Choosing Rhetorical Structures to Plan Instructional Texts

Leila Kosseim*
Druide informatique inc.
5515, chemin de la Côte-Saint-Luc
Montréal (Québec), Canada, H3X 2C6

Guy Lapalme
Département d'informatique et de recherche opérationnelle
Université de Montréal, CP 6128, Succ. Centre Ville
Montréal (Québec) Canada, H3C 3J7
e-mail:{kosseim, lapalme}@iro.umontreal.ca

May 1, 1998

Abstract

This paper discusses a fundamental problem in text generation: how to select the content to communicate and how to present it in a coherent text. In this research, we set out to determine automatically the semantic content and the rhetorical structure of French instructional texts. To do so, we performed a corpus analysis in order to capture natural linguistic phenomena. From the corpus analysis we determined 9 *semantic carriers* typically communicated in instructional texts and 7 *rhetorical relations* used to present them. From this analysis, we then developed 2 sets of heuristics: *content heuristics* that determine what semantic carriers should be communicated in the text and *presentation heuristics* that determine how the semantic carriers should be organized rhetorically in order to create a coherent and natural text. These heuristics are based on 5 types of constraints: conceptual, semantic, rhetorical, pragmatic and intentional constraints.

In order to verify these heuristics, we developed the SPIN system which performs all steps of text generation but focuses on the determination of the content and the rhetorical structures.

keywords: text generation, text planning, Rhetorical Structure Theory, instructional texts

*work done at the Université de Montréal

Introduction

This paper discusses a fundamental problem in text generation: the determination of the semantic content to be communicated and its organization into a coherent text. This research has studied instructional texts, an area of great potential for natural language generation; these texts are widely available and understandable by many readers, they are usually well structured. Compared to other narrative texts, they can be represented "objectively" by relations between states and operations which could be built by an automatic planner. This view of instructional texts as a hierarchy of operations and sub-operations was confirmed by psychological studies [Dixon, 1987, Dixon et al., 1988, Donin et al., 1992].

We analyzed a corpus of French instructional texts in order to capture natural linguistic phenomena without over-simplifying them. Although our study was done on French texts, we believe that our results are also applicable in English in the sublanguage of instructional texts:

> *One is therefore drawn to conclude that English and French technical texts show the strongest parallels because the text purpose is more similar here than in descriptive texts. Weather reports, recipes and aviation manuals, which show the strongest parallels, all have very well-defined test purpose.*[Kittredge, 1982, p. 135]

This can also be confirmed in this paper where we have given a literal translation (in this type) after each example from our corpus. These translations are almost always acceptable English instructions. On the other hand, other research have shown preferences in rhetorical relations according to the language studied [Delin et al., 1996].

Through the corpus analysis, we identified 9 meanings (which we call *semantic carriers*) and 7 rhetorical relations that are used to present them. Semantic carriers and rhetorical relations make up the larger part of instructional texts. We then developed a set of content heuristics to select the semantic carriers to be communicated and a set of presentation heuristics to select the appropriate rhetorical relations. Both sets of heuristics are based on conceptual, semantic, rhetorical, and pragmatic constraints, and also take into account the knowledge and intentions of the reader. We implemented these heuristics into an automatic text generator called SPIN[1]. SPIN performs all steps of the text generation process from the conceptual determination to the lexico-typographical choices; however, the expertise of SPIN is in the semantic and the rhetorical selection.

Section 1 of this paper discusses the advantages of separating the semantic and rhetorical levels. Section 2 presents the corpus used in our analysis. Section 3 discusses

[1] SPIN stands for "Système de Planification d'INstructions".

the semantic carriers identified through the corpus analysis. Section 4 shows the presentation heuristics. Finally, an overview of the SPIN system and its results are presented and evaluated.

1 Separating the Semantic and the Rhetorical Levels

In our model of generation of instructional texts, we first determine the sequence of operations to execute in order to reach a particular goal; i.e. the conceptual representation of the instructions. We then choose which information will be given in text and which one will be left implicit: we call this the semantic content of the text. We then choose the rhetorical structure of the text. This separation of the semantic and rhetorical levels allows a better rhetorical diversity than in RST or schema based approaches which combine content determination and rhetorical structure determination in a single step. This distinction between semantic and rhetoric concerns is much more important for us than the distinction between their processes, for example Moore and Paris [Moore and Paris, 1993] also separate these concerns by allowing both semantic and rhetorical operations in their planners.

In [Delin et al., 1993], it has been demonstrated that in multilingual instructions, the same information can be conveyed using different rhetorical structures depending on the language of communication. More generally, many researchers have argued that within a unilingual environment the mapping between the semantic and the rhetorical levels is many-to-many [Moore and Pollack, 1992, Korelsky and Kittredge, 1993]. It thus becomes necessary to have linguistically motivated guidelines on how to organize a text's content; that is, how to map a semantic representation onto the most appropriate rhetorical structure.

In French instructional texts, sentences like the following appear quite often:

(1) a. Brancher le cordon d'alimentation du magnétoscope dans une prise secteur 120V et appuyer sur l'interrupteur POWER. *Le voyant* POWER *s'allume et l'horloge commence à clignoter.*
Plug the electrical cord of the video-tape recorder in a 120V outlet and press on the POWER button. The POWER light is turned on and the clock starts to blink.

b. Revisser l'écrou capuchon sur la lyre *pour ne pas le perdre.*
Screw the nut-cap on the lamp-shade holder so that you do not lose the it.

c. *Vous pouvez voir le niveau de volume* en observant la barre rouge sur la gamme de 15 barres affichées sur l'écran.

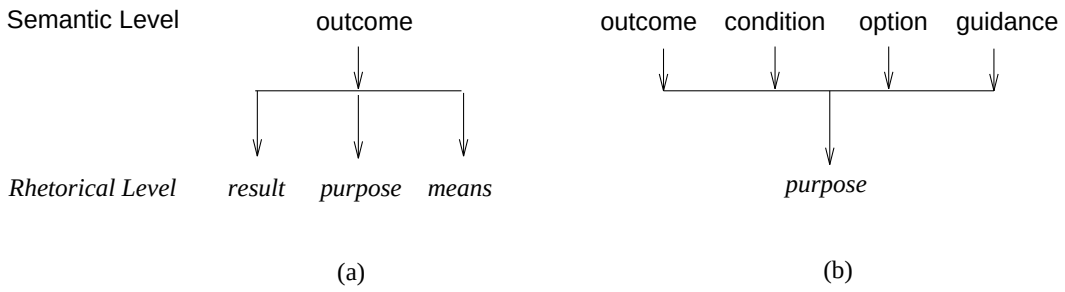


Figure 1: Many-to-many mapping between the semantic and the rhetorical levels

> You can see the volume by observing the red bar on the 15-bar scale displayed on the screen.

In these three examples, the same semantic information is conveyed by the expression in italics: it expresses the OUTCOME[2] of some action. However, these OUTCOMEs are communicated through different rhetorical relations. In the case of (1a), a *result* is used; in (1b) a *purpose* is used; and in (1c) it is the nucleus of a *means* relation. This mapping is shown in Figure 1a.

Inversely, one rhetorical relation can be used to convey different semantic information. For example, the relation of *purpose* in (1b) communicates the OUTCOME of an action; while in (2a) it communicates a CONDITION on an action; in (2b), it communicates the OPTIONAL NATURE of an action; and finally in (2c) it communicates a GUIDANCE[3] on how to perform an action. This is illustrated in Figure 1b.

(2) a. *Pour [vérifier] un commutateur ordinaire [...]*, touchez la vis de la borne de cuivre avec la pince du vérificateur.
To [check] an ordinary switch [...], touch the screw of the copper terminal with the clip of the tester.

b. Tirer la roue et le pneu; *pour vous aider*, poussez fermement le flanc du pneu avec votre pied.
Pull the wheel and the tire; to ease the task, firmly press on the side of the tire with your foot.

c. Tourner cette touche à droite et à gauche *pour minimiser les parasites*.
Turn this knob clockwise and counter-clockwise to minimize interference.

[2]In this paper, semantic carriers are written using SMALL CAPS while rhetorical relations are written using *italics*

[3]A GUIDANCE is equivalent to [Di Eugenio and Webber, 1995]'s notion of pragmatic overloading.

These examples illustrate the many to many mappings between semantic and rhetoric levels. These mappings are an important problem that should be dealt within a text generator. Before describing our approach, we first describe the corpus study that gave us the necessary cues to make appropriate choices of mappings.

2 The Corpus

To describe the natural process of generating instructions, we have taken a corpus of "correct" texts and analyzed it both at the semantic and the rhetorical levels. In choosing texts for our corpus study, we only considered "correct" instructions, i.e. instructions that we considered clear and understandable. Some manuals are badly structured or show a bad choice of words. We only considered original French texts or translations that did not seem "biased" by their original language. All texts were analyzed by the first author so there are no cross-reader dependencies. We collected instructional texts from 15 different sources from every-day appliances and how-to books. We analyzed only the procedural parts of these texts. The corpus is made up of 79 procedures ($\approx$13,300 words) that have different communicative goals, domains and intended readers. When one procedure was divided into sub-procedures, we counted as one a procedure for reaching a single goal (e.g. oil changing in a car). If the procedure had a less precise goal (e.g. using a remote control) then each procedure was counted separately. Table 1 lists the texts of the corpus along with some textual, conceptual, lexical and typographical characteristics. References to the original texts are given in [Kosseim, 1995]. The corpus was divided into 3 classes according to the communicative goal of the texts.

Execution texts are characterized by their simplicity and are aimed at an immediate execution. These texts are typically short, have a low level of lexical specialization, have strong typographical cues (ex. bullets, alert icons, ...) and are characterized by their conceptual simplicity, they required little judgment and simple instruments.

Comprehension texts are characterized by their complexity and are aimed at an eventual execution; their goal is to explain, not to tell. These texts have longer procedures, have more specialized terms, have less typographical cues and generally require more judgment and more instruments.

Hybrid texts exhibit characteristics of both execution and comprehension texts.

The corpus is rather varied with respect to the discourse domains and the communicative goals. The analysis of a less diversified corpus would certainly have permitted of more precise results, but their applicability to other domains would not have been

Type	Domain	Length (words)	Lexical specialization	Typo-graphical cues	Required judgment	Required instruments
Execu-tion	emergency respiratory care	50	low	strong	average	low
	cooking recipes	370	low	strong	low	average
	assembling a dresser	270	low	strong	average	average
	using a video tape recorder	55	low	strong	low	low
	using a television	70	low	strong	low	low
	fixing electrical appliances	100	low	strong	average	low
	car maintenance	100	low	strong	low	strong
Compre-hension	wine-making	250	strong	average	strong	strong
	glass painting	550	average	low	strong	strong
	glass blowing	120	strong	low	average	strong
Hybrid	Mazda car manual	120	low	strong	low	low
	techniques of rotin	255	average	average	average	average
	techniques of photography	200	average	average	average	average
	organizing a hard drive	200	average	average	low	low
	restoration of antiques	820	average	low	average	average

Table 1: Characteristics of the corpus

possible. Following [Mellish, 1988], our aim is the define general rules for the generation of instructions in all domains, then particular rules for adapting the text to a specific domain and reader.

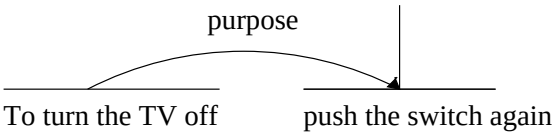


Figure 2: An RST relation

Instructional texts are characterized by stereotypical relations between parts of the text. Like [Rösner and Stede, 1992, Vander Linden, 1993], we used Rhetorical Structure Theory[Mann and Thompson, 1988] (RST) as a basis for the analysis of our corpus. RST was developed to identify semantic and pragmatic relations between adjacent portions of text and is defined in terms of

relations between a main part, the *nucleus* and an auxiliary part the *satellite*

schemata that specify the structural composition of a text

Figure 2 shows the schema corresponding to a *purpose* relation where the satellite *To turn the TV off* is linked to the nucleus *push the button again*. In RST, the horizontal lines span the linguistic expression, the vertical line show the nucleus and the arc between them indicates a relation between them. We use RST both as a descriptive tool for analyzing texts and as a constructive tool in the selection of the textual structure.

3 The Semantic Content

By viewing the conceptual representation of an instructional text as a hierarchy of plans, we realized that not all the information available in the conceptual representation is given in the text. Therefore, the goal of the semantic analysis of our corpus was to find out:

1. What type of information is typically communicated in instructional texts.
2. What constraints influence the communication (or non-communication) of this information.

Our analysis was done in terms of semantic carriers. Semantic carriers identify the meaning of textual expressions[4]. They are called *carriers* because they are elements of meaning that are directly mapped onto the satellite of a RST relation [Mann and Thompson, 1988]. But some carriers refer to relations whose nucleus and satellite will be determined at the rhetorical level when the most appropriate relation is chosen. To define the semantic carriers, we have been inspired by the semantic relations of *type generation*, *instance generation* and *enablement* of [Goldman, 1970]. [Delin et al., 1996] show how these relations can be used as the basis for the generation of multilingual instructional texts. These relations, however, do not cover the range of semantic entities found in instructional texts; thus, we have developed more specific entities particular to this discourse genre. In our corpus, one non-procedural and eight procedural semantic carriers have been identified. We give an informal definition and examples of these relations:

Non procedural relation

ATTRIBUTES OF OBJECTS do not participate in the instructions but merely give some background information to the reader

Procedural relations

REQUIRED OPERATIONS describe the steps to follow in order to execute the instructions

MATERIAL CONDITIONS describe a situation when the operations are to be followed, see (2a).

OUTCOMES show what should happen after an operation has been executed

(3) Appuyer sur la touche PLAY. *La lecture normale apparaît.*
Touch the play button. Normal reading appears.

[4]The granularity of the analysis is consistent with a typical RST analysis.

OPTIONS give operations that may be done but are not necessary, see (2b)

GUIDANCES show how or why an operation is executed while guiding the execution, see (2c)

CO-TEMPORAL OPERATIONS indicate that more than one operation have to be done at the same time

(4) [...] *effleurer la surface du verre, tout en le faisant tourner.*
touch lightly the glass surface, while making it turn

OPERATION PREVENTIONS are negative sentences indicating operations that should not be performed. They are realized grammatically by *preventative expressions* [Vander Linden and Di

(5) Utilisez une clé en croix pour desserrer les écrous de la roue [...] *N'enlevez pas complètement les écrous.*
Use a cross key to loosen the nut of the wheel [...] Do not completely remove the nuts.

EVENTUAL OPERATIONS indicate operations that might be executed by mistake or without any awareness from the user

(6) *Si vous tentez d'accéder au disque "C >"*, vous n'y parviendrez pas.
If you try to access the *"C >"* disk, you will not succeed

	Identification criteria								
Semantic carrier	procedural notion	operation	mandatory execution	immediate execution	individual execution	involved in instance generation	involved in type generation	positive operation	voluntary choice
ATTRIBUTE	no								
REQUIRED OPER.	yes	yes	yes	yes	yes	no		yes	
MATERIAL COND.	yes	no				no			
OUTCOME	yes	no				yes	yes		
OPTION	yes	yes	yes			yes	yes	yes	
GUIDANCE	yes	yes	yes			yes	no	yes	
CO-TEMPORAL OPER.	yes	yes	yes	yes	no	no		yes	
OPTION	yes	yes	no						yes
PREVENTION	yes	yes			yes	no		no	
EVENTUAL OPER.	yes	yes	yes	no		no			no
OTHER	yes	yes	no						

Table 2: Functional criteria used to identify the semantic carriers

The semantic carriers have been identified only through functional criteria and, thus, are independent of their syntactic form. Table 2 shows the criteria we used for identifying

the semantic carriers. For example, a REQUIRED OPERATION is defined as a semantic element that refers to a procedural aspect of the text, specifies an operation rather than a state, is mandatory, is performed individually and not in parallel with another operation, is not involved in an instance generation relation (as defined by [Goldman, 1970]) and refers to a positive operation (eg. *Do A*, rather than *Don't do A*). These criteria are generally sufficient to identify a semantic carrier, however, in some cases the context or knowledge of the reader is not sufficient to verify if a criterion is satisfied or not. For example:

(7) *Pour pallier à cet inconvénient*, peindre ces surfaces transparentes avec les couleurs à l'huile et attendre qu'elles sèchent.
To avoid this inconvenience, paint these transparent surfaces with the oil-based colors and wait until they dry out.

Here, *to avoid this inconvenience* is a postcondition to the operation *to paint*, but from the context, it is unclear as to whether it is mandatory or optional to reach this postcondition. In the first case, we are dealing with an OUTCOME, in the second case an OPTION is specified. In these situations, we choose the most salient interpretation in context.

Table 3 [5] shows the number of occurrences of the semantic carriers in our corpus. From this table, we can see that about half the content of the texts (52%) are made up of REQUIRED OPERATIONS; this is not surprising as the main goal of an instructional text is to indicate to the reader what actions to take to achieve some goal. Table 3 also shows the frequency of the semantic carriers by types of texts (execution, comprehension and hybrid). We can see that semantic carriers are influenced by the type of text. For example, execution texts contain many less attributes but more required operations than comprehension texts. These results are not surprising, considering that an execution text tells what to do; while an comprehension text explains.

More detail about the semantic analysis of our corpus can be found in [Kosseim, 1995].

Once a semantic element has been chosen, it may be expressed from a given point of view, but the mapping between these two levels is not direct as we showed in section 1. The next section describes the presentation heuristics for choosing the best rhetorical relations in a given context.

[5] In this paper, all percentages have been rounded off to the nearest non-decimal value. Their sum can therefore differ for 100.

	Entire corpus Entire corpus		Execution Texts	Hybrid Texts	Comprehension Texts
Semantic carrier	**Number of occurrences**	%	%	%	%
ATTRIBUTE	158	11	3	17	95
REQUIRED OPERATION	762	52	65	40	29
MATERIAL CONDITION	164	11	11	12	9
OUTCOME	136	9	7	13	9
GUIDANCE	124	8	9	8	8
CO-TEMPORAL OPERATION	45	3	1	4	7
OPTION	34	2	2	3	3
PREVENTION	21	1	1	2	2
EVENTUAL OPERATION	15	1	1	1	2
OTHER	12	1	0	0	5
Total	1471	≈ 100	100	100	100

Table 3: Frequencies of semantic carriers

4 The Rhetorical Structure

As instructional texts exhibit a rather stereotypical structure, the set of rhetorical devices is rather limited compared to the whole spectrum of RST relations. [Vander Linden, 1993] and [Rösner and Stede, 1992] have identified the rhetorical relations typically used. For Vander Linden, the most important RST relations are *temporal sequence*, *precondition* which we call *c-condition*, *purpose*, *result* and *concurrence*. Roesner and Stede identified other relations: *until* and *alternative* for which we found very few occurrences ($< 1\%$) so we did not take them into account. We also decided to combine their *step-sequence* with the usual *sequence*. We now give more details on the rhetorical relations we kept in our study:

sequence is a multinucleic relation where nuclei follow each other in the text;

c-condition combines RST's relations of circumstance and condition; Roesner and Stede and Vander Linden call this a *precondition* but we already use this term in its AI planning definition;

elaboration present additional information about the nucleus;

purpose in which by doing the satellite, the reader is better equipped to do the nucleus; Vander Linden does not distinguish this from the *goal* but we do;

result includes volitional and non-volitional results of RST;

means presents a situation to be realized by means of another activity;

Rhetorical relation	Number of occurrences	%
sequence	770	52
c-condition	172	12
elaboration	170	12
purpose	118	8
result	98	7
means	97	7
concurrency	45	3
Total	1471	$\approx$ 100

Table 4: Frequency of rhetorical relations

Name of heuristic	Semantic Carrier	Rhetorical Relation						
		sequence	*c-condition*	*elaboration*	*purpose*	*result*	*means*	*concurrency*
AT	ATTRIBUTE			100%				
RO	REQUIRED OPERATION	98%	1%		1%			
MC	MATERIAL CONDITION	2%	90%		4%	4%		
OU	OUTCOME				28%	68%	4%	
GD	GUIDANCE				31%		69%	
CO	CO-TEMPORAL OPERATION							100%
OP	OPTION		21%		79%			
PR	PREVENTION	86%					14%	
EO	EVENTUAL OPERATION		73%				27%	

Table 5: Global mapping between semantic carriers and rhetorical relations in the corpus

action concurrency is multinucleic like the sequence but the nuclei must be done at the same time.

Table 4 shows the results of the rhetorical analysis; while the mapping between semantic carriers and rhetorical relations in our corpus is shown in Table 5. As we can see, almost all REQUIRED OPERATIONS are presented within a *sequence* (98% of the time), but an OUTCOME can be presented through 3 different rhetorical relations: a *purpose* (29%), a *result* (68%) or a *means* (4%). This study enabled us to determine heuristics for choosing appropriate rhetorical relations to express semantic carriers. These heuristics are based on several factors given in the next section.

4.1 Criteria for choosing heuristics

During the corpus analysis, we took into account 5 types of constraints in order to identify which rhetorical relation was most appropriate to present a semantic carrier.

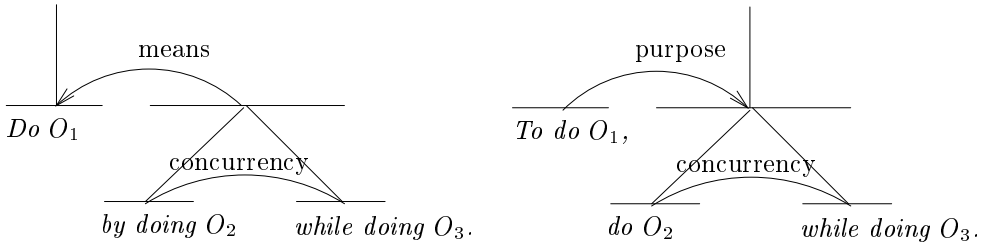
Conceptual constraint: As many studies is psychology suggest (ex. [Dixon, 1987, Dixon et al., 1988, Donin et al., 1992]), the content and structure of the concep-

tual representation of a procedure should be taken into account in order to construct instructions that are easy to interpret. An important conceptual criteria is the notion of *basic level operations* which can informally be defined as operations that are specific enough to be descriptive, yet general enough to be meaningful. More specifically, according to [Rosch, 1978], basic-level operations seem to be operations that are most easily remembered. Along the same lines, [Pollack, 1986] defines the notion of domain-basic act-types. For Pollack, *typing a character on a keyboard* cannot be considered a basic-level operation in the computer domain because it is too specific; just like *typing a string of characters* or typing anything at all. In this domain, a basic-level operation would be *issuing a command*. Pollack stipulates that within a discourse domain there exists a set of basic-level operations, and these may be agent-specific. We use the notions of basic-level operation of Rosch and Pollack, to explain the communication of certain operations in instructional texts. Indeed, basic-level operations are included in the text because the writers take for granted that readers have an easily accessible mental representation of them and because they ease the reader's recall since the readers can easily construct a memory representation of them.

Semantic constraint: The most important factor in determining what rhetorical relation to use is what semantic information we wish to convey. For a particular semantic carrier, only a subset (see Table 5) of the rhetorical relations are acceptable. For example, a MATERIAL CONDITION cannot be conveyed through a *means* or a *concurrency*.

Rhetorical constraints: As semantic carriers are mapped into a portion of a rhetorical relation (the satellite or the nucleus), it is necessary to ensure that the other portion of the rhetorical relation will be filled by another semantic carrier, in order to have a well-formed relation. The choices of rhetorical relations to present semantic carriers are therefore co-dependant. Not finding a filler for the other portion of the rhetorical relation is not a concern, because the selection of the semantic carriers already made sure that they can either stand alone (as is the case with REQUIRED OPERATIONS or that the element they refer to will also be communicated in the text, thus providing a filler for the other portion of the rhetorical relation.

When looking for a filler to the other portion of the rhetorical relation, it should be kept in mind that in instructional texts, some rhetorical relations seem to co-occur while some other combinations are never found. For example, if two MATERIAL CONDITIONS are to be presented to constrain the same operation, a *c-condition* will be used for the "easiest" one to verify and a *result* will convey the other, for example:



a. Unlikely structure

b. Preferred structure

Figure 3: Preferred rhetorical structure

(8) *S'ils [les écrous] portent la marque "L", ils ont le filetage à gauche* et vous devez les dévisser ....
If they [the nuts] have an "L" mark, *they have a left-hand thread*, and you must unscrew them ...

However, one will never find a *concurrency* related to its nucleus by a *means*. This unlikely form is shown in:

(9) * Do RO_1 by doing RO_2 while doing RO_3.

and in Figure 3a. To convey the same information, a *purpose* related to a *concurrency* is preferred, as in:

(10) To do RO_1, do RO_2 while doing RO_3.

and in Figure 3b.

Pragmatic constraints: This constraint takes into account specifications of the nature of the procedure (i.e. pragmatic characteristics of the operations and states of the procedure) to select a rhetorical relation. This includes the optionality and degree of desirability of an operation (if an optional line of operations is generally desirable, it will be conveyed differently than one rarely chosen), the level of danger of a negative operation and the internal/external status of states.

Intentional constraints: What a "generic" reader believes about the operations and states of the procedure and his pursued goals greatly influence how information is conveyed in the text. This is why a model of the imagined reader's knowledge and intentions must be taken into account to generate appropriate relations. For example, in:

Semantic Carrier	Rhetorical Relation						
	% sequence	*% c-condition*	*% elaboration*	*% purpose*	*% result*	*% means*	*% concurrency*
ATTRIBUTE			100-100-100				
REQUIRED OPERATION	98-98-100	1-1-0		1-1-0			
MATERIAL CONDITION	3-0-0	90-92-86		3-8-0	3-0-14		
OUTCOME				20-28-27	77-70-64	3-2-9	
GUIDANCE				37-27-10		61-73-90	
CO-TEMPORAL OP.							100-100-100
OPTION		19-27-14		81-73-86			
PREVENTION	100-88-50					0-12-50	
EVENTUAL OP.		100-100-20				0-0-80	

Table 6: Mapping between semantic carriers and rhetorical relations in execution, hybrid and comprehension texts

(11) a. *Si l'on souhaite une ligne plus large*, alors s'attarder sur le verre de façon à laisser s'écouler plus de couleur.
If you wish a thicker line, stay on the glass longer so that more paint can flow.

b. *Pour une ligne plus large*, alors s'attarder sur le verre de façon à laisser s'écouler plus de couleur.
To have a thicker line, stay on the glass longer so that more paint can flow.

If the two relations in italics present the semantic carrier of OPTION, a relation of *c-condition* (11a) is preferred for novice readers as the optional aspect is explicit. A relation of *purpose* (11b) does not convey the optionality as explicitly and can be mistakenly interpreted by a novice reader as a mandatory goal to be achieved.

Note that in our corpus, we did not found special selection rules for different types of text. Table 6 shows how semantic carriers are presented in the 3 types of texts analyzed: execution, hybrid and comprehension texts. The table shows that most semantic carriers are not influenced by this factor. For example, REQUIRED OPERATIONS are presented by a sequence 98% of the time in execution texts and hybrid texts and all the time in comprehension texts. However, the semantic carriers of GUIDANCE, PREVENTION AND EVENTUAL OPERATION do seem influenced by the textual type. In the case of PREVENTION and EVENTUAL OPERATION, the number of occurrences in our corpus is too low to verify or contradict this claim. In the case of GUIDANCES, we did not conduct a full statistic study but a cursory examination of the corpus seems to indicate that they are influenced by the textual genre.

In the next section, we will discuss the presentation heuristics based on the criteria we identified in this section.

4.2 The Presentation Heuristics

We now give the criteria for selecting an appropriate rhetoric relation to communicate a semantic carrier. The following rules are given in order of preference which is also the order in Table 5, i.e. as soon as semantic carriers satisfy a criteria, we choose the corresponding rhetorical relation. These heuristics are implemented in SPIN which will be presented in the next section.

4.2.1 Attributes of Objects

Attributes of objects are among the semantic carriers that are always presented through only one rhetorical relations. According to our corpus, the relation of *elaboration* is always used. For example:

(12) *Une lampe à une seule douille comporte habituellement une lyre qui tient à une barre de retenue par des manchons.*
A single-socket lamp usually has a shade holder attached to a retaining bar by a few fittings.

4.2.2 Required Operations

An required operation can be presented by 3 rhetorical relations:

- a *sequence* of actions (98 % of the time):

  (13) *Mettre le magnétoscope sous tension* et *placer le sélecteur* TV/VCR *sur* "VCR".
  Turn the recorder on and place the TV/VCR selector on VCR.

- a *purpose* (1 %):

  (14) Vérifiez le commutateur de la douille ou réinstallez la douille *pour [ensuite] vérifier le commutateur du socle*
  Check the socket switch or install the socket again to [then] check the switch on the base.

- a *c-condition* (1 %), as in examples (15a) and (15b),

  (15) a. Retrait de la lyre:
  [...] sur certaines lampes, il faut enlever la douille *avant de soulever la lyre.*
  Dégagement de la douille:

[...]

Removal of the lamp-shade holder:
[...] on certain lamps, the socket must be removed before lifting the lamp-shade holder.
Removal of the socket
[...]

b. Appuyer sur la touche OTR pour specifier l'heure d'enregistrement.
Lorsque la touche est enfoncée une fois, 'PM 10:35 (30 min)' sera affiché.
Lorsque la touche est enfoncée deux fois, 'PM 11:05 (1h)' sera affiché.
Press the OTR button to specify the recording time.
When the button is pressed once. 'PM 10:35 (30 min)' will be displayed.
When the button is pressed twice. 'PM 11:05 (1h)' will be displayed.

The presentation heuristics for REQUIRED OPERATIONS are:

RO-1 A *c-condition* is used in 3 distinct cases:

RO-1a semantic constraint: Two REQUIRED OPERATIONS RO_1 and RO_2 are to be communicated.

conceptual constraint: RO_1 and RO_2 should be executed in a particular order.

intentional constraint: The reader believes that RO_1 and RO_2 should be executed in the reverse order.

This is the case in (15a) where *the removal of the lamp-shade holder* and *the removal of the socket* are, most of the time, to be executed in that order (as indicated by the titles of the procedures) but on some lamps the same operations must be executed in the reverse order. If the order of execution is reversed, a *c-condition* is chosen for one of the operation instead of a *sequence* because it puts the emphasis on the temporal order of operations.

RO-1b semantic constraint: A REQUIRED OPERATION RO is to be communicated.

conceptual constraint: The agent *of* RO is not necessarily the agent of rest of the procedure.

rhetorical constraint: A *result* or a *sequence* will be related to RO in the text to ensure it a nucleus.

In this case, RO is presented by a *c-condition* because this relation does not mention explicitly or implicitly, who must perform the operation. In addition,

we say that a *result* or a *sequence* must be related to RO in the text in order to ensure a nucleus for the *c-condition*; as in

(16) a. Lorsque RO est fait, ceci se produira.
When RO is done, this will happen.

b. Lorsque RO est fait, faites RO_2.
When RO is done, do RO_2.

RO-1c semantic constraint: A REQUIRED OPERATION RO is to be communicated along with its OUTCOMES OU_i.

conceptual constraint: RO is divided into a set of sub-operations RO_i who have postconditions OU_i.

rhetorical constraint: The OUTCOMES OU_I will be communicated in the text by *results* to ensure a nucleus for RO_i.

In this case, even if the reader knows how to perform RO, its sub-operations RO_i must be included in the text in order to have a nucleus for the *results* OU_i. These sub-operations RO_i are presented by temporal *c-conditions*. This is the case in (15b), where the sub-operations *press the button once* and *press the button twice* should normally not appear because they are too primitive to be communicated to the user as part of the operations to perform. They are nevertheless communicated in the text for rhetorical reasons, not mainly for semantic ones although we could argue that this redundant information could be useful.

RO-2 A *purpose* is used if:

semantic constraint: Two REQUIRED OPERATIONS RO_1 and RO_2 are to be communicated.

conceptual constraint: RO_1 is a basic-level operation; while RO_2 is a precondition to RO_1.

pragmatic constraint: RO_2 is a modifiable condition.

rhetorical constraint: RO_2 can and will be presented within a *sequence*.

As defined by [Vander Linden, 1993], a condition is modifiable if the agent must check it or perform an operation so that it becomes true. This is the case in (14) where, although, there is no hierarchical relation between the two operations *check the socket switch or install the socket again* and *check the switch on the base*, they are related rhetorically by a purpose, but semantically by a sequence.

RO-3 In all other cases, an REQUIRED OPERATION RO is presented inside a *sequence*; as in (13).

4.2.3 Material Conditions

A material condition can be presented by 4 rhetorical relations:

- A *c-condition* (90 % of the time):

  (17) Au laboratoire d'IA, *si vous êtes sur la machine nommée* NIL, vous êtes sur une station SPARC-10; utilisez la procédure des SPARC-10.
  In the AI laboratory, if you are on the machine called NIL, you are on a SPARC-10 station; use the procedure for a SPARC-10.[6]

- A *purpose* (4 %):

  (18) *Pour [vérifier] un commutateur ordinaire* [...], touchez la vis de la borne de cuivre avec la pince du vérificateur.
  To [check] an ordinary switch, touch the screw of the copper terminal with the tester.

- A *result* (4 %):

  (19) Au laboratoire d'IA, si vous êtes sur la machine nommée NIL, *vous êtes sur une station* SPARC-10; utilisez la procédure des SPARC-10.
  In the AI laboratory, if you are on the machine called NIL, you are on a SPARC-10 station; use the procedure for a SPARC-10.[7]

- a *sequence* (2 %):

  (20) Introduire la cassette (*vérifier que la languette de la vidéocassette n'a pas été enlevée.*)
  Insert the cassette (check that the tab of the video cassette has not been removed.)

MC-1 A *result* and a *c-condition* are used if:

semantic constraint: Two equivalent material conditions are to be communicated.

pragmatic: The condition that is difficult to evaluate presents the main condition justifying the next line of actions. This state is always presented by a *result* and the state that is easier to verify uses a *c-condition*.

[6] Example outside our corpus.

[7] Example outside our corpus.

This is the case in (17) and (19) where the expression *in the AI laboratory* defines the context in which the equivalence of the states is verified.

MC-2 *sequence* is used if:

semantic constraint: A MATERIAL CONDITION MC is to be communicated, and MC is not related to the operation it constrains by an enablement [Di Eugenio, 1993].

pragmatic constraint: MC is a modifiable condition.

This is the case in (20). This phenomenon seems to apply both in French and in English. In fact, Vander Linden notes that conditions that specify a modifiable state are presented by a *sequence* [Vander Linden, 1993]. He calls this phenomenon a *rhetorical promotion*. Indeed, what is considered a condition on an operation at the semantic level is promoted to a full-fledge agent-action at the rhetorical level. Instead of being presented as the satellite of a relation, it becomes a nucleus.

MC-3 A *purpose* or a *c-condition* with ellipsis of the verb is used when:

semantic constraint: A MATERIAL CONDITION MC is to be communicated.

pragmatic constraint: MC pertains to the nature a device.

In this case, the MATERIAL CONDITION is presented by either *purpose* or a *c-condition* (see ex. (18)).

MC-4 In all other cases, MATERIAL CONDITIONS are presented by a *c-condition*, as in (17).

4.2.4 OUTCOMES

An OUTCOME can be presented by:

- a *result* (68 %):

(21) a. Brancher le cordon d'alimentation du magnétoscope dans une prise secteur 120V et appuyer sur l'interrupteur "POWER". *Le voyant* "POWER" *s'allume et l'horloge commence à clignoter.*
Plug the electrical cord of the recorder in a 120V outlet and press the POWER button. The power light is turned on and the clock starts to blink.

b. Engagez le levier de vitesse rapidement dans chacune des positions [...], *cela fait circuler le liquide de transmission.*
Put the gear-shift lever rapidly in each position [...], this makes the transmission liquid circulate.

- a *purpose* (28 %):

  (22) *Pour protéger les bornes contre la tension*, nouez les extrémités séparées du cordon.
  To protect the terminals from contacting each other, tie the extremities of the wire away from each other.

- a *means* (4 %):

  (23) Vous pouvez voir le niveau de volume *en observant la barre* ROUGE *sur la gamme de 15 barres affichées sur l'écran.*
  You can see the volume level by observing the bar on the 15-bar scale displayed on the screen.

  In this case, the OUTCOME (*to see the volume level)* is presented in a *sequence* of actions and the operation that brings it about (*to observe*) is related to it by a relation of *means*. Here, the OUTCOME is promoted to the nucleus position of relation; while *to observe* is demoted to the satellite position.

[Delin et al., 1994] note that in English, an OUTCOME can be presented by a *sequence*, but this phenomenon has not been found in our French corpus.

The choice between a *purpose*, a *result* and a *means* does not seem to depend on cooccurrences of rhetorical relations but on whether the outcomes are desirable or not. [Kosseim, 1995] describe many examples that can be categorized by the following heuristics:

OU-1 A *result* is always used if

semantic constraint: An OUTCOME *OU* is to be communicated.

pragmatic constraint: *OU* specifies the reaction of a device. In that case, it is generally a non-desirable side-effect (as in (21a)) or the reader cannot guess that *OU* is desirable (as in (21b)).

OU-2 A *purpose* is used if:

semantic constraint: An OUTCOME *OU* is to be communicated.

pragmatic constraint: The reader knows or can guess that *OU* is desirable and the method described in the instruction is the normal way to realize the OUTCOME (as in (22)).

A *means* is also used if the reader cannot guess that the OUTCOME is desirable and if the operation that brings about the OUTCOME is the normal method to achieve it. However, in this case, a *result* is always available. As our corpus only includes 4 % of *means*, but 68 % of *result* that present OUTCOMES, we decided to always use an *result*.

4.2.5 GUIDANCES

A GUIDANCE is made up of a parent operation RO_p and its sub-operations RO_i which "guide" its execution. A GUIDANCE can be presented by:

- A relation of *means* (69 %); in this case the nucleus is constituted of the parent-operation:

  (24) a. La paraison est centrée soit *en utilisant le marbre*, soit *en la roulant dans une forme creusée dans du bois, appelée "mailloche".*
  The glass blob is centered either by using the marble or by rolling it in a wooden concave form called "mailloche".

  b. *Avec un tournevis plat*, grattez la saleté accumulée sur le contact.
  With a flat-headed screwdriver, scrape off the dirt accumulated on the contact.

  c. Régler la ceinture *en la tirant par la languette*.
  Adjust the seat belt by pulling it by the strap.

- A relation of *purpose* (31 %); in this case the nucleus is made up of the sub-operations:

  (25) a. Si la boîte-pont est munie d'une jauge d'huile, tirez-la hors du tube de remplissage, essuyez-la, réinsérez-la complètement et retirez-la de nouveau *pour lire le niveau d'huile*.
  If the oil sump has a dip stick, pull it out of the filling tube, wipe it, insert it again entirely and remove it again to read the oil level.

  b. Tourner cette touche à droite et à gauche *pour minimiser les parasites*.
  Turn this knob left and right to minimize interference.

  c. Rouler le pinceau en le tirant vers soi *de façon à reformer la pointe*.
  Twist the paint brush while pulling it towards you in order to form a pointed tip.

It is interesting to note that for grammatical reasons, all operations of the same level of abstraction are presented by the same rhetorical relation; the GUIDANCES in:

(26) Pour faire RO_{gp}, faites RO_p en faisant RO_1 et RO_2.
To do RO_{gp}, do RO_p by doing RO_1 and RO_2.

involve operations from three different levels of abstraction: grand parent (RO_{gp}), parent (RO_p) and children (RO_1 and RO_2). However, the operations of the same level (RO_1 and RO_2) are both presented by the same relation (a *means*).

The heuristics involved in the choice of the rhetorical relation are:

GD-1 A *means* is used if:

GD-1a semantic constraint: A REQUIRED OPERATION RO is to be communicated along with a set of guiding operations (GUIDANCE) G_i.

conceptual constraint: G_i is a set of sub-operations of an operation RO.

rhetorical constraint: RO can and will be presented by a *sequence*.

pragmatic constraint: At least one member of G_i presents the use of an instrument.

In this case all members of G_i are presented through a means, so that operations at the same level of abstraction are presented by the same rhetorical relation. For example, in (24a), the use of *marble* dictates the choice of a *means* for all the sub-operations.

GD-1b semantic constraint: A REQUIRED OPERATION RO_p is to be communicated and only one of its sub-operation G_i is to be communicated as a GUIDANCE to RO_p.

conceptual constraint: RO_p is the parent operation of G_i and RO_p is a basic-level operation.

rhetorical constraint: RO_p can and will be presented by a *sequence*.

In this case, the GUIDANCE is generally seen top-down (the nucleus is formed of RO_p and the satellite of G_i) by a relation of *means*, as in (24c).

GD-2 A relation of *purpose* is used if:

GD-2a semantic constraint: A REQUIRED OPERATION RO_p is to be communicated along with a set of guiding operations (GUIDANCE) G_1 and G_2.

conceptual constraint: RO_p is the parent operation of operations G_1 and G_2 which should be executed co-temporally.

rhetorical constraint: RO_p can and will be presented by a *sequence*.

This is the case in (25c) where the sub-operations (*twist* and *pull*) should be performed co-temporally. The only way to communicate this temporal aspect

is to use a *concurrency* to relate G_1 and G_2, and a *purpose* between O_p and the pair (G_1, G_2).

GD-2b semantic constraint: A REQUIRED OPERATION RO_p is to be communicated along with a set of guiding operations (GUIDANCE) G_i.

conceptual constraint: RO_p is the parent operation of operations G_i.

pragmatic constraint: No operation of G_i indicates the use of an instrument.

rhetorical constraint: RO_p can and will be presented by a *sequence*.

This is the case in (25a).

GD-2c semantic constraint: A REQUIRED OPERATION RO is to be communicated along with a single GUIDANCE G.

conceptual constraint: RO and G are both operations; G is the sub-operation of RO and G is a basic-level operation.

rhetorical constraint: RO can and will be presented by a *sequence*.

This is the case in (25b).

4.2.6 CO-TEMPORAL OPERATIONS

Similarly to ATTRIBUTES, CO-TEMPORAL OPERATIONS are always presented through a single rhetorical relation. In the case of CO-TEMPORAL OPERATIONS, a *concurrency* is always used:

(27) *Rouler le pinceau en le tirant vers soi* de façon à reformer une pointe.
Twist the paint brush while pulling it toward you in order to form a pointed tip.

One should distinguish the relation of *concurrency* and the relation of *means* which, in French, are both realized grammatically by a gerund (eg. *en soufflant*). In the case of a *concurrency*, it is possible to add the adverb *tout* before the gerund (eg. *tout en soufflant*) without modifying the meaning of the expression, while a relation of *means* cannot be realized by *tout* + gerund. In English, the distinction is more obvious as a *means* is realized by a gerund; while the *concurrency* is realized without the preposition *by* but with a preposition like *while*, *meanwhile*, ... [Vander Linden, 1993].

4.2.7 OPTIONS

OPTIONS can be presented in the text by:

- A *purpose* (79 %):

(28) Tirer la roue et le pneu; *pour vous aider*, poussez fermement le flanc du pneu avec votre pied.
Pull the wheel and the tire; to help you, firmly press on the side of the tire with your foot.

- A *c-condition* (21 %):

(29) *Si l'on souhaite une ligne plus large*, alors s'attarder sur le verre de façon à laisser s'écouler plus de couleur.
If you wish a thicker line, stay longer on the glass in order to let more color flow.

The choice of rhetorical relation only depends on semantic and intentional constraints:

OP-1 A *purpose* is used if:

semantic constraint: An OPTION OP is to be communicated.

intentional constraint: OP will probably be followed by the agent.

rhetorical constraint: A *sequence* will be related to OP in the text to ensure it a nucleus.

This is the case in (28).

OP-2 On the other hand, a *c-condition* is used if:

semantic constraint: An OPTION OP is to be communicated in the text

intentional constraint: OP is as likely to be followed as to be rejected by the agent.

rhetorical constraint: A *sequence* will be related to OP in the text to ensure it a nucleus.

This is the case in (29).

The *c-condition* gives an explicit choice to the agent by using expressions like *if you wish* while *purpose* expresses the OPTION less explicitly, thus restricting the possibility of rejection.

In this section we will not discuss PREVENTION and EVENTUAL OPERATION because our corpus only contained 15 and 12 occurrences of these semantic carriers, barely enough to develop heuristics of a more general nature.

This section identified the presentation heuristics we have extracted from the examples found in our corpus. These rules have been implemented in SPIN described in the next section.

5 The SPIN system

SPIN performs all levels of text generation going from a conceptual representation to a French formatted text. The emphasis of SPIN was put on the text planning stage: the semantic carrier and the rhetorical relation structurers. The other modules are rather straightforward but by building a complete system, we have shown the feasibility of the overall generation system while making sure that hard problems in the semantic and rhetorical stages were not "shoved off" to another module. We make sure that the planning stage can indeed give all the necessary information for the semantic structurer and that the output given by the rhetorical structurer is sufficient to obtain a readable text.

SPIN follows the linear architecture shown in Figure 4. Although this type of architecture prevents us from having lower-level decisions influence upper-level ones, it was chosen for its implementation simplicity. However, SPIN is written in Prolog so that, if a choice made at one level cannot be realized by lower-level components, Prolog's built-in backtracking mechanism allows the previous component to try another possibility until the specification is realisable by the lower-level components. The generation of a text is performed through the components shown in Figure 4:

The task planner is used to plan the procedure to achieve a particular goal. To achieve this, it constructs a *conceptual representation of the task* using a non-linear AI planner and a library of schemas of operators [Sacerdoti, 1977]. This technique is typically used in the generation of instructional texts (see for example, [Mellish, 1988], [Dale, 1992], [Paris and Vander Linden, 1996a]) because the structure of the resulting plans is hierarchical, similarly to the structure of the text. The result of this step is a tree-like structure of operator schemas necessary to achieve the top-level request. For example, given the top-level request, `record_by_OTR(object: X)` and a library of operator schemas like the one shown in Figure 5, the task planner builds a hierarchical conceptual representation of the task linking operation and suboperations in a tree structured representation. Each level of the tree is a plan at a given level of abstraction.

The semantic carrier structurer first selects the semantic carriers that will be expressed while leaving others implicit. SPIN takes into account a model of the reader indicating which operations are known. This model is updated when an information is selected for expression; the nature of the task is described by properties such as the fact that an operation is dangerous, optional or irreversible, the fact that the procedure is to be followed at the time of reading, etc Figure 6 show the set of relevant semantic carriers computed by the procedure.

These information are then reordered so that preparation operations (indicated

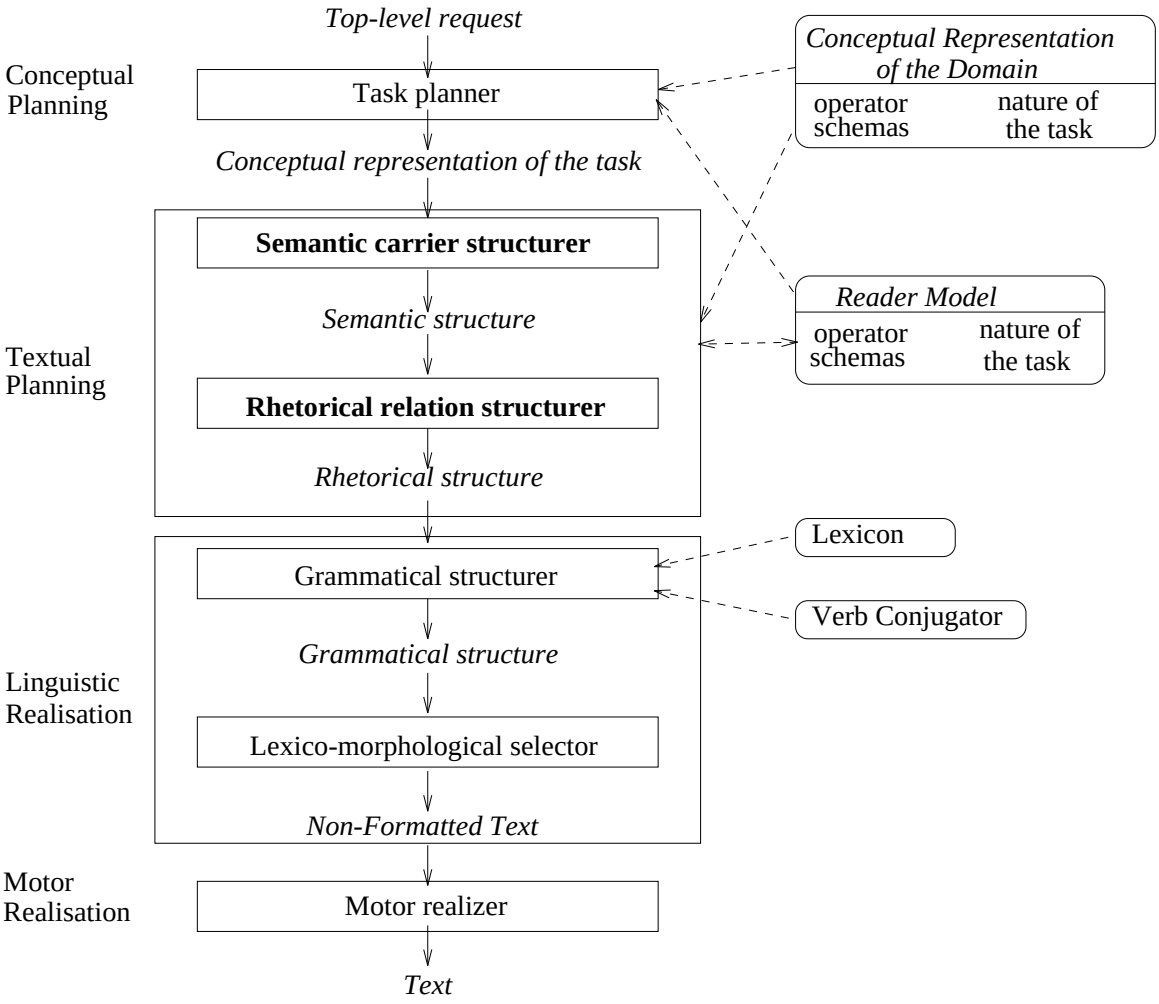


Figure 4: Architecture of SPIN

operator:	`record_by_OTR(objet: X)`
precondition:	`inside(object: cassette, location: recorder)`
body:	`set(object: tape speed, to: SP)`
	`select(object: channel 4)`
	`specify(object: duration time of 1 h 30 min)`
	`press(object: Timer button)`
success postcondition:	add: `programmed(objet: X)`
failure postcondition:	add: `not(programmed(objet: X))`

Figure 5: Example of operator schema

```
[title(programmer(obj: enregistrement, qual: d'1h 30 min par une touche incrémentielle)),

 req_op(régler(obj: sélecteur de vitesse de bande, dest: SP)),
 option(meilleure qualité d'image, régler(obj: sélecteur de vitesse de bande, dest: SP)),
 req_op(sélectionner(obj: canal 4)),
 guidance(sélectionner(obj: canal 4), appuyer(dest: touche de canal)),
 req_op(appuyer(dest: touche TIMER, qual: dans un délai de 9 secondes)),
 req_op(appuyer(dest: touche de canal)),
 req_op(appuyer(dest: touche OTR, qual: 3 fois)),
 req_op(enfoncer(obj: touche OTR, qual: 1 fois)),
 outcome(PM 10:35 (30min), enfoncer(obj: touche OTR, qual: 1 fois)),
 req_op(enfoncer(obj: touche OTR, qual: 2 fois)),
 outcome(PM 11:05 (1 h), enfoncer(obj: touche OTR, qual: 2 fois)),
 req_op(enfoncer(obj: touche OTR, qual: 3 fois)),
 outcome(PM 11:35 (1 h 30min), enfoncer(obj: touche OTR, qual: 3 fois))]
```

Figure 6: Semantic carriers of the vcr text

```
[title(programmer(obj: enregistrement, qual: d'1h 30 min par une touche incrémentielle)),

req_op(régler(obj: sélecteur de vitesse de bande, dest: SP)),
option(meilleure qualité d'image, régler(obj: sélecteur de vitesse de bande, dest: SP)),
req_op(sélectionner(obj: canal 4)),
guidance(sélectionner(obj: canal 4), appuyer(dest: touche de canal)),
req_op(appuyer(dest: touche de canal)),
req_op(appuyer(dest: touche OTR, qual: 3 fois), action),
req_op(enfoncer(obj: touche OTR, qual: 1 fois)),
outcome(PM 10:35 (30min), enfoncer(obj:touche OTR, qual: 1 fois)),
req_op(enfoncer(obj: touche OTR, qual: 2 fois)),
outcome(PM 11:05 (1 h), enfoncer(obj: touche OTR, qual:2 fois)),
req_op(enfoncer(obj: touche OTR, qual: 3 fois)),
outcome(PM 11:35 (1 h 30min), enfoncer(obj: touche OTR, qual: 3 fois)),
req_op(appuyer(dest: touche TIMER, qual: dans un délai de 9 secondes))]
```

Figure 7: Semantic structure of the vcr text

```
[title,
 purpose(programmer(obj: enregistrement, qual: d'1h 30 min par une touche incrémentielle), _)
 paragraph,
 sequence(régler(obj: sélecteur de vitesse de bande, dest: SP)),
 purpose(meilleure qualité d'image, régler(obj: sélecteur de vitesse de bande, dest: SP)),
 sequence(sélectionner(obj : canal 4)),
 means(appuyer(dest : touche de canal),
 sequence(appuyer(dest: touche OTR, qual: 3 fois)),
 c_condition(enfoncer(obj: touche OTR, qual: 1 fois), appuyer(dest: touche OTR, qual: 3 fois)
 result(PM 10:35 (30min), enfoncer(obj: touche OTR, qual: 1 fois)),
 c_condition(enfoncer(obj: touche OTR, qual: 2 fois), appuyer(dest:touche OTR, qual: 3 fois))
 result(PM 11:05 (1 h), enfoncer(obj: touche OTR, qual: 2 fois)),
 c_condition(enfoncer(obj: touche OTR, qual: 3 fois), appuyer(dest: touche OTR, qual: 3 fois)
 result(PM 11:35 (1 h 30min), enfoncer(obj: touche OTR,qual: 3 fois)),
 sequence(appuyer(dest: touche TIMER, qual: dans un délai de 9 secondes))]
```

Figure 8: Rhetorical relations of the `vcr` text

as such in the task representation) are presented first, then autonomous tasks are presented. The remaining tasks are then presented while keeping adjacent task depending upon a single parent thus decreasing the number of focus changes. Figure 7 shows the list of semantic carriers where for example the `req_op(appuyer...)` is now correctly placed at the end of the list.

The rhetorical relation structurer takes over the semantic structure and applies the presentation heuristics described in section 4 to select the most appropriate rhetorical relation to communicate a semantic carrier. This process constructs an RST-like text representation. Figure 8 shows an example of semantic structure of the `vcr` text. Once the list of rhetorical relations selected, it must be reordered to take into account some cooccurrence constraints and to position the satellites in relation with their nucleus. Vander Linden [Vander Linden, 1993] has thoroughly studied this aspect in English and in our corpus, the relative position of the elements of a rhetoric relation seem to follow the same rules. This step finally adds appropriate punctuation signs between relations: each nucleus is presented in a single sentence. Figure 9 illustrates the final rhetorical structure of the `vcr` text.

The grammatical structurer selects the appropriate grammatical structures to present the rhetorical relations between the satellites and their nucleus. To realize this

```
[title,
 purpose(programmer(obj: enregistrement, qual: d'1h 30 min par une touche incrémentielle), _)
 paragraph, item,
 purpose(meilleure qualité d'image, régler(obj: sélecteur de vitesse de bande, dest: SP)), vi
 sequence(régler(obj: sélecteur de vitesse de bande, dest: SP)), point, item,
 sequence(sélectionner(obj: canal 4)),
 means(appuyer(dest: touche de canal), sélectionner(obj: canal 4)), point, item,
 sequence(appuyer(dest: touche OTR, qual: 3 fois)), point, nl, space,
 c_condition(enfoncer(obj: touche OTR, qual: 1 fois), appuyer(dest:touche OTR, qual: 3 fois))
 result(PM 10:35 (30min), enfoncer(obj: touche OTR, qual: 1 fois)), point, nl, space,
 c_condition(enfoncer(obj: touche OTR, qual: 2 fois), appuyer(dest: touche OTR, qual: 3 fois)
 result(PM 11:05 (1 h), enfoncer(obj: touche OTR, qual: 2 fois)), point, nl, space,
 c_condition(enfoncer(obj: touche OTR, qual: 3 fois), appuyer(dest: touche OTR, qual: 3 fois)
 result(PM 11:35 (1 h 30min), enfoncer(obj: touche OTR, qual: 3 fois)), point, nl, point, ite
 sequence(appuyer(dest: touche TIMER, qual: dans un délai de 9 secondes)), point]
```

Figure 9: Rhetorical structure of the `vcr` text

```
[titre,
 gn(programmer(obj: enregistrement, qual: d'1h 30 min par une touche incrémentielle)),
 paragraph, item,
 forme_pour(meilleure qualité d'image), virgule,
 actif_direct(régler(obj: sélecteur de vitesse de bande, dest:SP)), point, item,
 actif_direct(sélectionner(obj: canal 4)),
 gerondif(appuyer(dest: touche de canal)), point, item,
 actif_direct(appuyer(dest: touche OTR, qual: 3 fois)), point, nl, space,
 lorsque([enfoncer(obj: touche OTR, qual: 1 fois)]), virgule,
 forme_textuelle(PM 10:35 (30min)), point, nl, space,
 lorsque([enfoncer(obj: touche OTR, qual: 2 fois)]), virgule,
 forme_textuelle(PM 11:05 (1 h)), point, nl, space,
 lorsque([enfoncer(obj: touche OTR, qual: 3 fois)]), virgule,
 forme_textuelle(PM 11:35 (1 h 30min)), point, nl, point, item,
 actif_direct(appuyer(dest: touche TIMER, qual: dans un délai de 9 secondes)), point]
```

Figure 10: Grammatical structures of the `car` text

```
Programmation d'un enregistrement d'1h 30 min par une touche incrémentielle

- Pour une meilleure qualité d'image, réglez le sélecteur de vitesse de bande sur SP.
- Sélectionnez le canal 4 en appuyant sur la touche de canal.
- Appuyez sur la touche OTR 3 fois.
  Lorsqu'elle est enfoncée 1 fois, PM 10:35 (30min).
  Lorsqu'elle est enfoncée 2 fois, PM 11:05 (1 h).
  Lorsqu'elle est enfoncée 3 fois, PM 11:35 (1 h 30min).
- Appuyez sur la touche TIMER dans un délai de 9 secondes.
```

Figure 11: Final `vcr` text output from SPIN

step, we have adapted Vander Linden's analysis of English instructions to our French corpus, and in addition, we took into account the textual type as a criterion for choosing grmmatical structures. Indeed, our corpus analysis revealed a strong correlation between the textual genre (execution, hybrid or comprehension texts) and the grammatical form of rethorical relations; thus confirming [Hartley and Paris, 1996]'s results. Figure 10 shows the output of the grammatical structurer.

The lexico-morphological selector takes over the grammatical structure and selects the lexemes to be used. The words are chosen (in the current implementation, a one-to-one correspondence exists between concepts and words) and declined. This component is also responsible for producing anaphoric expressions based on [Tutin, 1992]'s work. For example, these rules enabled SPIN to generate

```
Appuyez sur la touche OTR 3 fois.
Lorsqu'elle est enfoncée 1 fois, PM 10:35 (30min).
```

where the pronoun *elle* refers to `touche OTR` already introduced in the text. SPIN can also produce personal pronouns, partial repetitions and ellipsis of noun phrases.

The motor realizer finally realizes the typographical layout of the text. Elisions, capitalizations, etc are performed. SPIN can produce paragraph-structured procedures and itemized lists to produce the text given in Figure 11.

6 Evaluation of the Heuristics

Today, in natural language generation, there exists no accepted methodology to validate a generation theory. The problem is amplified by two phenomena: the lack of a canonical

form for the input and sometimes the output of a generation system and the difficulty to define the set of "correct" output [Walker, 1989, Bates et al., 1994].

We do not believe that a word-for-word comparison of a language generation theory's output and a "natural" text is an appropriate evaluation method. Indeed, if the human writer has chosen a different linguistic form (semantic, rhetorical, etc) from the theory to be validated, it does not necessarily imply that the theory is wrong. In fact, it might only demonstrate the many to many mapping between the different levels of text representations which allows for the richness and flexibility of natural language.

A valid and more objective evaluation method consists of giving subjects a set of "natural" and generated texts and asking them to evaluate them according to specific criteria (ex. informational content, text coherence, comprehensibility, etc). This method is particularly interesting for instructional texts as the subjects can be asked to actually perform the prescribed procedure. Criteria like their execution errors and reading time can be measured to evaluate the quality of the texts and compare them to "natural" ones. This method, although interesting, needs an involved experimental setup and experience in interpreting psychological performances that we do not have. It would also imply separating the text output quality from the cognitive ability of the individual readers. Given all these difficulties, we did not pursue further in this approach.

In order to validate our heuristics, we have thus followed two other more popular evaluation methods: a comparison with other instruction-generation systems and a global qualitative comparison of the generated texts with their "natural" counterparts.

6.1 Comparison with previous work

Two previous research projects have built instructional text generators: [Mellish, 1988] and [Dale, 1992]'s EPICURE. We have thus taken their output texts and managed to reproduce them through SPIN.

A comparison of both texts, presented in detail in [Kosseim, 1995], shows that SPIN uses better content selection heuristics than that of Mellish and that it can generate comparable output to the one produced by Epicure in the domain of cooking recipes although it was not optimized for this application.

An important research in the generation of instructional texts is the DRAFTER project [Paris et al., 1995, Paris and Vander Linden, 1996b]. In DRAFTER, the text planner determines the content to be included in the text as well as its rhetorical structure through [Moore and Paris, 1993]'s text planner. Similarly to SPIN's, DRAFTER's text planner allows the same conceptual information to be presented through different rhetorical relations; but in addition DRAFTER determines its choice according to the language of communication. In DRAFTER, the choice of the semantic information to be conveyed and the rhetorical relations to be used is performed at the same time through discourse

Text Generated by SPIN

Desserrage des écrous de la roue

- N'enlevez pas les écrous complètement.
- Desserrer les écrous avec une clé en croix.
- S'ils portent la marque L, les écrous ont le filetage à gauche, tourner les écrous dans le sens des aiguilles d'une montre.
- S'ils ne portent aucune marque, tourner les écrous en sens contraire.
- Si les écrous sont difficiles à dévisser, mettre un peu d'huile pénétrante, attendre quelques minutes, essayer de nouveau.

Original Text

Desserrage des écrous
Utilisez une clé en croix pour desserrer les écrous de la roue. S'ils portent la marque L, ils ont le filetage à gauche et vous devez les dévisser en tournant dans le sens des aiguilles d'une montre. Pour les écrous qui ne portent aucune marque, tournez en sens contraire. Si les écrous sont difficiles à dévisser, mettez un peu d'huile pénétrante, attendez quelques minutes puis essayer de nouveau. N'enlevez pas complètement les écrous.

Translation of the original text

Loosening of the nuts
Use a cross-bar to loosen the nuts of the wheel. If they have an L mark, they have a left-hand thread and you must unscrew them by turning clockwise. For nuts that have no mark, turn counter-clockwise. If the nuts are difficult to unscrew, put a small amount of penetrating oil, wait a few minutes, then try again. Do nut remove the nuts.

Figure 12: Example of a text generated by SPIN

strategies. However, although both steps are performed simultaneously, the discourse strategies distinguishes the semantic and rhetorical operators, similarly to SPIN. We believe, that a major contribution of SPIN is the explicit identification and specification of the criteria used to select, on the one hand, the semantic carriers, and on the other, the rhetorical relations. Through this work, regardless of the text planning technique used in the generation process, each criterion can be evaluated to identify the best semantic carrier and rhetorical relation to be used.

Although no formal test as been done to compare SPIN and DRAFTER's French outputs, we believe that they are quite comparable, as both systems are based on the same linguistic observations.

6.2 Comparison with natural texts

We also took 3 texts (Figure 12 shows one of them) outside our corpus of analysis and generated them by SPIN, then compared, on a qualitative basis, the two versions. If we compare the content and rhetorical structure of the texts in Figure 12 we can see for each aspect:

semantic content From the semantic point of view, the text generated by SPIN is the same as the original text except for one semantic carriers. The REQUIRED OPERATION *vous devez les dévisser* found in sentence 3 of the original text, was not communicated by SPIN. This is because SPIN was told that the operation was not a basic-level one. In light of this, the system did not deem it useful to communicate it in an execution-oriented text.

rhetorical structure From the rhetorical point of view, SPIN did not always pick the same rhetorical relations present in the original text. The rhetorical relations that are different are:

1. A *means* is used in sentence 2 of the generated text:

   (30) Desserrer les écrous avec une clé en croix.
   Loosen the nuts with a cross-bar.

   while the original text used a *purpose*:

   (31) Utilisez une clé en croix pour desserrer les écrous de la roue.
   Use a cross-bar to loosen the nuts of the wheel.

   The underlying semantic carrier in both sentences a GUIDANCE, as the use of a cross-bar *guides* the agent in determining how to loosen the nuts. The heuristic **GD-1a** was responsible for the choice of the relation of *means* in that case as the use of a particular instrument was communicated. According to our corpus analysis, SPIN's choice is the most common one made, and to us, does not seem odd at all.

2. A *c-condition* is used in sentence 4 of SPIN's version:

   (32) S'ils ne portent aucune marque, tourner les écrous en sens contraire.
   If they have no mark, turn the nuts counter-clockwise.

   while a *purpose* is used in the original text:

   (33) Pour les écrous qui ne portent aucune marque, tournez en sens contraire.
   For nuts that have no mark, turn the nuts counter-clockwise.

   In both sentences, a MATERIAL CONDITION is communicated. the heuristic **MC-3** was responsible for SPIN's choice. Recall that in the case of material conditions that pertain to the nature of a device, **MC-3** suggests the use of a *purpose* or a c-condition, but has no preference. SPIN thus picked randomly the relation of *c-condition*, which in our opinion sound very natural.

[Kosseim, 1995] has analyzed two other texts that show that the output of SPIN is comparable to the ones found in the current instruction booklets.

6.3 Global evaluation of SPIN

The most important contribution of SPIN is the explicit separation of two important questions in textual planning. The *semantic level* corresponding to the "What to say?" problem and the *rhetoric level*, the "How to organize it" problem, are considered separately while being linked in many ways. These two levels had not been considered separate before. For example, in a schema or RST based approach for textual planning, these questions are more tightly linked which can bring some problems especially in multilingual generation where rhetorical choices might differ between languages.

SPIN's heuristics are based on a thorough linguistic corpus study and thus do take into account many interesting phenomena of the sublanguage of instructional texts.

SPIN is a complete system that goes from a high level description of a task to well formed short (about 12 relations) French texts instructing how to perform the task. The longer texts we have analyzed are linked by a sequence relation and ruled by the same global textual rules. We believe that we can reuse the selection and presentation heuristics described here for longer texts but we would need a strategy for separating a global task into subtexts to be generated individually.

7 Conclusion and Further Research

This article has presented the results of an investigation and the implementation of a system for the generation of instructional texts. We take the view that the planning of instructional texts must be a 2-stage process: selecting its semantic content then its rhetorical structure. Indeed, one semantic carrier can be presented through different rhetorical relations, while the same rhetorical relation can carry different semantic contents. This article has emphasized the results of this second step: how to select the most appropriate rhetorical relation in French instructions. First, we have introduced the 9 semantic carriers typically found in instructional texts, and the 7 rhetorical relations used to present them. The presentation heuristics have then been specified for each semantic carrier in order to show how the most appropriate rhetorical relation can be chosen automatically. These heuristics are based on the notion of basic-level operation and 5 types of constraints. The SPIN system, implementing the heuristics, has then been presented in order to validate the research.

Among the questions raised by this research, one can consider the influence of the discourse domain and of the language of communication on the selection of the content and the structure. For example, cooking recipes have a higher percentage of REQUIRED OPERATIONS and *sequences* than instructions from other domains. We have attributed this phenomenon on the communicative goal of the text (execution text versus comprehension texts), but even among execution texts, cooking recipes have a particularly

stereotypical content and structure. The discourse domain may in fact have a lot to do with this.

The language of communication and cultural context may also play a role in the choice of textual content and structure in instructions. These factors may very well influence the semantic content of instructional texts as different cultures have a different level of knowledge on a particular domain and have different sets of values. Warnings and safety issues, that are required to be communicated in North-American instructions, may seem frivolous in another country where the culture and legislation is different. The number of PREVENTIONS, for example, will therefore be influence by this factor. As to the structure of the texts, [Delin et al., 1993] have already noted that in multilingual instructions, the same content may be presented by different rhetorical means according to the language of communication. These very interesting issues remain to be investigated.

Acknowledgment

The work presented here is based on the first author's Ph.D. thesis. We are greatly indebted to the thesis' examiners: Cécile Paris, Richard Kittredge, Jacques Malenfant and Paul Bratley whose invaluable comments greatly improved the text. We also thank the anonymous referees for their valuable comments on a previous version of this paper. This work was supported by a scholarship and grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[Bates et al., 1994] Bates, L., Hovy, E., and Senneff, S. (1994). (invited panelists). Panel discussion: Evaluating Generation Systems, 7^{th} International Workshop on Natural Language Generation. unpublished.

[Dale, 1992] Dale, R. (1992). *Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes.* ACL–MIT Press Series in Natural Language Processing. The MIT Press.

[Delin et al., 1994] Delin, J., Hartley, A., Paris, C., Scott, D., and Vander Linden, K. (1994). Expressing Procedural Relationships in Multilingual Instructions. In *Proceedings of the 7^{th} International Workshop on Natural Language Generation*, pages 61–70, Kennebunkport, Maine.

[Delin et al., 1993] Delin, J., Scott, D., and Hartley, T. (1993). Knowledge, Intention, Rhetoric: Levels of Variation in Multilingual Instructions. In Rambow, O., editor,

Proceedings of the ACL Workshop on Intentionality and Structure in Discourse Relations, pages 7–10, Ohio State University.

[Delin et al., 1996] Delin, J., Scott, D. R., and Hartley, A. (1996). Language-specific mappings from semantic to syntax. In *Proceedings of the sixteenth International Conference on Computational Linguistics*, pages 292–297, Copenhagen, Denmark. COLING.

[Di Eugenio, 1993] Di Eugenio, B. (1993). *Understanding Natural Language Instructions: A Computational Approach to Purpose Clauses.* PhD thesis, University of Pennsylvania.

[Di Eugenio and Webber, 1995] Di Eugenio, B. and Webber, B. L. (1995). Pragmatic overloading in Natural Language instructions. *Submitted for publication.* Available through the authors.

[Dixon, 1987] Dixon, P. (1987). The Processing of Organizational and Component Step Information in Written Directions. *Journal of Memory and Language*, 26(1):24–35.

[Dixon et al., 1988] Dixon, P., Faries, J., and Gabrys, G. (1988). The Role of Explicit Action Statements in Understanding and Using Written Directions. *Journal of Memory and Language.*

[Donin et al., 1992] Donin, J., Bracewell, R., Frederiksen, C., and Dillinger, M. (1992). Students' Strategies for Writing Instructions, Organizing Conceptual Information in Text. *Written Communication*, 9(2):209–236.

[Goldman, 1970] Goldman, A. (1970). *A Theory of Human Action.* Princeton University Press, Princeton, NJ.

[Hartley and Paris, 1996] Hartley, A. and Paris, C. (1996). Two sources of control over the generation of software instructions. In *Proceedings of the* 34^{th} *Annual Meeting of the Association for Computational Linguistics*, pages 192–199, University of California, Santa Cruz. ACL.

[Kittredge, 1982] Kittredge, R. (1982). Variation and homogeneity of sublanguages. chapter 4, pages 107–137. Walter de Gruyter, Berlin.

[Korelsky and Kittredge, 1993] Korelsky, T. and Kittredge, R. (1993). Towards stratification of RST. In Rambow, O., editor, *Proceedings of the ACL Workshop on Intentionality and Structure in Discourse Relations*, pages 52–55, Ohio State University.

[Kosseim, 1995] Kosseim, L. (1995). *Planification de textes d'instruction: Sélection du contenu et de la structure rhétorique.* PhD thesis, Université de Montréal.

[Mann and Thompson, 1988] Mann, W. and Thompson, S. (1988). Rhetorical Structure Theory: towards a functional theory of text organization. *TEXT*, 8(3):243–281.

[Mellish, 1988] Mellish, C. (1988). Natural Language Generation from Plans. In Zock, M. and Sabah, G., editors, *Advances in Natural Language Generation*, Communications in Artificial Intelligence Series, chapter 7, pages 131–145. Pinter Publishers, London.

[Moore and Paris, 1993] Moore, J. and Paris, C. (1993). Planning Text for Advisory Dialogues: Capturing Intentional, and Rhetorical Information. *Computational Linguistics*, 19(4):651–694.

[Moore and Pollack, 1992] Moore, J. and Pollack, M. (1992). A Problem for RST: The Need for Multi-Level Discourse Analysis. *Computational Linguistics*, 18(4):537–544.

[Paris and Vander Linden, 1996a] Paris, C. and Vander Linden, K. (1996a). Building knowledge bases for the generation of software documentation. In *Proceedings of the sixteenth International Conference on Computational Linguistics*, pages 734–739, Copenhagen, Denmark. COLING.

[Paris and Vander Linden, 1996b] Paris, C. and Vander Linden, K. (1996b). Building Knowledge Bases for the Generation of Software Documentation. In *Proceedings of the Sixteenth International Conference on Computational Linguistics*, volume 2, pages 734–739, Copenhagen, Denmark. COLING.

[Paris et al., 1995] Paris, C., Vander Linden, K., Fischer, M., Hartley, A., Pemberton, L., R., P., and Scott, D. (1995). A Support Tool for Writing Multilingual Instructions. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95)*, volume 2, Montréal, Canada.

[Pollack, 1986] Pollack, M. (1986). *Inferring Domain Plans in Question-Answering.* PhD thesis, University of Pensylvania.

[Rosch, 1978] Rosch, E. (1978). Principles of Categorization. In Rosch, E. and Lloyd, B., editors, *Cognition and Categorization*, pages 27–48. Lawrence Erlbaum, Hillsdale, NJ.

[Rösner and Stede, 1992] Rösner, D. and Stede, M. (1992). Customizing RST for the Automatic Production of Technical Manuals. In Dale, R., Hovy, E., Rösner, D., and Stock, O., editors, *Aspects of Automated Natural Language Generation*, Lecture Notes in Artificial Intelligence, pages 199–214. Springler–Verlag, Trento, Italy.

[Sacerdoti, 1977] Sacerdoti, E. D. (1977). *A Structure for Plans and Behavior*. Elsevier, New York.

[Tutin, 1992] Tutin, A. (1992). *Étude des anaphores grammaticales et lexicales pour la génération automatique de textes de procédures*. PhD thesis, Département de linguistique et philologie, Université de Montréal.

[Vander Linden, 1993] Vander Linden, K. (1993). *Speaking of Actions: Choosing Rhetorical Status and Grammatical From in Instructional Text Generation*. PhD thesis, University of Colorado.

[Vander Linden and Di Eugenio, 1996] Vander Linden, K. and Di Eugenio, B. (1996). Learning micro-planning rules for preventative expressions. In *Proceedings of the Eighth International Natural Language Generation Workshop*, pages 11–20, Sussex, UK.

[Walker, 1989] Walker, M. (1989). Evaluating Discourse Processing Algorithms. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics (ACL-89)*, Vancouver.