitions in Centering Theory Extension

most highly ranked element of sentence n+1 potential focus list either ($Cb(S_{n+1}) \neq Cp(S_{n+1})$).

These transitions are summarised in table 5.

The transition ranking becomes (see rule 2 of the centering theory): Continuing > Retaining > Smooth-Shift > Rough-Shift. Making this finer distinction is useful for the resolution of pronoun references in some dialogues. For example, in the following dialogue, the centering theory cannot predict the references in the last sentence (see example 8) but the extended theory can (see example 9). The finer distinction also provides finer constraints for text generation.

Example 8 Centering Theory (adapted from Brennan et al. (1987), p. 157)

Brennan drives an Alfa Romeo.
Cf = {Brennan, Alfa Romeo}, Cp = Brennan
She drives too fast.
centering: Cb = Brennan, Cf = {Brennan}, Cp = Brennan, Continuing, she = Brennan
Friedman races her on weekends.
Cb = Brennan, Cf = {Friedman, Brennan, weekends}, Cp = Friedman, Retaining, her = Brennan
She often beats her.
Cb = Friedman, Cf = {?, ?}, Cp = ?, Shifting, she = ?, her = ?

Example 9 Extended Theory (from Brennan et al. (1987), p. 157)

Brennan drives an Alfa Romeo.
Cf = {Brennan, Alfa Romeo}, Cp = Brennan
She drives too fast.
Cb = Brennan, Cf = {Brennan}, Cp = Brennan, Continuing, she = Brennan
Friedman races her on weekends.
Cb = Brennan, Cf = {Friedman, Brennan, weekends}, Cp = Friedman, Retaining, her = Brennan
She often beats her.
Cb = Friedman, Cf = {Friedman, Brennan}, Cp = Friedman , Smooth-Shift, she = Friedman, her = Brennan

---

Common problems

There are a number of remaining problems with Sidner's theory and the centering theory (more are given in Walker et al. (1997b)):

- As they are purely based on sentence form, semantic incoherence is not detected. Example 2 (see section 2.a) is not considered as being particularly

Mode	Focus Level	Example
Pronominal	High	Your having called him up.
Name	Medium	Your having called Mark up.
Description	Low	Your having called your son up.
Implicit	Zero	Your having called.

Table 6: Focus levels (from Reichman (1981), p. 120)

incoherent. Detecting semantic incoherence would require one to reason about the meaning of the sentences. Sidner proposes to add a reasoning module with the sole purpose of accepting or rejecting the propositions made by the focusing algorithm (thus having a limited search space to explore), but such a module has not yet been implemented.

- The relation between global and local focus is not clearly explained by the theories (Grosz and Sidner 1995). The consequences of global shifts on the local focus remain unknown. Nor do we know the interaction between local focus pops, which are in fact non-local actions, and global focus.

These two problems are partly tackled by the following theories.

2.c.2 Based on dialogue moves: Reichman (Reichman 1985)

This theory completes Reichman's global focus approach (see section 2.b.5). Each focus space, which contains all the elements in global focus, is divided into four categories. These categories correspond to different levels of local focus. Depending on these levels, entities will have different realizations. The levels are described in table 6 along with their effect on realization (called mode in the table). The rules governing the attribution of focus level to the different entities are based on those of Sidner (1979). Here are the most important (Reichman 1981, p. 73):

- The subject of a sentence, of a there-insertion[14] clause, of a cleft, pseudo-cleft or topicalised clause[15], is assigned a high focus level.
- a non subject referenced by name is assigned a medium focus level.
- A non subject referenced by description is assigned a low focus level.
- An entity referenced by name after a previous reference by description is assigned a high focus level.
- If an entity's high focus level assignment is usurped by another entity then the old high focus constituent is reassigned a medium focus level.

An important difference though between this theory and Sidner's theory is the recognition of the global focus shifts' influence on the local focus. Changes in

[14] An "there-insertion" sentence has the following form : *There is* NP. For example, "There is a research group that works on knowledge-based systems" is a there-insertion sentence.

[15] A "topicalised" clause is a clause in which one constituent has been placed unexpectedly at the start. For example, "To John I gave the strawberries" is a topicalised clause with subject "John".

COCAINE				
	FUNCTION	get high	RESULT	get stoned
	PRICE	middle	ENABLE	find a dealer
	OWNER	nil	USER	nil
	AVAILABILITY	hard to get	PROBLEM	illegal
BUY				
	SCRIPT	shopping	ASSOC	nil
	RESULT	own	ACTOR	nil
	EFFECT	nil	OBJ	nil
	FREQUENCY	high	UNUSUAL	2

Table 7: Conversation/ associational categories (from Schank (1977), p. 433 and 438)

global focus level cause local focus changes. For example, "initial focus level assignments to entities in a digression space, which were previously mentioned in the context that this space interrupted, are carried over from that space" (Reichman 1985, p. 75) since the entities are still fresh in the mind of the dialogue participants. On the other hand, closing a space results in "the removal of all elements from focus, which is reflected in their reassignment to a zero focus level" (Reichman 1985, p. 82) since the elements in a closed focus space are not supposed to be spoken about again.

2.c.3 Based on semantics: Schank (Schank 1977)

The theory presented by Schank deals with semantic coherence between sentences. It is based on the dialogue participants' perception of the relations between topics.

As in Sidner's theory, the focus is defined by a relation between two sentences. A first sentence defines a focus set composed of the objects, persons, locations, actions, states and time mentioned in it. A second sentence modifies that set in two ways. First, a subset of the initial set, called the reduced old topic, is extracted. It represents the concepts that are common to the two sentences. Second, a new set is created. This second set, called the new topic, represents the new concepts that have been introduced in the second sentence. The speaker has then a choice, when continuing the discourse, to pick the new focus in either set. The speaker can maintain the old focus by choosing to discuss an element of the reduced old topic or shift the conversation to an element of the new topic. Even when maintaining the old focus, the speaker will have to acknowledge the new topic, since ignoring it would be certainly considered as very impolite. The acknowledgement can been done for example by rephrasing the old focus from the new focus point of view. In all cases, the new sentence provides a list of focuses that will be used to redefine the reduced old focus and new topic. The discourse can therefore be modelled by a binary tree where each node represents a sentence and the two branches represent the maintain focus choice and the focus shift choice respectively. If the conversation runs into problems, e.g., the topic becomes uninteresting, a speaker can jump back to a previous point of choice in the discourse and develop another branch of the tree. Such moves are usually marked in the dialogue by cue words (Schank 1977, p. 428).

Example 10 (from Schank (1977), p. 432)

I just got some cocaine.
Where did you find a dealer?

The main issue of the Schank theory is the process of selecting the new possible topics based on the reduced old topic, i.e., what are the shifts allowed in the dialogue ? Schank proposes to associate conversation/ associational categories with each concept. These categories describe the elements that are related to a concept and their level of interest. For example, "cocaine" and "buy" have the conversation/ associational categories given in table 7. These categories can be viewed as putting interesting items in implicit focus (see section 2.b.1). Based on these categories, the speaker can select an interesting shift in the conversation. In example 10, the speaker realises that a major problem in having some cocaine is to get it (the value of the availability category is "hard to get"). Therefore an interesting question is to find out where a dealer can be found ("find a dealer" being the value of the enable category).

On the other hand, example 11 seems odd since the actor of a buy action is usually not important (the value of the actor category for buy is nil).

Example 11 (from Schank (1977), p. 439)

I just bought a new car.
I know someone else who buys things.

The main problem with this theory is the necessity to have conversation/ associational categories for each possible concept that can be spoken about in a discourse. Moreover, the value of these categories may vary from one person to the other, depending on their cultural background.

2.c.4 Based on domain: McCoy & Cheng (McCoy and Cheng 1991)

The theory of McCoy and Cheng is somewhat similar to that of Schank. A first sentence puts an element into focus. Depending on the nature of that element (object, action, etc.) certain focus shifts are allowed as shown in table 8. The speaker selects one of these shifts and puts a new element in focus. Depending on the nature of this new element and on the previous dialogue history, new shifts will be allowed. This process continues for the rest of the discourse. In this approach, the focuses are organised as a tree, called the *focus tree*: nodes represent the focus of sentences and arcs represent shifts between them. The topic at the current point in the dialogue is the list of focuses from the node representing the current sentence up to the root node of the tree. That is to say: the topic is the current focus seen in the perspective of the preceding dialogue.

Type	Focus shift candidates
Object	• attributes of the object, • actions the object plays a prominent role in (e.g., is actor of)
Attribute	objects which have the attribute, more specific attribute
Setting	• objects involved in the setting • actions which typically occur in the setting
Action	• actor, object, etc., of the action – any participant role • purpose (goal) of action, next action in some sequence, sub-actions, specialisations of the action
Event	actions which can be grouped together into the event

Table 8: Candidate focus shifts (from McCoy and Cheng (1991), p. 112)

As in Schank's theory, the speaker has the possibility to stop developing the current node and instead expend a previous node. In that case the speaker will have to mark this move in the dialogue so that hearer can understand how the discourse is evolving, i.e., the marks needs to be more and more explicit as the dialogue deviates from its normal flow. Such marks may be cue words, tense shift, anaphora usage or pronoun shifting. McCoy and Cheng do not explain how cue words should be chosen for marking shifts (see section 4.d for related work on cue words).

2.c.5 Summary

Three classes of theories can be distinguished from the presentation (see table 1):

- Those based on dialogue moves and sentence construction.
- Those based on dialogue participant's beliefs.
- Those based on real world organisation.

Each kind of theory is incomplete by itself: for example, current theories based on sentence form need to be complemented by some semantic checker (see section 2.c.1) and theories based on semantic relations do not explain how the sentences are organised. Bridging the gap between the theories seems necessary to obtain a general theory of focus. Another issue that remains largely unexplained is the relation between local and global focus. Reichman's theory is an interesting attempt at integrating these two focuses. However, more work needs to be done to understand the influence between these two phenomena. Finally the relation between focus and belief is still a major problem because it requires participants to share the same frame of reference.

2.d Global and local focuses revisited

From the preceding presentations, we can notice several differences between local and global focus. They reflect the fact that the local focus is concerned with consecutive sentence coherence, while the global focus is concerned with discourse part coherence.

	Global focus	Local Focus
Deal with	High-level coherence	Consecutive sentences coherence
Contain	Several elements	One element
Shifts mainly constrained by	Dialogue goals and moves	Sentence structure
Shifts mainly marked by	Cue words and Anaphora	Anaphora

Table 9: Distinction between global focus and local focus

In particular:

- At any point, only one thing can be in local focus. On the other hand, the global focus may contain several things.
- Evolution of the local focus is restricted by sentence form and general knowledge constraints. How the global focus evolves is restricted by dialogue goals and moves.
- The local focus evolution is marked in the dialogue by linguistic phenomena such as anaphora. Evolution of the global focus is usually marked in the dialogue by cue words. Anaphora can also be used to mark a global shift since reference to a concept in a stacked focus space can cause this space to become active again.

The distinction between global and local focus is summarised in table 9.

3 Applications

In this section we describe some practical applications of the notion of focus. These applications are related to our research on dialogues in requirements elicitation systems. In this kind of dialogue, it is of the utmost importance to keep users interested in the dialogue while presenting what the system is doing. Using focus theories, it is possible to know what information should, or should not, be presented and when to present it.

For example, local focus theories (see section 2.c.1) can be used to improve text generation. Here are two examples demonstrating this.

Example 7 (adapted from Grosz et al. (1995), p. 214)

a. John has been acting quite odd.
b. He called up Mike yesterday.
c. Mike was studying for his driver's test.
d. He was annoyed by John's call.

Consider again example 7 (repeated here for convenience). Suppose that the first two sentences have already been produced and that we want to say that Mike was studying for his driver's test. Should "Mike" be realised by a pronoun or not, i.e., should we use sentence c as it is in example 7 or can we say "*He* was studying for his driver's test" ? If we chose the latter option, rule 1 of the centering theory would put a preference on interpreting "he" as being the focus of sentence b, i.e., John. Therefore, the realisation of Mike by a pronoun is impossible.

Now suppose that only the first sentence has already been produced and that we want to add to sentences b and c that John has been working very hard recently. How should we organise the text ? Since we want to speak about Mike in sentence c and the only sentence putting Mike in the potential focus list is sentence b, these two sentences should stay together. As a result we have to introduce our new information immediately, producing the following text:

a. John has been acting quite odd.
a'. He has been working hard recently and is quite tired.
b. He called up Mike yesterday.
c. Mike was studying for his driver's test.

Focus theories have been used in various applications (e.g., see Dale (1992); Huang (1994a,b); Mellish et al. (1998); Mittal et al. (1998); Reed and Long (1997a,b); Reed et al. (1997) for recent applications). In this section we first present two applications where the notion of focus is applied to the production of multi-sentential explanations. For this kind of explanation, focus rules need to be applied to produce coherent texts of high-quality. Otherwise, the explanations are no better than the series of disconnected sentence-long utterances of early systems. The two applications we review present information contained in databases to users. This domain is one in which several focus theories have been applied with relative success. It represents the main application domain of focus theories in natural language generation. Then we describe the application of the notion of focus to create collaborative agents with dialogue processing capabilities. These agents are used to help system users carrying out tasks. This application shows that the focus theories can be applied to more interactive domains than the production of explanations and gives some hints for the next generation of applications using focus.

Since the notions of global and local focus explain only partially the coherence of discourse, they are often used in conjunction with other theories of discourse. We pay particular attention to describing the integration of the notion of focus with such other techniques.

3.a Schema-based: McKeown (1985b)

The aim of this application is to answer users' questions about the content of a database. An example of an answer generated by the system is given in table 10. The system constrains the text generation by applying global and local focus rules when using predefined answer schemata.

More precisely, the system starts by selecting a set of schemata that can be used to answer the question of the user. Each schema corresponds to a strategy to answer the question. Examples of schemata are given in table 11.

At the same time, the system collects all the knowledge it has on the particular topic of the question by querying its database. Finally, the system finishes this first step by selecting a unique schema from the set of schemata based on

What is a ship ?
1. A ship is a water-going vehicle that travels on the surface. 2. Its surface-going capabilities are provided by the DB attributes DISPLACEMENT and DRAFT. 3. Other DB attributes of the ship include MAXIMUM SPEED, PROPULSION, FUEL (FUEL CAPACITY and FUEL TYPE), DIMENSIONS, SPEED DEPENDENT RANGE and OFFICIAL NAME. 4. The DOWNES, for example, has MAXIMUM SPEED of 29, PROPULSION of STMTURGRD, FUEL of 810 (FUEL CAPACITY) and BNKR (FUEL TYPE), DIMENSIONS of 25 (DRAFT), 46 (BEAM), and 438 (LENGTH) and SPEED DEPENDENT RANGE of 4200 (ECONOMIC RANGE) and 2200 (ENDURANCE RANGE).

Table 10: "What is a ship" (from McKeown (1985a), p. 30)

the available information from the database. Therefore, at the end of this step, the system has selected a strategy to present its answer and it has partitioned the knowledge that will be used for this presentation, i.e., it has set the global focus of the answer.

The second step consists in actually producing the answer in a coherent way, i.e., in respecting local focus constraints. In each schema, choice points enable the system to vary its presentation. Therefore, the system is able to select a presentation based on the available knowledge and on the preceding utterances. The rules used by McKeown to make this selection are the following:

- Changing the current focus to a potential focus as defined in Sidner's theory (see section 2.c.1) should be preferred to maintaining the focus. Indeed, if the potential focus is not presented immediately, the system would have to re-introduce it at a later point, which might be difficult (McKeown 1985b, p. 64).
- Maintaining the current focus should be preferred to shifting back to a stacked focus. Indeed, it is better to present a topic completely before going back to a previous one, otherwise the topic would have to be re-introduced (McKeown 1985b, p. 66).
- If several shifts are still possible, select the new topic which is most semantically related to the current focus (McKeown 1985b, p. 67).

We now show on the text of table 10 how the two steps apply. Suppose that the user asks what is a ship. The system, based on its available knowledge and

schema	use
identification	used to provide a definition
attributive	used to illustrate a particular point about a concept
compare and contrast	used to describe something by contrasting a positive point against a negative one

Table 11: Examples of schemata (from McKeown (1985a))

Identification schema
Identification (class & attribute/function) {Analogy/Constituency/Attributive}* Particular-illustration/Evidence+ ...
'{}' indicates optionality, '/' indicates alternatives, '+' indicates that the item may appear 1 to n times and '*' indicates that the item is optional and may appear 0 to n times.

Table 12: Partial identification schema (based on McKeown (1985a), p. 12 and 19–20)

presentation strategies, selects the identification schema. At the same time the knowledge that is relevant to the schema is selected from the database. This sets the global focus of the explanation: only the information in the relevant knowledge pool can be presented. The first two actions of the schema are the identification of the concept to explain and the choice between providing an analogy, a constituency, an attributive, a piece of evidence or a particular-illustration phrase (see table 12). The first action results in the first sentence of the answer describing a ship as a water-going vehicle. To carry out the second action, the system must apply the focus constraints since several focus shifts are possible. The particular-illustration move is ruled out because it requires information not included in the relevant knowledge pool. To choose between the remaining shifts, the system applies the local focus rules. All the available moves would lead to focus on the ship except for the evidence move which makes information on the "surface-going capabilities" salient[16]. Since the surface-going capabilities are a member of the potential focus list, this shift is preferred to the others and the system describes these capabilities in the second sentence of the answer. The system then goes through the rest of the schema applying the focus rules at all the choice points.

3.b Based on RST: Hovy and McCoy (1989)

Hovy uses RST relations to generate text (Hovy 1988, 1991). RST relations are relations that hold between parts of a text (Mann and Thompson 1987). Examples of such relations are: elaboration when one part gives some more information on the other, or evidence when one part gives reasons to believe the other. Hovy considers text generation as achieving a communication goal and the RST relations as operators that can be used to realise this goal. Each relation is defined by its intended effect and its constraints on the two parts of the discourse it links. These constraints can be seen as sub-goals that must be achieved for the relation to hold. By using a planner, the system can then find a plan that will achieve the communication goal and generate the corresponding discourse. This can be viewed as an improvement over McKeown's approach where predefined schemata are used to generate the text. In Hovy's approach,

[16]For each fact that could be chosen by a move, a default focus indicates what is most likely to be focused on.

RST Generated Navy text
Knox is en route to Sasebo. It is at 79N 18E heading SSW. It is C4. It will arrive 4/24, and will load for four days.
Joint RST and Focus Generated Navy text
With readiness C4, Knox is en route to Sasebo. It is at 79N 18E heading SSW. It will arrive 4/24 and will load for four days.

Table 13: Texts generated without and with focus constraints (from Hovy and McCoy (1989))

the schemata are created by the planner ,which offers more flexibility.

A problem with this approach is that achieving a discourse goal is only one part of generating a discourse. Coherence must be taken into account (Hovy and McCoy 1989; Hovy 1993). This is where focus rules can be used to constrain the text produced by the planner so that it is coherent. The focus approach taken by Hovy is that of McCoy and Cheng (see section 2.c.4): a focus tree is constructed in parallel to the RST relation tree created by the planner. Focus rules are then applied at two points in the generation of the text. They are first applied to limit the generation. By default, the planner tries to insert as much information as it can in the text. However, adding information can sometimes make the discourse incoherent. Focus rules are used to restrict the growth of the text by constraining the information that can be added. A new node is added to a node in the RST plan only if it is allowed by the corresponding node in the focus tree. The second point where focus rules are applied is in the ordering of the nodes in the plan. As explained at the beginning of this section, a RST relation links two parts of a text. These two parts can be ordered in one way or the other, e.g., an evidence can be presented before or after the phrase it supports[17]. Therefore, focus rules can be used to improve the text coherence by choosing between the possible text part orderings.

By applying focus rules to the generation of text, Hovy observed some improvement in the text quality. Two texts generated respectively without and with focus constraints are presented in table 13. Readiness is an attribute of being en route. It is therefore better, based on the focus tree, to introduce the readiness when speaking about the destination of Knox rather than separately.

A similar planning approach using RST relations was taken in the EES project (Moore and Swartout 1991; Paris 1991). Although focus rules were not used to constrain the text generation, in this project the notion of focus was employed to interpret follow-up questions to the text presented.

3.c Based on dialogue intentions: Collagen (Rich and Sidner 1997)

Collagen is an application-independent toolkit used to implement collaborative interface agents with discourse processing capabilities. These agents help the users of a system by conversing with them and studying their actions. The agents can act themselves directly on the system. They have access to a database

[17]This is not completely true. RST relations place preferences on the order in which the two parts should be presented (Mann and Thompson 1987, p. 16)

USER: Propose scheduling a trip.
AGENT: OK.
(User identifying route of scheduling a trip as San Francisco to Dallas to Boston)
USER: Add San Francisco to the route.
USER: Add Dallas to the route, allowing 97 itineraries.
USER: Add Boston to the route, allowing 100+ itineraries.
AGENT: How should a trip on the route be scheduled ?
USER: Propose scheduling on the route via working forward.
USER: What should be done next ?
AGENT: Propose you work on San Francisco to Dallas leg.
...

Table 14: Collagen agent dialogue (adapted from Rich and Sidner (1997))

of application-dependent plans that can be applied to the system to carry out some goals. We focus here on their dialogue capabilities.

Collagen agents decompose a dialogue into segments according to the Grosz and Sidner global focus theory (see section 2.b.3). Each segment is therefore associated with its intention and the entities in focus during its realization. Segments are put into a stack when they begin and are popped when they finish. Based on this decomposition and on the plans they have access to, agents propose actions to be performed. To carry out this task, the discourse generation algorithm produces an agenda of prioritised actions that would contribute to the current discourse segment purpose (DSP). The priority is computed thanks to the focus of attention (among other factors). For example if objects used in the action considered are already known, this action will receive a high-priority. Collagen agents therefore ensure that the action proposed is always coherent with the current user preoccupations.

Collagen agents also make use of the same mechanism to reduce the need for natural language understanding. Rather than tackling unrestricted natural language inputs, the agents propose a dynamic set of inputs to the users. This set is computed from the current discourse context and the possible actions that the user can do in the given situation. The agent-user interactions are therefore adapted to the current needs of the user and are easy to understand since they follow dialogue rules. This approach, although less powerful than having a fully-fledged natural language understanding module, seems cost-effective and well-adpated with the idea of having a focus toolkit ready to be used in a variety of applications.

An example of agent-user communication is given in table 14. The agent and the user are trying to schedule a trip using an application displaying a map and computing the possible itineraries.

3.d Summary

The main observation of this section is that, in limited forms, the notion of focus can be implemented. It can provide useful rules to select how information is presented. It can also be integrated with other theories or techniques such as RST relations. However a second observation is that in order to work every implementation requires ad hoc solutions to un-resolved problems in its focus mechanism. This is for example the case with McKeown's additional rules to constrain the focus shifts by preferring a potential focus over the current focus and the current focus over stacked focuses (see section 3.a)[18]. This is due to the fact that focus theories have been mainly created for text interpretation and not text generation. Generation requires more constraints than interpretation because the system has to make a choice about the possible focus shifts rather than just recognising them.

4 Remaining issues

As we have seen in the previous section, the theories of focus are not complete. Some ad hoc solutions must be implemented to obtain working systems. For example, presentation strategies such as those used by McKeown are needed to supplement focus theories.

In this section we review problems that remain to be solved to obtain a complete and applicable theory of focus. The integration with other discourse theories is particularly important. Another problem is the actual implementation of the theories: choices have to be made to obtain computable models of the theories.

4.a How many relations?

Many focus theories are based on relations such as dominance and satisfaction-precedence (see section 2.b.3), dialogue moves (see section 2.b.5) or semantic links (see section 2.c.3). An issue that must be considered is the evaluation of the advantages and drawbacks of having very few or very many relations (Hovy and Maier 1992; Knott and Dale 1996). Adding relations allows finer grained theories of focus. For example, the richer set of relations in Reichman's theory allows a more precise classification of cue words than in Grosz's theory. However too many additions can result in superfluous complexity. For example, there is not much point in differentiating many relations if they are not expressed by different means in the dialogue. More research on this issue is needed.

4.b User modelling and focus

Theories of focus are only one aspect to be considered in text generation. Another important one is user modelling. Not only should the text be coherent but it also should be of interest to its readers (Kok 1991; Kobsa 1993). A user model would allow the system to produce the right amount of information expressed with the right vocabulary. For example, a user model could restrict the global focus by eliminating the knowledge that a user already knows. This would result in an answer centred on the points the user is ignorant about. A user model could also complement the local focus rules in choosing between competing focus shifts, again depending on what the user is interested in.

[18]These rules could be considered as not just ad hoc rules but as general principles. (They are indeed used in various systems.) However we consider them as ad hoc because they are not yet part of a theory of focus.

User modelling can also be used to take into account the users' beliefs or the perspective they take on a topic (McCoy 1988). These aspects have a strong influence on the focus (see section 2.b.1). Their integration is therefore necessary to produce high-quality texts.

4.c Knowledge representation and focus

An issue related to the question of the number of relations is the question whether a dedicated knowledge representation should be used to apply the focus rules. If the relations used by the rules are equivalent or can be deduced from the representation used to represent the data, a unique representation is sufficient. This is for example the case for McKeown's database interface (see section 3.a). In this application the information in the database is sufficient to determine the focus relation between the data. If this is not the case, a separate representation needs to be created. This has some drawbacks, such as the replication of information, but also some advantages beyond the capability of applying the focus rules. Indeed, it is probable that the new representation based on a linguistic theory will be compatible with some other existing linguistic representations such as an upper-model[19] (Bateman 1990, 1991). This would help in the generation of the text (McKeown and Swartout 1988, p. 25–32). The need for a dedicated representation depends on the data represented as well as the focus theory used. Evaluation of the advantages and drawbacks of having a dedicated knowledge representation is therefore needed.

4.d Cue words

Cue words have been recognised as playing a major role in dialogue (see sections 2.b.1 and 2.b.5). In particular, they help the dialogue interpretation process by pointing out focus shifts. However, little work has been done in determining when cue words should be used to mark focus shifts and how to select them. Cohen (Cohen 1984, 1987) proposes that cue words are used:

- to specifically re-direct the attention of the hearer to an earlier part of the dialogue. An example of such a cue word is *returning to*.
- to indicate how the following utterances are related to the current focus. Examples include *first* or *second* to introduce a parallel focus and *in particular* to introduce a more specific subject.

This is confirmed with the cue words proposed by the different focus theories.

Cohen then states that these two types of cue words can be used either to ease the dialogue interpretation in which case they are optional or to allow an interpretation that would be otherwise impossible in which case they are mandatory. An interpretation is considered to be impossible without cue words if another one exists which requires less effort. (The calculation of the amount of effort needed to interpret a dialogue depends on the dialogue theory used.) The further the intended interpretation is from the least costly one, the stronger the use of cue words needs to be.

[19] An upper-model is an ontology based on linguistic principles

This is similar to the proposition made by McCoy and Cheng (see section 2.c.4) who propose to use cue words when the focus does not closely follow the path predicted by their rules.

This issue was also alluded to by Linde and Goguen (Linde and Goguen 1978). In particular they propose to classify the strength of cue words on the following criteria:

- The more a cue word looks like a sentence, i.e., the more independent it is, the stronger it is.
- Cue words in front of sentences are stronger than those placed within sentences.
- The longer and more explicit a cue word is, the stronger it is.

For example, the cue in the first sentence of table 15 is stronger than the cue in the second. It can then be used to mark a bigger deviation from the expected flow of discourse.

Strong cue word: On the other hand, we have to...
Weak cue word: We also have to...

Table 15: Cue word strength (Linde and Goguen 1978, p. 238)

In spite of these pieces of work, the use of cue words, especially in text generation, is still not fully understood. Systematic work like Moser and Moore (1996) and Knott and Dale (1996) may shed light on this issue.

4.e Internal and external phenomena

Most theories of global focus are based on phenomena that are "external" to the dialogue: task structure, speaker's intentions. Much less attention has been paid to the influence of "internal" phenomena that can influence the dialogue. For example, if I say, during the elicitation of information on a research group, "A research group involves its researchers, its projects and its publications", there is an expectation that the dialogue will deal with these three points. More generally, the mention of a list of elements in a dialogue seems to cause a shift to each element in the rest of the dialogue[20]. More research is also needed to determine what are the internal phenomena that can take place in the dialogue, their effects and their relations with external factors.

4.f How is the information put into focus?

As mentioned in section 2.b.1, the global focus usually contains elements present in the discourse as well as implicit elements. The latter are used to determine if a new phrase appearing in the text is related to the current focus or not.

[20] We have noticed this phenomenon in elicitation dialogues about research groups. Sometimes, when no information about an element has to be elicited, the element is just put into focus before being immediately popped ("Well, in fact we do not need to speak about this"). This shift seems to be necessary to fulfil the expectation set by the mention of the element in the list.

If it is the global focus remains unchanged. If not, the global focus is shifted to the new topic. Therefore an important problem is to know what should be included in implicit focus. Including too few elements would result in numerous unnecessary focus shifts whereas including too many elements would preclude real focus shifts. Some theories propose domain and culture dependent solutions, e.g., Schank's conversation/ associational categories, while others propose domain-independent rules, e.g., Cheng and McCoy's candidate focus shifts. More research needs to be carried out to determine what should be considered in implicit focus. This would also give some insights on the local focus and global focus boundary: when is a shift local and when is it global ? Moreover, it is important to know how to interpret partial or incomplete references. Should they be considered as referring to elements in (explicit or implicit) focus or should they be considered as being new elements, possibly shifting the focus ? This problem is similar to deciding what is new or given in a dialogue (Clark and Haviland 1977). Clark proposes that a piece of information is considered given if the listener can build "an inferential bridge [to it] from something already known". Clark does not have however firm suggestions on how a bridge is built. More research needs to be carried out on this topic too.

4.g How does the focus evolve?

Quite often, focus rules restrict what can be said next but a choice remains to be made between the possible developments. This is why additional rules have to be created when generating text. McKeown uses additional focus constraints and schemata (see section 3.a) and Hovy uses RST relations (see section 3.b). Taking into account the intentions of the system and those of the users as well as who has the initiative in the dialogue could also restrict the generation process (Smith 1996). Research should be carried out to determine if a general theory can explain these different approaches and point out their commonalities.

5 Conclusion

In this paper, we have presented the notion of focus. We have described various theories explaining this notion. We have classified them into two groups depending on the level of focus they explain and we have presented the information they use to predict the focus. Then we presented actual applications of the theories in two domains: presentation of database data and cooperative agent dialogue. Finally we described remaining issues in the understanding of the notion of focus. To conclude, we can summarise the current state of the art and the expected evolution of focus use in a few points:

- Theories of focus explain a large number of focus related linguistic phenomena. Problems, such as the relation between local focus and global focus, remain but the theories are sufficiently mature to allow actual applications of the notion.
- The application of the notion of focus raises engineering problems: how to represent the information used by the focus rules, how many relations are needed, etc. Current approaches have developed ad hoc solutions to these problems. As experience is gained on the use of the notion of

focus, guidelines should appear and the development of applications may be easier.

- The notion of focus will certainly be used in conjunction with other techniques. It may become part of a set of tools used to produce high-quality interaction with users. Such an approach is exemplified by the Collagen toolkit that implements focus rules and is used to create user-system interfaces.

References

John A. Bateman. Upper modeling: A general organization of knowledge for natural language processing. In *Proceedings of the Workshop on Standards for Knowledge Representation Systems, Santa Barbara, USA*, March 1990.

John A. Bateman. The theoretical status of ontologies in natural language processing. In *Proceedings of the workshop on Text Representation and Domain Modelling – Ideas from Linguistics and AI, Berlin, Germany*, October 1991.

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

Herbert H. Clark and Susan E. Haviland. Comprehension and the given–new contract. In Roy O. Freedle, editor, *Discourse Production and Comprehension*, chapter 1, pages 1–40. Ablex publishing corporation, Norwood, New Jersey, USA, 1977.

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

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

Robert Dale. *Generating referring expressions: constructing descriptions in a domain of objects and processes.* The MIT Press, 1992.

H. P. Grice. Logic and conversation. In Peter Cole and Jerry Morgan, editors, *Speech Acts*, volume 3 of *Syntax and Semantics*, pages 41–58. Academic Press, New York, 1975.

Barbara J. Grosz. The representation and use of focus in dialogue understanding. Technical Report 151, SRI International, July 1977.

Barbara J. Grosz. Focusing and description in natural language dialogues. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, *Elements of discourse understanding*, chapter 3, pages 84–105. Cambridge University Press, 1981.

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 Karen E. Lochbaum. Intersegment relations and intentions. In Donia Scott and Eduard Hovy, editors, *NATO Workshop: Burning Issues in Discourse, Maratea, Italy*, pages 20–26, April 1993.

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

Barbara J. Grosz and Candace Lee Sidner. Plans for discourse. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, chapter 20, pages 417–444. The MIT Press, 1990.

Barbara J. Grosz and Candace Lee Sidner. Lost intuitions and forgotten intentions. Technical Report 95-02, Lotus, 1995.

Raymonde Guindon. Anaphora resolution: short-term memory and focusing. In *Proceedings of the* 23^{rd} *annual meeting of the Association for Computational Linguistics, Chicago, Illinois, USA*, pages 218–227, July 1985.

Eduard H. Hovy. Planning coherent multisentential text. In *Proceedings of the* 26^{th} *annual conference of the Association for Computational Linguistics, Buffalo, New York, USA*, pages 163–169, June 1988.

Eduard H. Hovy. Approaches to the planning of coherent text. In Paris et al. (1991), chapter 3, pages 83–102.

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

Eduard H. Hovy and Elisabeth Maier. Parsimonious or profligate: How many and which discourse structure relations? Technical Report RR-93-373, University of Southern California, Information Science Institute, August 1992.

Eduard H. Hovy and Kathleen F. McCoy. Focusing your RST: A step toward generating coherent multisentential text. In *Proceedings of the eleventh annual conference of the Cognitive Science Society, Ann Arbor, Michigan, USA*, pages 667–674, August 1989.

Xiaorong Huang. Planning argumentative texts. In *Proceedings of the* 15^{th} *International Conference on Computational Linguistics, Kyoto, Japan*, pages 329–333, August 1994a.

Xiaorong Huang. Planning reference choices for argumentative texts. In *Proceedings of the* 7^{th} *International Workshop on Natural Language Generation, Kennebunkport, Maine, USA*, pages 145–152, June 1994b.

Alistair Knott and Robert Dale. Choosing a set of coherence relations for text-generation: a data-driven approach. In Giovanni Adorni and Michael Zock, editors, *Trends in natural language generation: an artificial intelligence perspective*, number 1036 in Lecture Notes in Artificial Intelligence, pages 47–67. Springer-Verlag, 1996.

Alfred Kobsa. User modeling: Recent work, prospects and hazards. In M. Schneider-Hufschmidt, T. Kühme, and U. Malinowski, editors, *Adaptive User Interfaces: Principles and Practise*. Elsevier, 1993.

André J. Kok. A review and synthesis of user modelling in intelligent systems. *The Knowledge Engineering Review*, 6(1):21–47, 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*, pages 190–196. IEEE Computer Press, April 1998.

Willem J.M. Levelt. *Speaking: from intention to articulation*. The MIT Press, 1989.

Charlotte Linde. Focus of attention and the choice of pronouns in discourse. In Talmy Givón, editor, *Discourse and Syntax*, volume 12 of *Syntax and Semantics*, pages 337–354. Academic Press, New York, 1979.

Charlotte Linde and J.A. Goguen. Structure of planning discourse. *Journal of social biological structures*, 1:219–251, 1978.

William C. Mann, Madeline Bates, Barbara J. Grosz, David D. McDonald, Kathleen R. McKeown, and William R. Swartout. Text generation: The state of the art and the literature. Technical Report RR-81-101, University of Southern California, Information Science Institute, December 1981.

William C. Mann and Sandra A. Thompson. Rhetorical structure theory: A theory of text organization. Technical Report RS-87-190, University of Southern California, Information Science Institute, June 1987.

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

Kathleen F. McCoy and Jeannette Cheng. Focus of attention: Constraining what can be said next. In Paris et al. (1991), chapter 4, pages 103–124.

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

Kathleen R. McKeown. *Text generation – Using discourse strategies and focus constraints to generate natural language text*. Studies in Natural Language Processing. Cambridge University Press, 1985b.

Kathleen R. McKeown and William R. Swartout. State of the art. In Michael Zock and Gérard Sabah, editors, *Advances in Natural Language Generation – An Interdisciplinary Perspective*, volume 1, chapter 1, pages 1–51. Pinter Publishers, London, 1988.

Chris Mellish, Mike O'Donnell, John Oberlander, and Alistair Knott. An architecture for opportunistic text generation. In *Proceedings of the nineth International Workshop on Natural Language Generation, Niagara-on-the-Lake, Ontario, Canada*, August 1998. (to be published).

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*, 1998. (to be published).

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

Johanna D. Moore and William R. Swartout. A reactive approach to explanation: Taking the user's feedback into account. In Paris et al. (1991), chapter 1, pages 3–48.

Megan G. Moser and Johanna D. Moore. Investigating cue selection and placement in tutorial discourse. In *Proceedings of the* 33^{rd} *annual meeting of the Association for Computational Linguistics, Cambridge, Massachusetts, USA*, pages 130–135, June 1995.

Megan G. Moser and Johanna D. Moore. Toward a synthesis of two accounts of dialogue structure. *Computational Linguistics*, 22(3):409–420, 1996.

Cécile L. Paris. Generation and explanation: Building an explanation facility for the explanaible expert systems framework. In Paris et al. (1991), chapter 2, pages 49–82.

Cécile L. Paris, William R. Swartout, and William C. Mann, editors. *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Kluwer, 1991.

Cécile L. Paris, Michael R. Wick, and William B. Thompson. The line of reasoning versus the line of explanation. In Michael R. Wick, Cécile L. Paris, William R. Swartout, and William B. Thompson, editors, *Proceedings of the AAAI Workshop on Explanation, Saint Paul, Minnesota, USA*. American Association for Artificial Intelligence, 1988.

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*, 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*, August 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*, March 1997b.

Chris A. Reed, Derek P. Long, and Maria Fox. Context and focusing in argumentative dialogue planning. In *Proceedings of the International Conference on Context, Rio de Janeiro, Brazil*, pages 88–99, February 1997.

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.* MIT Press, Cambridge, MA, USA, 1985.

Charles Rich and Candace Lee Sidner. Collagen: When agents collaborate with people. In *Proceedings of the First Conference on Autonomous Agents, Marina del Rey, California, USA*, February 1997.

Colette Rolland. Modeling the requirements engineering process. In *Proceedings of the third European-Japanese Seminar on Information Modelling and Knowledge Bases, Budapest, Hungary*, June 1993.

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*, September 1994.

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

Candace Lee Sidner. Towards a computational theory of definite anaphora comprehension in english discourse. Technical Report TR-357, MIT, 1979.

Candace Lee Sidner. Focusing and discourse. *Discourse Processes*, 6:107–130, 1983.

Ronnie W. Smith. Initiative-dependent features of human-computer dialogs in a task-assistance domain. In *Proceedings of the Energy Information Management Conference*, pages 41–48, 1996.

Richard W. Southwick. Topic explanation in expert systems. In B. Kelly and A. Rector, editors, *Research and Development in Expert Systems V*, pages 47–57. Cambridge University Press, 1989.

Richard W. Southwick. Explaining reasoning: an overview of explanation in knowledge-based systems. *The Knowledge Engineering Review*, 6(1):1–19, 1991.

Marilyn A. Walker. *Informational Redundancy and Resource Bounds in Dialogue.* PhD thesis, University of Pennsylvania, December 1993.

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

Marilyn A. Walker. Limited attention and discourse structure. *Computational Linguistics*, 22(2):255–264, 1996b.

Marilyn A. Walker. Centering, anaphora resolution, and discourse structure. In Walker et al. (1997a).

Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors. *Centering in discourse*. Oxford University Press, 1997a.

Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince. Centering in naturally-occuring discourse: An overview. In *Centering in discourse* Walker et al. (1997a).

Michael R. Wick. Explanation as a primary task in problem solving. *The Knowledge Engineering Review*, 9(1):78–82, 1994.