Refining Instructional Text Generation After Evaluation

Fiorella de Rosis*
Università di Bari, Italy

Floriana Grasso†
Heriot-Watt University, Edinburgh, U.K.

Dianne C. Berry‡
University of Reading, U.K.

In this paper, we describe how user-adapted explanations about drug prescriptions can be generated from already existing data sources. We start by illustrating the two-step approach employed in the first version of the NL generator, and the limitations of generated texts, that we discovered through analytical and empirical evaluations. We claim, that, although style refinement would be needed in these texts, a particular care should be devoted to implementing some of the persuasion techniques that doctors employ in their explanations. This would require either throughly revising the text planning techniques employed, or passing to a multistep generation architecture. We justify why we selected this second alternative, and propose some heuristics to repair problems found in the first version of the generator. Some final considerations about the advantages of this approach and the possibility of generalizing it to other domains conclude the paper.

Keywords. evaluation, text planning, sentence planning.

1. Introduction

Text generation systems have been developed in many application areas. In the first prototypes, the generation process was made up of a phase of Text Planning, in which information content and order and rhetorical structure were established, followed by a phase of Linguistic Realization, in which the plan was translated into an understandable, coherent message (Hovy, 1993). More recently, the limitations of this approach have become apparent (Jacobs, 1992), and several authors are now investigating how to bridge the gap between the two phases. Methods proposed introduce a phase whose most popular name is 'Sentence Planning' between the previous two (Hovy and Wanner, 1996; Meteer, 1991). In this intermediate phase, 'local' refinement techniques are applied to increase the quality of the generated text. For example: determining the sentence scope, aggregating clauses or elements internal to a clause, defining how to realize rhetorical relations and so on (Rambow and Korelsky, 1992; Dalianis and Hovy, 1996; Wanner and Hovy, 1996; Reiter, 1994). As most systems that have been developed are still in a prototypical phase, not many of them have been evaluated in a real world environment, to assess whether the texts produced correspond to what speakers and addressees would expect from them. Moreover, existing evaluations often consist of an examination of the results driven by linguistic criteria.

* Dipartimento di Informatica, Via Orabona 4, 70126 Bari; `derosis@gauss.uniba.it`
† Department of Computing and Electrical Engineering; `floriana@cee.hw.ac.uk`
‡ Department of Psychology; `d.c.berry@reading.ac.uk`

It is our opinion that, although there is a real need for style refinement in the generation of a complex text, other communicative requirements may take priority in specific cases. For example, in instructional texts, which are the subject of this paper, the persuasion strength is essential to ensure that explanations achieve their goal: all generation phases (includeing sentence planning) should therefore be aimed at making texts really convincing to the addressee.

In this paper, we describe our experience in a specific application domain: generation of explanations about drug prescription. Evaluation of generated texts, in this context, by a two-phase process (planning and linguistic realization) demonstrates once again that an intermediate phase is needed; such a phase involves revision of the plan at both 'local' and 'global' levels and should be guided by persuasion rather than, or in addition to, style improvement purposes. The local and global plan revision techniques that we propose are represented in a set of rule-based heuristics, which we think can be easily generalized to other instructional texts; an alternative to this multi-step approach would require a revision of text planning methods and knowledge representation which still appears to be theoretically difficult.

The work reported in this paper is the continuation of a European Research Project, OPADE (de Zegher et al., 1994), which started in 1991 and was aimed at prototyping a decision support system in the domain of drug prescription. A text generator was designed as part of the Project, from a corpus of explanations provided by doctors. Generated texts were submitted to informal and formal evaluations, which demonstrated the potential value of the explanation system but also revealed a number of methodological limitations. In this paper, after describing the structure of our generator (Section 2), we examine these limitations (Section 3) and the features of the generator which are responsible for them (Section 4). In illustrating the new version of the generator (Section 5), we describe how we could solve some of the problems simply by revising plan operators and linguistic realization methods, whereas others required more complex *Plan Revision* techniques. In Section 6, we describe these techniques by showing examples of results and by contrasting them with related research. We finally say something about the current state of the system (Section 7) and discuss whether the proposed approach could be generalized to other instructional texts (Section 8).

2. Explanation Generation in OPADE

OPADE is a 'watch-dog' decision support system to be employed in primary and hospital care (Séné et al., 1995). Like in any medical decision support activity, explanations are an important by-product of this system, as they substantiate its suggestions, by favouring, at the same time, compliance and correct treatment performance (Cawsey, Binsted, and Jones, 1995; Cawsey, Webber, and Jones, 1997; Buchanan et al., 1995). To achieve their goals, these messages have to be formulated in a clear and convincing way, and their arguments have to be adapted to the addressee. An additional desirable requirement for such systems would be to generate these texts by exploiting information from already established sources, rather than requiring ad hoc data collection. OPADE tried and met this requirement, one of its strengths being to consider the decision support as one of the components of a wider medical information system, which included, among other sources, a European Drug Database based on the information which is routinely produced in European Union countries.

2.1 User and task analysis

The main purpose of the explanation sheet is to contribute to ensuring that the prescription is performed, and is performed correctly. Explanations may be addressed to the three participants who contribute, with different roles, to put the prescription in effect: doctors, nurses and patients. The generator, then, had to produce texts with distinct characteristics, for the three types of users. We call *indirect users* (IU) these final users of the explanation, as they do not interact with the system, whereas the *direct user* (DU) is the prescribing physician who validates the texts before delivering them to the IUs.

Text generation design, in OPADE[1], relied on a preliminary user and task analysis and, subsequently, on evaluation. As a first step, we examined the information needs of nurses and patients and tried to find out what information doctors think should be included in explanation texts (Berry, Gillie, and Banbury, 1995; OPADE, 1993). In comparing the results of the two studies, we discovered a gap between the patients' *information interests* and the doctors' *propensity to talk about particular topics*. The most significant example of this discrepancy concerns side effects, which are considered to be 'of high interest' to the patients, whereas doctors tend to omit them or downplay them, probably because they fear that knowing about them might negatively influence patient compliance (Berry et al., 1997). Data collected from the two cited studies were employed to set up the two user models, for the Indirect and the Direct User.

A corpus of explanations was also collected from a group of physicians in U.K. (our representatives of DUs). Those who agreed to take part in the study were asked to make a prescription for a given clinical case and then explain it, orally, to three hypothetical addressees: a patient, a nurse and a colleague. The explanations were recorded on a tape recorder and later transcribed. The corpus of English explanations that were eventually employed for the design of the system included 36 texts, produced by 6 doctors on 4 clinical cases (two cases of angina in a male and a female aged patient, a case of tuberculosis and a case of subacute thyroditis). The transcripts of the explanations were analyzed by using discourse segmentation and tagging techniques (Walker and Moore, 1997) and by applying Rhetorical Structure Theory (RST) (Mann, Matthiessen, and Thompson, 1989). For the purposes of this paper, only explanations addressed to patients are considered: for a more detailed description of the variations among texts addressed to different IUs, the reader may refer to (De Carolis et al., 1996).

Figure 1 shows an example of the transcript of an explanation addressed to a patient, for a case of angina (Mr Fictif is a fictitious name).

This text is representative of the texts collected: the structure of these texts is rather independent of the doctors who produced them; they mainly varied in the degree of detail with which individual items are treated. Though we asked physicians to 'formulate a prescription for a case illustrated in a scenario, and to explain this prescription', all doctors began their explanation by illustrating the health problem, with a large use of 'enlistening techniques': this corresponds to what other studies proved to be an essential condition for making a medical text acceptable to patients (Forsythe, 1995). Only after this introduction, did doctors illustrate their prescription by describing, for each drug, the expected positive effects, the possible negative ones (side effects, contraindications, interactions with other drugs), and the administration details.

2.2 Text planner

The design of OPADE's text generator had to respect the previously mentioned requirements of using existing data sources and matching direct and indirect users' needs at the same time. The first requirement had significant consequences for the availability of data: some of the items that might have been useful for explanation purposes were not available

1 For conciseness purposes, the name OPADE is often used, in the paper, to refer only to the text generator component.

Case description

A 62 year old man complains of central chest pain, radiating to the left shoulder and arm. The pain occurs on climbing steep slopes or more than two flights of stairs. He does not have to do either of these frequently, and can walk at least three miles on the flat without difficulty. If the pain does occur, it will disappear after a couple of minutes rest. He is otherwise very well, and is not unduly worried about his symptoms. Tests confirm coronary disease but no other specific clinical syndrome abnormalities. The patient has been diagnosed as suffering from exercise induced angina in a mild form.

Explanation to patient

Well Mr Fictif, your tests have come back, and they confirm that you have angina. This is a condition where the heart muscle, when you exercise, can't get enough oxygen; it is just like a cramp in your leg and, when you rest and the heart slows down and relaxes, then it gets the oxygen it needs and the pain goes away.

This condition is very mild, as you are able to do a lot of exercise before the pain comes on, but I do think we should treat it, and I'd like to give you three separate drugs for this.

The first drug is ASPIRIN, just ordinary ASPIRIN, which I want you to take four times a day by mouth, and this has been shown to reduce the chance of having a heart attack with your angina. The only possible side effect this may have is to upset your stomach a little bit, and I must check that you don't have an ulcer before I prescribe it to you.

The second drug is called ATENOLOL. This is a beta blocker which you may have heard of. It's probably the best drug for preventing the pain coming on, and you just need to take one tablet in the morning. The only worry I have about prescribing this drug is that, if people have asthma, it can make that worse, and I would ask you whether you had asthma before I would let you take it. It is possible that it may make your arms and legs go a little bit cold.

The third thing I need to give you is a spray, which is to treat the attacks when they occur; you take one or two puffs of this spray under your tongue when the pain comes on. The pain should go away very quickly. If the pain disappears within a minute anyway when you stop exercising, you don't have to use this drug, but you can use it as often or as little as you need.

Figure 1
A transcript of an explanation addressed to a patient for a case of angina.

in our databases, and the values assumed by others were different from those needed. This is, typically, the case of the general drug 'indication', which was sometimes different from the reason why that drug was selected in the specific case (we'll see an example later on, for ASPIRIN). Also, the first requirement considerably constrained the surface generation process. We had to produce linguistically acceptable texts by employing data which ranged from single words to complete clauses or sentences. In addition, we had no control whatsoever over the content of the database fields with respect, for instance, to verbal forms or similar issues. As we knew from the beginning that these bindings would result in various limitations in the generated text content, style and lexical correctness, we decided to concentrate on high level processing rather than on surface generation. To compensate for this limitation, we gave the DUs the opportunity to revise interactively the generated text, before delivering it to the IU. There is support for this decision from Reiter, who claims that "... much of the real value-added of many NLG systems may be in the high-level processing, not in ensuring correct syntax..." and that all real world NLG systems that he is aware of allow the human user to modify the generated text (Reiter, 1995).

OPADE's explanation component has the 'classical' architecture of first generators: a text planner and a surface generator, with the peculiarity of two User Models (see

Fig. 2). Planning aims at including in the text what is important for at least one of the users, though controlling the document complexity.

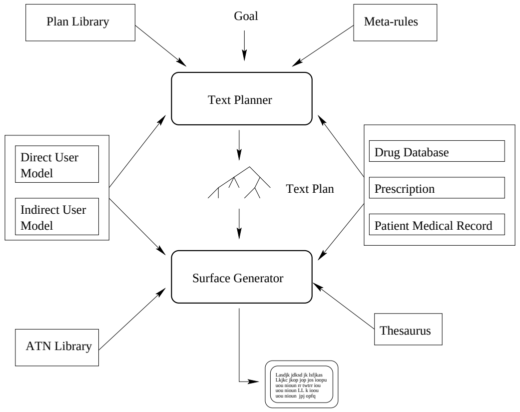

Figure 2
OPADE's text generator architecture.

The design of the plan operator library followed the RST analysis: each operator represents a rhetorical relation applied between a nucleus and one or more satellites. However, the planning process is driven by intentional rather than rhetorical knowledge. An example of an operator is given in Fig. 3: the Header slot represents intentional knowledge, and guides the goal expansion. Constraints on the applicability of the operator help selecting among concurrent operators; they are classified either as Domain or Intentional, though, a part from their semantics, there is no substantial difference in how the two categories are dealt with, and they were separated only for optimization purposes. The Decomposition slot represents how the goal is decomposed: the subgoals assume the roles of nucleus and satellites of the relation expressed in the Rhetorical Relation (RR) slot, according to the RST definition of the relation. Also, each satellite may be labelled either as 'obligatory' or 'optional', all nuclei of the RRs being obligatory.

The operator in Fig. 3, for example, specifies that a speaker (`S`) may include the definition of one of the side effects (`?x`) of a generic drug (`?y`) when the addressee (`IU`) does not know about it; this may be achieved by the speaker informing the addressee about the side effect's sign, whereas its severity, frequency and intensity may complete the description if needed. If at least one of the Satellites is included, it is linked to the Nucleus by a RR of Elaboration Object Attribute. The effect of the successful application of the operator is that the addressee will know about the side effect.

The planner proceeds by first determining a minimal text plan, to which details are added progressively. In the first step, only obligatory nodes are expanded. In the subsequent steps, satellites are expanded according to their level of importance to the users: first, those which are very important for both of them; next, those which are important for at least one of them, the considered level of importance (high, medium, low) decreasing with the increasing level of detail. Notice that, as all the nuclei of every relation are mandatory, the planner produces, from the first step, a 'coherent' discourse, according to Mann, Matthiessen and Thompson's (1989) definition of the 'essence' of


```
-----------------------------------------------------------------------
Header:                    (DefineSingleSideEffect S IU ?x ?y)
Domain Constraints:        (Drug ?y) AND (Side_effect ?x ?y)
Intentional Constraints: NOT (Know_about IU ?x)
Effects:                   (Know_about IU ?x)
Decomposition:             (Inform S IU (Sign ?x))          Nucleus
                           (Inform S IU (Severity ?x))      Sat  opt
                           (Inform S IU (Frequency ?x))     Sat  opt
                           (Inform S IU (Intensity ?x))     Sat  opt
Rhetorical Relation:       ElaborationObjectAttribute
-----------------------------------------------------------------------
```

Figure 3
A plan operator.

a text: the subsequent steps only add details to this essence. After each step, the plan complexity is evaluated, to decide whether to go on. This complexity is a function of the number of 'leaves' generated so far: it is not estimated in absolute terms, but relative to the complexity of the prescription to which it refers. The number of health problems diagnosed and the number of drugs prescribed thus contribute to establishing an 'acceptable' complexity value for the explanation. A small metarule set drives the planning process by formalizing decision criteria on whether to proceed to the next detail step. These metarules enable us, in particular, to decide whether to give more importance to what the DU wants to say (Scrivener mode), to what the IU wants to know (Advocate mode) or to a compromise between the two attitudes (Judge mode). For examples of these metarules refer, again, to (De Carolis et al., 1996). The consequence of our planning method is that there is no unequivocally defined relationship between information available and message content. Some side effects of a drug might be omitted, as well as some secondary details of the administration mode, and so on. Only when the planning process has been concluded is the message content established in a precise way.

2.3 Surface generator

To allow some variability in the texts, though using pre-established phrase fragments, the surface generator was built as a library of augmented transition networks (ATN) (Woods, 1970; Miller, 1984). The exploration of the plan tree, in a depth-first fashion, guides the selection of a network from the library: each RR associated with intermediate nodes corresponds to the start state of an ATN. When a network is invoked, actions associated with the arcs may cause the production of a piece of text or the invocation (`PUSH`) of a new ATN. When a pushed ATN reaches its final state, the control is returned to the one which called it. The leaves of the plan tree, which correspond to basic utterance types, activate ATNs producing the corresponding piece of text, after collecting the information from the appropriate Database.

Adaptation to direct and indirect users is made in this phase too, by associating further conditions on ATN's arcs. For instance, each item can be realized in a 'verbose' or a 'concise' way according to the DU preferences; the language may be defined (in the current version, either English or French); the way leaves are verbalized is customized to the IU's features such as age, sex, level of instruction; the linguistic markers of the RRs are tailored to the depth of subtree of which they are the root, and so on.

An example of the ATN for the CauseEffect is given in Fig. 4. In the figure, soon after being invoked, the ATN `PUSH`es the correspondent network for the analysis of the first subgoal of the relation. When control returns to it, conditions about the category of the subgoal analysed (Satellite or Nucleus) lead to different states of the network, each causing the production of different linguistic markers. Afterwards, the analysis of the second subgoal is performed, which is the ATN's concluding task. In the figure, some conditions on the arcs are omitted for the sake of clarity, and only few phrasing alternatives are presented.

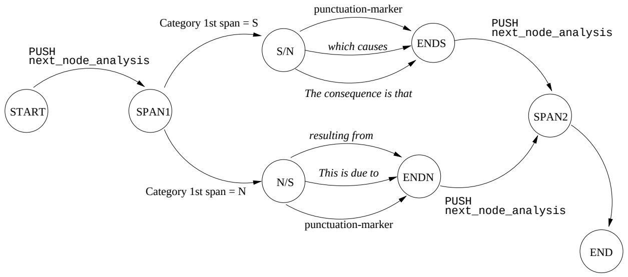

Figure 4
An ATN for the relation of CauseEffect.

2.4 An example of output

An example of the top part of a plan tree is shown in Fig. 5; this plan includes the following main components:

- a *Description of the Patient Status*, which summarizes the main problems that the prescription intends to solve, and is aimed at persuading the patient about the need for treatment;
- a *Request to perform the treatment*, with information about the number of drugs in the prescription; this is a 'leaf' of the plan-tree, and therefore is framed into a rectangle;
- a detailed description of the treatment plan. Drugs are considered in sequence by introducing, for each of them:
    - a *Request to administer the individual drug*, with a description of its characteristics;
    - a segment aimed at persuading that the drug is useful and not harmful, with an *Illustration of its positive effects* (efficacy) and of the *negative* ones (side effects, contraindications, interactions with other drugs).
- a segment aimed at enabling the patient to perform the treatment correctly, with a detailed *Description of drug administration modalities*.

In the figure, goals encircled with an oval represent intermediate goals, expanded in a further subtree, whereas goals in a box represent leaf nodes.

Figure 6 shows an example of text generated from this plan. It is subdivided, for explicative purposes, into discourse segments (DS1,...,DS8) which are the expansion of the encircled goals in Fig. 5, from left to right, with the exclusion of the ExplainContraindication one. A closing sentence (DSf) is added at the end of each explanation. Note also,

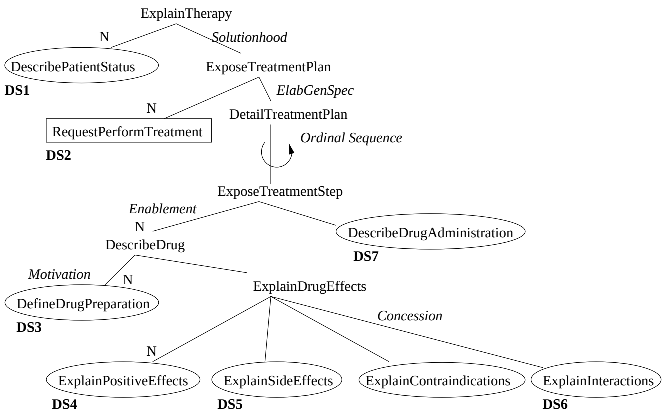

Figure 5
The top part of a plan tree.

in DS1, how the application of the ATN in Fig. 4 caused the linguistic marker *resulting from* to be used to express a relation of Cause-Effect.

Grammatical errors are present in the text: together with the poor style, they are due to the above mentioned problem of using unverified pieces of texts. In addition, as we anticipated in Sect. 2.2, the indication for the prescription of ASPIRIN, which is the standard one contained in the Drug Database, does not obviously apply to this case.

2.5 Related works

OPADE's generator is focussed on producing a text with an acceptable content and rhetorical structure rather than in a linguistically tidy form. As in the majority of Health Care Explanatory Systems, its aim is to reduce the workload of health staff (in our case, direct users) though considering the needs of indirect users.

The first release of OPADE's text generator follows the line of several previous systems (among others Moore and Paris (1992), Maybury (1992), Hovy (1993)). Its most relevant feature is that the considered application field is not a 'toy domain'. The corpus of texts examined is not large: however, differently from other generators (see for instance Vander Linden and Martin (1995)), we collected ad hoc texts rather than examining published documents. Such a task was not simple: some time was spent between the first contact with doctors and the actual collection of the texts, and the analysis of oral material, although the texts showed a rather constant intentional and rhetorical structure, had to deal with a large variety of degree of detail. As mentioned before, the system's strength is the use of existing data sources, which makes it potentially more acceptable to users (Reiter and Osman, 1997); however, this is also the main cause of the limitations in the generated texts, especially from the surface point of view. Nevertheless, as direct users would have taken the responsibility of revising texts before delivering them in any case, we preferred to leave doctors the last word on the generated explanation, by developing a friendly interface with which they can easily make changes to the text content and style.

That does not mean, however, that the system will accept passively the DU's changes:

Comments on the drug prescription for Mr Fictif

DS1 You have been diagnosed as suffering from a mild form of what we call 'angina', that is a spasm of chest resulting from over-exertion when heart is diseased. In addition, you have elevated cholesterol.

DS2 To solve these problems, there are two drugs I would like you to take.

DS3 The first one is Aspirin, which is a analgesic, that is it relieves the pain.

DS4 This drug has been shown to reduce the risk of heart attack and prevents abnormal blood clotting within five minutes.

DS5 The only problem is that this drug can be associated with some side effects. The first one is bleedings, that is a slight hemorrhage; it can be a serious side effect but occurs infrequently. In the unlikely event that this occurs, I suggest you to stop treatment. The second one is allergic reaction; it can be a serious side effect but doesn't happen commonly; it occurs especially when the patient is suffering from what we call 'atopic disease', that is a hypersensitivity to substances having a basis of hereditary predisposition. In case you notice that this problem occurs, you should stop treatment.

DS6 I must also warn you that this drug may interfere with many others. One interaction is with Vitamin K inhibitors; this could cause increased bleeding risk. Therefore, please follow carefully your prescription, do not take other drugs without telling your physician and do not stop treatment without your physician's advice.

DS7 You have to take this drug by mouth; take it with water, at regular intervals. If you miss a dose, do not modify rhythm or quantity of the following ones. Regarding administration of this drug, the treatment steps are the following. The first one is a attack treatment. The dosage is 1gr in 2 intakes for 1 day. You have to take one tablet of 500mg at 8am and one tablet of 500mg at 8pm. The second one is a maintenance treatment. The dosage is 500mg in 1 intake for 7 days. You have to take one tablet of 500mg at 20.

DS8 The second drug I will give you is Glyceril Trinitrate...

......

DSf These are the main treatment suggestions I wanted to give you. If you need any more information, please do not hesitate to ask me directly.

Figure 6
An example of text generated.

by using the intentional structure represented in the discourse plan, the generator has the ability to reconstruct the reasons for its choices and to interpret the request of change as a conflict. It can therefore undertake (if needed) a negotiation dialogue with the DU to justify its choices (de Rosis et al., 1996; de Rosis and Grasso, 1997).

Another distinguishing feature of OPADE's generator is the use of two User Models in which, as in (Kolln, 1995) not only knowledge and belief are represented, but also IUs' 'knowledge interests' and DUs' 'propensity to talk about' and 'style preferences'.

Planning by increasing levels of detail allows maximization of the impact of the message on the addressee by controlling, at the same time, its complexity; this has some similarities to Zukerman and McConachy's (1995) approach to plan avoiding 'boredom' and 'overload', although in their system the DUs are the final message receivers, and may interact with the system to get more details or clarifications, whereas OPADE generates a one shot document.

3. Evaluation

Generated texts were submitted to several types of evaluation. A first, *informal* evaluation was made at the end of the Project, by representatives of OPADE partners and final users of the five countries involved in the Project (Belgium, France, Italy, Swe-

den and UK). People interviewed considered it 'surprising' that different explanation texts addressed to doctors, nurses and patients could be obtained without an ad hoc database organization. Generated texts were judged to reproduce rather faithfully the overall content and rhetorical organization of corpus texts, whereas their style was much less satisfying. Subsequently two more formal studies were performed:

- in a first *analytical* evaluation, the artificial texts were compared once again with the corpus of explanations provided by doctors, focussing, this time, on the identification of the speaker's aims and argumentation techniques;

- in a second *empirical* study, a selection of produced texts was presented to a sample of patients, by varying information content and order, in order to draw some hints on how cognitively and intentionally effective texts should be structured.

The following subsections present the result of the two studies.

3.1 Cues from analytical evaluation

There are several limitations in the generated texts: the 'Describe drug administration' segment, which is aimed at instructing the IU on how to perform treatment, includes overloading repetitions which would provide a rich material to those who study how to produce 'clear and concise' texts. We concentrated, however, on the persuasion power of these texts, and will focus the next sections on this aspect, to detail changes that are demanded to our generator. Although the computer generated explanations reproduce the overall intentional and rhetorical structure of natural messages, several elements contribute to increase the persuasion strength of these messages, which are not reproduced in the artificial ones.

3.1.1 Enlistening techniques in the 'Describe patient status'. The main goal of the explanation text is obtained by linking the two discourse segments *Describe Patient Status* and *Expose Treatment Plan* by a RR of Solutionhood. The Description of the Patient Status opens the explanation: doctors show their 'participation' to the summarized patient problems by adding empathy to sentences:

Ex1: '*Unfortunately* or *I'm afraid* you have an infection...' (when the severity is 'high').

They also try to minimize these problems by de-emphasizing their seriousness:

Ex2: 'you have got *a little bit* of diabetes...' / '*a little bit* of fluid on the chest' / '*a little bit of too much* cholesterol...'

In addition, when the disease is mild and the patient might presume that treatment is not needed, the Solutionhood RR which links the two main discourse segments becomes a Concession, to convey the idea that this is not the case:

Ex3: 'This condition is mild, *but* we have to treat it....'

3.1.2 Persuasion elements in the 'Request to perform the treatment'. The goal *Request to perform the treatment* is achieved by a simple sentence which provides information about the number of drugs in the prescription. In the corpus of texts examined, elements of persuasion are introduced, in this segment, by means of several techniques:

1. *Predisposing favourably the Addressee towards the Request:* by anticipating the positive aspects of treatment (especially when the diagnosis is serious), if treatment has a high expected effectiveness:

Ex4: '*So, the good news* is that we do have tablets that are very effective against TB...' / 'The *good thing* is that we have *many drugs* which can be used to tackle it...' / 'I hope that with the treatment we'll give you, *we can bring all these things back to normal*...'

or if it is, at least, expected to have a positive effect on symptoms:

Ex5: 'We can give you something *to make you feel more comfortable*...' / 'I'm sure *you'll feel better* once you've been put on treatment...'

2. Preventing a non collaborative attitude towards treatment: by anticipating aspects on which a high compliance is needed, especially if treatment is long and complex:

Ex6: 'We have to undertake *quite a long course of treatment*' / 'The problem with this infection, is that it takes *a very long time to eradicate*'

3. Promoting the remembering of significant items: by synthesizing information items which are common to several drugs, *before* detailing the treatment:

Ex7: '...two drugs which are very similar: *they both aim to open the blood vessels up*.' / 'I'm going to ask you to take *two tablets*...' / '...*some tablets* to take *twice a day*...' / '...two tablets *to take regularly*...' / '...and you will have to take them *for some months*'.

or by means of a *final synthesis* of this detailed description:

Ex8: 'So, in summary, we'll give you *tablets* to *help the spasm of the blood vessels*...'

In the previous examples, the communicative goal of *Requesting to perform the treatment* is not directly realized with a "Request" surface speech act, but with a more complex discourse segment which aims at reinforcing the intention and memory of the Addressee.

3.1.3 Persuasion elements in the 'Request to administer the drug'. This purpose is achieved by minimizing about the drug dosage:

Ex9: '...*just one* tablet' / '...one *small* dose of...' / 'The dose is *very small*...we might increase it later' / '...*a little* spray...' / '...*a little* tablet...'

or by emphasizing the use of drugs with which patients are familiar and that they might know to be *harmless*:

Ex10: 'The drug that I'm going to ask you to take is Aspirin, *just ordinary Aspirin*...'

3.1.4 Strengthening or attenuation techniques in the 'Explain side effects of the drug'. This is achieved by minimizing negative aspects, such as the number of side effects:

Ex11: 'The *only worry I have* about prescribing this drug is that...' / 'The *only possible side effect* this may have is to...'

their seriousness:

Ex12: 'your arms and legs go *a little bit* cold' / '...it's to upset your stomach *a little bit*' / '...the spray can *sometimes* make you a *little bit* dizzy...'

or their frequency:

Ex13: '*Occasionally*, they can disturb you in your sleep.'

through detensifier adjectives or adverbs (like in the previous examples) or through aggregation of repetitions:

Ex14: 'these aren't things that happen commonly: most people that take the tablet don't have any problem...'

by emphasizing, on the contrary, positive aspects by means of on purpose repetitions:

Ex15: '*Again*, it's very unlikely to happen' / '*Again*, I would like you to reassure that...'

As to the frequency of the mentioned techniques: minimizations are the most frequent (all doctors use them, somewhere); introductory syntheses are employed in 50% of the cases, as well as emphasizing the positive aspects, in treatment and in side effects. Less frequent (about 30% of the cases) is the introduction of elements of empathy.

3.2 Cues from empirical studies

At the University of Reading, several empirical studies were carried out which examined the effects of varying the type of information given about the possible side effects of prescribed medication, and the order of information within the explanation. Specifically, the studies investigated whether judgements of how good explanations were thought to be, and of perceived likelihood of taking the medication, were affected by whether the medication was said to cause severe as opposed to mild side effects, and whether the information about possible side effects was presented relatively early or relatively late in an explanation. Memory for the presented information was also measured.

In the studies (which are reported in full in Berry, Gillie, and de Rosis (in press)), several texts were prepared, by combining the following components of an OPADE explanation: the patient's health status, the prescribed medication, information about the medication (its class and action), details of drug administration, its side effects and contraindications. The components were generated by the OPADE system. Eight different texts were prepared: two included descriptions of eight relatively severe side effects, two included descriptions of eight relatively mild side effects, two included descriptions of four relatively severe side effects, and two included descriptions of four relatively mild side effects. There were two different orders of presentation of information. All explanations began with information about the patient's health status and the prescribed drug. In one case this was followed by information about side effects, then contraindications, then drug administration. In the other case, information about drug administration preceded information about contraindications, which preceded information about side effects. The actual wording of the resulting explanations was tidied up to make sure that it read smoothly. Nearly 500 participants were tested in the studies.

The first finding of relevance to the present paper concerns the *order of presentation of information*. In general, participants recalled more information about drug administration when this was presented relatively early in the explanation (i.e. before information about the side effects). However, memory for information about side effects depended on both the order of the information and the number of side effects described. When there were only a few side effects, people remembered more about side effects if they came before rather than after the information about drug administration. In contrast, order had no effect when there were many side effects. The implication of this finding is that the optimal order of presentation of the information cannot be established in a rigid way, but depends on the overall explanation content (in this case, the amount of information about side effects).

The second finding from the Berry and colleagues study of relevance here concerns the ratings of 'perceived likelihood of taking the medication' and of 'how good the explanation is', as a function of the severity of the side effects. The experiments showed that people rated the explanations as being less good, and said they would be less likely to take the prescribed medication, when the explanation referred to the occurrence of severe side effects (such as diabetes, hypertension and thinning and crumbling bones) rather than mild side effects (such as tiredness, a rash and increased appetite). In addition, people's perceived need for additional information in the explanation also increased when they were presented with the explanations describing the severe side effects.

If we take the 'memory for information in the text' as a measure of the effect of the instructional text on the cognitive attitude of the addressees, and the ratings of 'perceived likelihood of taking the medication' as a measure of their intention to follow the treatment, we can conclude from these experiments that:

- as the order of information in the text affects the first attitude, one should define this order according to the importance which is attached to remembering each item;
- as the level of detail affects the second one, one should add details according to how important and how difficult it is to convince the patient about the suggested treatment.

4. Limitations of the First Version of OPADE's Generator

A preliminary question is: should we introduce in our artificial texts the same persuasion elements that doctors introduce in their texts? We think we should: artificial texts will be revised by doctors, who would not like a message with a much different tone from their own. On the other side, the Indirect User needs to be convinced even more, if the received explanation has an artificial origin. Scott and Sieckenius de Souza (1990) claim that "since artificial interlocutors clearly have fewer possibilities to make reliable assessments of their audience's ability to 'get the message' than do their human equivalents, then expressions of the message often need to be more explicit than would be ideal".

In approaching the release of a new version of OPADE, the limitations of the first version needed to be considered.

4.1 Limitations of the planning method

As it was said before, the information content and presentation order of the texts are established as a function of data available and users attitudes, through progressive incrementation of a minimal plan. When deciding whether to apply a specific plan operator, the plan expansion module does not know what the overall content of the final text will look like. For example: when the subtree concerning one of the side effects of a specific drug is being expanded, the planner does not know yet how many of the side effects of the drug will be actually mentioned in the explanation, let alone their seriousness, frequency and so on. The discourse structure and its content will be established entirely only when the plan is completed: problems mentioned in Section 3 will only then become manifest. In addition, items in the Decomposition slot of the plan operators are expanded in strict order rather than dynamically: consequently, to introduce context-dependent variability of the order of information presentation, plan operators should be duplicated.

4.2 Limitations of knowledge representation

In complex domains, planning is interleaved with evaluation to assess whether goals are fully satisfied or whether the achieved state contains 'objectionable results': in this case,

a replanning step is needed, to avoid the undesirable effects or, as for Hammond's CHEF (1990), to apply some repair heuristics to the plan. At the end of the planning process, a simulation step produces a detailed final state description: for example in CHEF's domain (cooking recipes) the state description involves the states of ingredients employed, compounds created and final products. This description is not purely qualitative, but is characterized by *quantitative* parameters, like 'savory intensity'. Comparing this description with a detailed description of the goal, enables discovering partial failures and recovering from them by appropriate repair strategies, such as altering event sequences, by inverting the order of two steps, detailing a step into several ones, adding new steps by an adjunct planning, or substituting old steps with new ones.

These repair strategies would be extremely useful for OPADE. However, in order to apply them, the discourse goal should be expressed as a detailed description of the intentional and knowledge state of the Indirect User, whereas in OPADE the goal is defined as a 'DU intention' (for example 'Explain Therapy'). A goal fails only when some essential information is not available and, at the end of the planning process, no evaluation of the result in quantitative terms can be made. Changes in the Indirect User's mental state after communicative acts are represented in a purely qualitative way in plan operators. For instance, in the example in Fig. 3, the 'Define single side effect' operator produces the same change in the knowledge status of the IU irrespective of whether satellites in the Decomposition are all expanded or not.

Nevertheless, a similar knowledge representation is common to the majority of discourse planners (see among others (André and Rist, 1995; Moore and Paris, 1992; Cawsey, 1992; Maybury, 1992)), with which OPADE shares the same problems:

1. *main effects* of elementary and complex communicative acts are *not represented adequately*: negative as well as positive effects should be mentioned, and should be described in a quantitative way (degree of persuasion and knowledge, and so on);
2. *side effects of acts are not mentioned;* for example: the 'Describe drug administration' increases the IUs' degree of knowledge about how to perform the action, but may affect, at the same time, their intention to perform the action;
3. *effects* produced by the communicative acts *are not correlated to the situation*; for example: the Request produces, in the Hearer, a lower increase of the degree of persuasion to perform the action if H is 'hostile' towards S, if the action is 'unpleasant', and so on.

These limitations in plan operator representation introduce incorrect hypotheses in the text planning process: subgoals are considered as mutually *independent* or *serializable*[2], whereas it is not difficult to make examples in which these hypotheses do not apply. This is especially true in persuasive texts, as opposed to informative ones: if the knowledge acquisition process can be considered as 'linear', and only 'knowledge overloading' factors might interfere with a correct learning, this is not the case of persuasive argumentation, where, if the text is complex (like in OPADE), the variation of user attitudes is not monotonic. Consider the following situation: I describe to a patient her health problems and tell her that she has to follow some effective treatment (she may well be convinced so far). Then I have to warn her that this treatment has some side effects... and she becomes much less convinced: to compensate this decrease of persuasion, I must introduce new

2 Two subgoals are independent if we can partition the complete operator set into subsets such that the operators in a given subset are only relevant to a single subgoal. A set of subgoals is serializable if there exists an ordering among the subgoals such that the subgoals can always be solved sequentially without ever violating a previously solved subgoal in the order (Korf, 1987).

arguments, whose strength should be a function of her health problems' seriousness and of treatment effectiveness.

To treat the graded nature of the persuasion process, we should represent degrees of belief in the IU model, and should attach strengths of argument to information items, as proposed by Reed, Long, and Fox (1996). Examples of strengths are 'numerical measures of the impact of an utterance on the Hearer' in (Haimowitz, 1991) and 'quantitative fuzzy measures' (a combination of 'support' and 'plausibility' of a proposition) in (Zukerman and McConachy, 1995). If modeled this way, the plan enrichment process involves some scalar value of 'mounting evidence' or 'increasing conviction', and stops when the system considers that the Hearer has been reasonably informed or persuaded. There are, however, several difficulties in applying such a method to argumentation to convince someone to perform an action: 'degrees of belief' become 'degree of desire or intention', and a sophisticated theory of persuasion is needed, to handle these quantitative measures, similar, for instance, to the utility theory proposed by Walker and Rambow (1994), able to deal with the cases of decreasing, other than increasing, strength.

Reed and colleagues renounce treating this variation, claiming that a quantitative theory of the persuasion process (how agreeing or disagreeing arguments combine) has not yet been established. And also from the discourse structure point of view, Walker and Rambow say that "when the text to be generated is more complex than the brief single-sentence utterance of a dialog system, hierarchical structure of the text interacts with attentional state in ways that have not been fully explored in the literature today". These authors conclude that their model "does not appear to be easy applicable to complex domains, mainly because it requires including in the U(ser)M(odel) features which are difficult to estimate, either subjectively or objectively" (1994).

4.3 Limitations of the surface generator

In contrast to the majority of linguistic realizers, which employ sophisticated surface generators, such as Penman (Penman, 1989), after setting up an ad hoc database, our surface generation method is based on ATNs and is built on an existing information system. Using pre-existing data sources considerably limits grammatical and lexical choices and reduces the possibility of introducing elements of empathy, emphasizing–de-emphasizing techniques and 'user-tailored' definitions of medical terms. We should stress here that the crucial issue is not the use of canned texts rather than Penman-like surface generators: people working at the HealthDoc project (DiMarco, Hirst, and Hovy, 1997) demonstrated that qualitatively good texts can be produced by relying almost exclusively on sentence planning techniques applied to pre-authored texts. Nevertheless, their sources are, once again, set up ad hoc for the purpose of generation, whereas OPADE has to do its best with the pieces of textual information that the European Drug Database and the other sources have.

Another characteristic of the surface generation with ATNs is that it is responsible for some tasks that other systems attribute to a Sentence Planning module; for example, the decision of how many items to put in a single sentence, which is implicitly taken by means of conditions on ATNs' arcs deciding about whether to terminate a sentence.

Finally, the lack of a sentence planning module prevents OPADE from making those stylistic improvements that would better render its texts.

5. Towards A New Version Of The Generator

To try and solve some of the problems mentioned in Section 3, and overcome some of the limitations shown in Section 4, the task of releasing a new version of the OPADE's text generator was undertaken.

In order to introduce persuasion elements in the text, a knowledge base was built upon the existing databases to connote data items in terms of their relevance for persuasion and memory. Moreover, the IU model was enriched by representing attitudes that might influence these aspects.

Also, the planning process was improved at various levels: plan operators and ATNs were refined and an intermediate phase between planning and surface realization was introduced, involving modules performing local and global plan revision.

5.1 Improving the organization of knowledge

Information items that should be considered in generating explanations may be categorized in two classes:

1. *relevant-for-compliance* items, such as:

   a: `(Severity ? disease ?patient)`,
   b: `(Efficacy-for-prognosis ? drug)`,
   c: `(TreatmentDuration ?drug ?patient)`,
   d: `(Frequency-Side-effect ?drug)`,
   e: `(Severity-Side-effect ?drug)`,
   f: `(Intensity-Side-effect ?drug)`

   These items may take 'unfavourable' values, such as:
   `(a: SEVERE, b: LOW, c: LONG, d: FREQUENT, e: SEVERE, f: STRONG)`
   or 'favourable' ones, such as:
   `(a: MILD, b: HIGH, c: SHORT)`;

2. *relevant-for-correct-treatment* items, such as:

   g: `(AdministrationForm ? drug)`,
   h: `(IntakeModality ?drug)`,
   i: `(Dosage ?drug ?patient)`,
   j: `(NumberOfIntakes ?drug ?patient)`

All these functions can be evaluated from OPADE's databases: those of the type: `(<attribute> ?disease ?patient)` from the patient record; those of the type: `(<attribute> ?drug)` from the drug database and those of the type: `(<attribute> ?drug ?patient)` from the prescription record.

Different IU's attitudes might influence the activation of persuasion techniques, when the mentioned information items assume favourable or unfavourable values; for example: `(CooperationAttitude ?patient)`, `(EmotionalState ?patient)` or `(Memory-Level ?patient)`.

These data are not currently available, but default values can be obtained by reasoning about other items in the clinical record such as age, sex, disease history and so on.

5.2 Enriching the plan operators library

New operators enriched the former OPADE's library to deal with some of the problems mentioned in Section 3.

For instance, to 'convince the patient that treatment is needed when the disease is not serious, we just add to the plan library a new operator whose Header is the 'main goal' and whose RR is a Concession: this RR produces a 'but' linguistic marker rather that an 'in order to solve this problem', like in the Ex3, Section 3.2. The planner selects the operator which best suits the context by verifying the Constraint slot, which, in the case of a Concession, includes the following condition:

`(Current-Patient ?p)}` $\wedge$
$\forall$ `?d ((Diagnosis ?d)}` $\wedge$ `(Diagnosed ?d ?p)):` $\neg$ `((Severity ?d ?p) SEVERE))`

whereas, in the case of a Solutionhood, includes:

`(Current-Patient ?p)` $\wedge$
$\exists$ `?d ((Diagnosis ?d)` $\wedge$ `(Diagnosed ?d ?p)): ((Severity ?d ?p) SEVERE))`

5.3 Enriching the planning meta-rules set

New heuristics were introduced to guide the planning by detail levels; these heuristics vary the degree of detail in the explanation of the patient's health status or of the drug side effects according to how serious they are (see results of the empirical evaluation study). For instance, the following rule can be used to expand to a further level of detail, even though both DU and IU consider, in principle, this information as irrelevant:

R1 IF one of the side effects of the prescribed drugs is severe AND the patient is anxious, THEN add details about long term effects of side effects and ways to prevent or alleviate them.

5.4 Revising the ATNs library

To introduce empathy elements in the text, we modify the generation of individual sentences by applying heuristics of the following type (after each rule, the example from Section 3 that the rule is meant to solve is expressed):

R2 IF the subtree 'Describe the patient status' includes 'relevant for compliance' items taking unfavourable values THEN add empathy to description of these items. (Ex1)

R3 IF the subtree 'Describe the patient status' includes 'relevant for compliance' items taking unfavourable values AND the patient is anxious THEN deemphasize these items. (Ex2)

R4 IF the patient is anxious THEN add empathy in the 'Request to administer the individual drug'. (Ex9)

R5 IF the subtree 'Explain positive effects of a drug' includes 'relevant for compliance' items taking favourable values THEN emphasize these items. (Ex4)

R6 IF the subtree 'Explain negative effects of a drug' includes 'relevant for compliance' items which take unfavourable values, AND (the patient is anxious OR non-cooperative), THEN de-emphasize these items. (Ex12)

These heuristics are similar to those employed by Haimowitz to produce 'emotionally empathetic responses' (1991). There is, in fact, a surprising similarity between our data and Haimovitz's examples of 'hypothetical dialogue', such as the use of '*only* three days' to downplay the length of hospital stay, or '*unfortunately*, the Bactrim does leave a rash'. Heuristics are implemented by introducing, on ATNs' arcs, conditions on the IU attitudes and on the domain KB, which establish when and how the sentence accent should be modified by detensifier or intensifier adverbs or adjectives. Example: one of the ways of adding, to a surface speech act of `(Inform S IU (Diagnosed ?d ID))`, some empathy and some de-emphasizing of unfavourable aspects produces the following sentence:

> *I'm afraid you've been diagnosed as suffering from a mild form of what we call angina (...)*

instead of the sentence in DS1 in Fig. 6.

Similar tricks can be applied to the ATNs that inform about test results ('a little bit of cholesterol'), recall symptoms or diseases ('a mild form of angina') or describe treatment ('a little tablet') or side effects of a drug ('a little bit cold').

5.5 Introducing a plan revision phase

After carefully considering pros and cons about the planning issue, we decided to apply *plan revision techniques* rather than revising the planning method. This choice is also supported by complexity considerations, as Hovy (1990) suggests: aggregate repetitions by post text-planning grouping rules rather than before structure planning, has the advantage that, while in the second approach, the algorithm should "inspect every pair of input elements for each aggregation rule" (with an order of n^2 operations per rule), "if performed after structuring, it needs to inspect only the pairs of elements within each leaf of the paragraph structure".

The plan revision modules can be classified either as local or global, depending on the portion of the plan tree that is involved. Here follows a brief description of the three tasks that were undertaken, while a more detailed description of the modules is delayed until the next section.

5.5.1 Treating repetitions. This is one of the main problems faced by the new, intermediate phase, and constitutes maybe the most widely recognized task the phase of Sentence Planning has to undertake. It is a *local* task, as it requires considering portions of the plan tree with limited depth. As we mentioned in Section 3, treatment of repetition plays a crucial role, for instance, in minimizing concern about side effects, and contributes considerably to the persuasion process.

5.5.2 Highlighting relevant subjects. As we saw in Section 3, introduction of synthesis elements has three main purposes:

- *increase the patient confidence* on the possibility of solving health problems mentioned in the previous discourse segment, before describing treatment;
- *predispose patients to a good compliance,* by anticipating the aspects of treatment which require their active participation;
- *promote correct treatment execution,* by anticipating aspects which are relevant for taking the drug correctly.

The technique we employed first finds out items to be highlighted, and then introduces a new discourse segment before the "Detail Treatment Plan" segment, in which these items are synthesized appropriately. It is a *global* plan revision task, as it needs to explore the whole plan tree, and introduces a new branch into the tree itself.

5.5.3 Reordering text spans. The main finding of empirical evaluation studies is that the order of presentation of Administration Modalities and Side Effects has to be established according to the number and severity of the side effects of prescribed drugs. The text planner is responsible for assuring that the most preferable ordering among items is chosen, but, as for other situations we mentioned before, it could be the case that the text produced does not achieve the mentioned requirements. A reordering task may, in this case, be necessary: it is, again, *global*, because large parts of the tree may be required to move.

6. Local And Global Plan Revision

Local and global plan revision transformed OPADE's generator from a two-step architecture (text plan + surface realization) into a multistep one. Heuristics employed in these tasks are *tree-rewriting rules* which apply to discourse subplans to produce new subplans. The *left-hand sides* of rules are logical combinations of conditions on the communication goal and the RR associated with the subplan root, the semantic properties of information associated with its leaves and the Addressee's attitudes. Their *right-hand sides* correspond to tree-restructuring algorithms, which vary according to the task to be performed.

Some definitions need to be given first. Let $L = \{l_1, \ldots, l_n\}$ be a collection of leaves of a plan tree T, then:

Definition 1: the *minimal subtree producing* L (MS_L) is obtained by tracing up T, starting from all the leaves belonging to L, until a unique node of T is reached, and then considering the whole subtree originating from that node. Notice that MS_L will produce, in general, other leaves that are not included in L.

Definition 2: the *pruned minimal subtree producing* L (PMS_L) is obtained by marking intermediate nodes of MS_L as follows:

1. let $N_k = \{n_1, \ldots, n_m\}$ be the set of nodes to mark at step k (where $N_0 = L$);
2. mark in MS_L each element belonging to N_k;
3. repeat until N_k is empty:

   3.1. for each element n_i of N_k:

   3.1.1. let F_{n_i} be the parent node of n_i;
   3.1.2. mark F_{n_i} in MS_L;
   3.1.3. if n_i was labelled as satellite in the RR associated with F_{n_i}, then mark, in MS_L, the most nuclear part[3] of F_{n_i};
   3.1.4. add F_{n_i} to N_{k+1};

   3.2. eliminate from N_{k+1} all repetitions and all elements equal to the root of MS_L and go to step 3;

4. prune out, from MS_L, all unmarked nodes;
5. trace down MS_L from its root, by pruning out all intermediate nodes which have only one child, and by connecting the initial to the final node in each simplified path, so as to restore connectivity in the tree.

The set of nodes in MS_L obtained after this procedure make up the *pruned minimal subtree producing* L, that is the minimal plan tree which insures generating a coherent text that includes all nodes in L.

Having completed the description of preliminaries, we are now ready to detail the three algorithms.

3 The *most nuclear part* of a tree is calculated as in (Mann and Thompson, 1988).

6.1 Treating repetitions

From examples in Section 3.2, it appears that the decision on whether to aggregate repetitions or to emphasize them should be made by considering the global aim of the sentence, not only its readability; in addition, considerations about the relevance of subjects help establishing the order in which items should be presented to maximize the persuasion effect.

In the discussion below, the following text segment, that was produced by the first release of our generator, is considered:

> *However, I must inform you that this drug may cause some side effects. The first one is nausea; it is serious, it occurs infrequently, in a strong form, in sensitive patients. The second one is headache; it is serious, it occurs infrequently, in a strong form, in sensitive patients. The third one is insomnia; it is not serious, it occurs frequently, in a strong form, in sensitive patients.*

This text includes several repetitions, some of which are concerned with 'positive' aspects, other with 'negative' ones. These repetitions originate from subtrees having a multinuclear RR associated with their root-node (in the example, an Ordinal Sequence). Their treatment may be described by the two following heuristics:

R7 IF the subtree 'Explain negative effects of a drug' includes favourable items with equal values, THEN emphasize them by stressing these repetitions.

R8 IF the subtree 'Explain negative effects of a drug' includes unfavourable items with equal values AND the patient is anxious towards the pain THEN aggregate these repetitions.

Implementing these heuristics requires classifying the objects mentioned in the nuclei according to their characteristics, described in the satellites; this can be made by an algorithm which consists of four main steps:

step 1: find the subtree whose root node is the 'Explain Side Effects'. The RR associated with this node is an Ordinal Sequence, and the sub-trees originated from it have a similar structure, each of them describing a single side effect. In the example, this step extracts, from the plan tree in Fig. 5, the SubTree shown in Fig. 7.

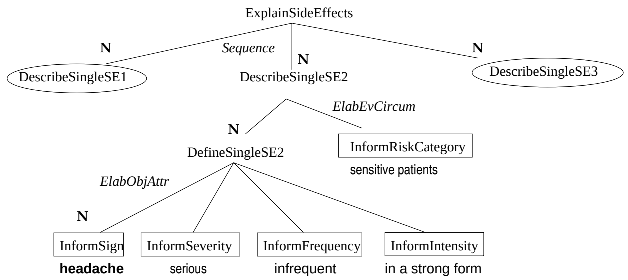

Figure 7
The subtree producing the side effect example.

step 2: classify leaves. The classification is made in two stages:

2.1: identify the *most nuclear part* of each sub-subtree and label these leaves as *objects*; label remaining leaves as *characteristics*.
In the example, there are three nuclei (one for each 'Inform Sign' leaf) whose values are nausea, headache and insomnia. Each of them has four characteristics: severity, frequency, intensity and risk-category.

2.2: classify objects according to the values of their characteristics, and identify characteristics whose values are the same in all objects (uniform characteristics) or only within a class (class-uniform characteristics).
In the example, frequency and severity are class-uniform characteristics; risk-category and action are uniform characteristics.

step 3: decide which characteristics to aggregate and which one to repeat.
In the example, *strong* intensity is an unfavourable item, *infrequent* frequency is a favourable one; the first one is therefore aggregated, the second one is duplicated.

step 4: restructure the subtree:

4.1: prune uniform unfavourable characteristics; add a new super-root (whose name is the name of the original root, (ii) whose attached RR is an *ElaborationObjectAttribute* and (iii) whose children are the original root (renamed) and as many children leaves as are the unfavourable uniform characteristics that have been singled out;

4.2: rearrange the nuclei in the original Ordinal Sequence according to classes of *similar* objects (that is, objects with equal, unfavourable, class-uniform characteristics). For each class which contains more than one object, create a new intermediate node and define an Ordinal Sequence among objects within it;

4.3: apply recursively 4.1 and 4.2 within each class.

Figure 8 shows how the subtree in Fig. 7 is transformed by this algorithm. Using the new ATNs library, empathy can also be added to the surface generation of this segment, producing the following text:

> *However, I must inform you that this drug may cause some side effects. A first group of them includes nausea, which occurs infrequently and only in particularly sensitive patients, and headache which, again, occurs infrequently and only in particularly sensitive patients; these side effects are both serious. Then, you may have insomnia: it is not serious but can be frequent; however, once again I would like to reassure you that it occurs only in particularly sensitive patients. All these side effects can occur in a strong form.*

6.2 Highlighting relevant subjects

Heuristics applied to find out and highlight relevant subjects in an introductory summary are the following:

R9 IF at least one leaf of the plan tree is a 'relevant for compliance' item AND this item takes a favourable value AND the patient is non-cooperative OR his/her memory-level is low, THEN substitute the 'Request to perform treatment' with a sentence which emphasizes the most significant of these items.

R10 IF at least one leaf of the plan tree is a 'relevant for correct treatment' item AND the patient is aged THEN substitute the 'Request to perform treatment' with a sentence

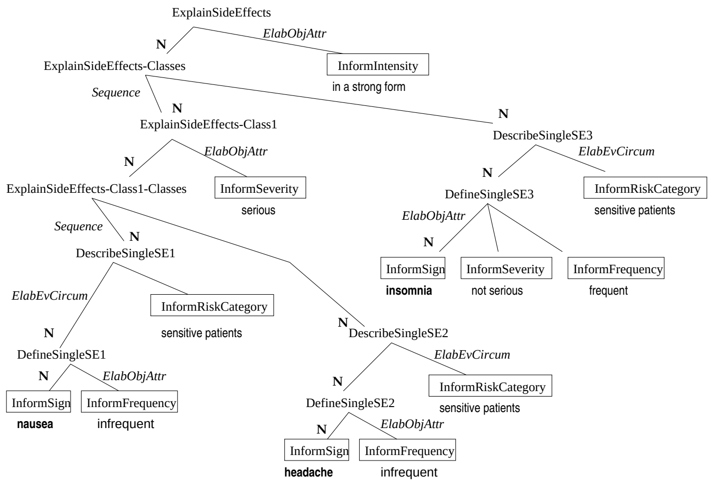

Figure 8
The subplan tree of Fig. 7 after the treatment of repetitions.

which introduces the most significant of them, by aggregating those which take the same values

An outline of the algorithm implementing these heuristics is:

step 1: identify 'relevant for compliance' items in the plan tree: for each subtree of the 'Detail treatment plan' subtree (which describes an individual drug) find out the leaves which are relevant for compliance and assume a favourable value; these will be typically found among the leaves of the ExplainPositiveEffects subtree. Some examples are `((EfficacyForPrognosis ?drug) HIGH)` or `((Efficacy-ForSymptoms ?drug) HIGH)`. Insert these items in a PERSUASION LIST;

step 2: identify 'relevant for correct treatment' items in the plan tree:

2.1: classify leaves of the 'Detail treatment plan' tree which are relevant for correct treatment according to their values. These will be leaves of the DescribeDrugAdministration subtree. For example: `(Administration-Form ?drug)`, `(IntakeModality ?drug)`, and so on.

2.2: find out leaves with the same value: these leaves will describe equal characteristics of administration for different drugs;

2.3: insert these values in a MEMORY LIST;

step 3: identify the list L_H of items to be highlighted, by selecting a few items from both PERSUASION and MEMORY lists;

step 4: introduce a highlighting summary: substitute the Request to Perform Treatment leaf with PMS_{L_H}, the pruned minimal subtree producing L_H. After applying the *Treating repetition algorithm*, this subtree will produce, in the surface generation phase, a summary sentence in which the main items relevant for compliance and for correct treatment will be highlighted.

To illustrate results of this algorithm, let us consider the subtree in Fig. 9(a), which is an expansion of Fig. 5, and let us imagine that the second drug (which is not described in this figure) is very effective, like the first one (see DS4) and has to be taken by mouth and with water, again, as the first one (see DS7). In this case, the highlighting algorithm (followed by some *adding empathy* in the surface realization) will produce a pruned minimal subtree which is made up of two identical components, each being described in Fig. 9(b). This will enable us substituting, the sentence in DS2:

> *To solve these problems, there are two drugs I would like you to take.*

with the following sentence:

> *The good thing is that, to solve these problems, we have some very effective drugs, which have to be taken by mouth, with water.*

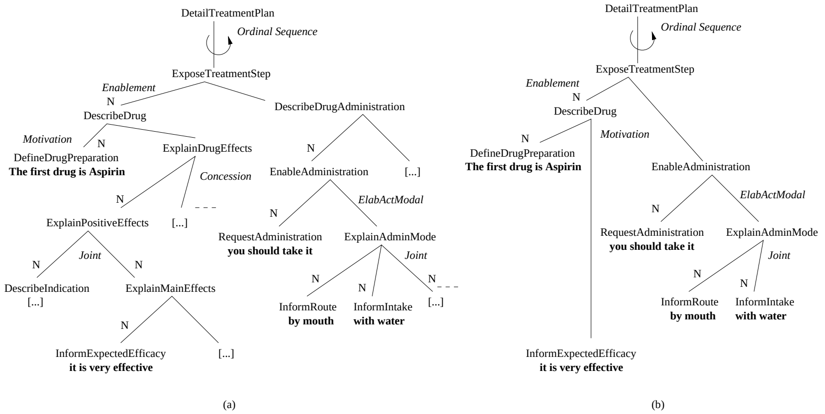

Figure 9
Extraction of the highlighting subtree.

6.3 Reordering text spans

To define when the order of two discourse segments of our text has to be changed, we apply some heuristics of the type:

R11: IF the side effects mentioned are many and most of them are serious THEN make sure that the administration details are mentioned before the side effects of the drug.

Implementing heuristics like this requires that two subtrees in the plan tree exchange their places. Although our heuristics are clearly domain dependent, we tried and made the swapping algorithm on the right side of the rule more general; to assure that the text produced after swapping is still coherent, the algorithm is not a mere prune and paste of the subtrees involved.

Some assumptions are needed before describing it:

Assumption 1: a RR existing between two large spans[4] of text also holds between the *most nuclear parts* of the two spans. A similar hypothesis can be found in (Marcu, 1996).

Assumption 2: if a text span is *the most nuclear part* of two different RRs, then the two RRs can be exchanged, and so their satellites, without loosing coherency. Some consideration about the strength of this assumption will be presented at the end of this Section.

Assumption 3: two text spans S_i and S_j are exchangeable if:
- they do not overlap;
- the minimal subtree generating each of them does not generate other leaves outside the spans themselves.
For instance, in Fig. 7, $S_i = \{InformSign + InformSeverity\}$ does not constitute a valid span, whereas adding $InformFrequency$ and $InformIntensity$ to it would meet the requirement.

Moreover, two *basic operations* on the text tree have to be defined.

Inversion of siblings: Let T be a subtree whose root node has the children $N = \{n_1, \ldots, n_m\}$; let $N_i = \{n_{i_1}, \ldots, n_{i_k}\}$ and $N_j = \{n_{j_1}, \ldots, n_{j_h}\}$ be two non overlapping subsets of N such that N_i *precedes* N_j (that is there exists the order: $\{\ldots n_{i_1} \ldots n_{i_k} \ldots n_{j_1} \ldots n_{j_h} \ldots\}$). Then an *inversion of the siblings* N_i *and* N_j *in* T (written as $\mathbf{Inv}(T,N_i,N_j)$) replaces T with a new subtree T' in whose root N_j *precedes* N_i (that is there exists the order: $\{\ldots n_{j_1} \ldots n_{j_k} \ldots n_{i_1} \ldots n_{i_h} \ldots\}$).

Exchange of satellite children: Let T_1 and T_2 be two subtrees, RR_1 and RR_2 the RRs associated with the roots of T_1 and T_2 respectively, Sa_1 and Sa_2 the sets of satellite children, and Nu_1 and Nu_2 the sets of nucleus children of the roots of T_1 and T_2 respectively. Then an *exchange of the satellites* Sa_1 *and* Sa_2 *between the two subtrees* T_1 *and* T_2 (written as $\mathbf{Exc}(T_1,T_2,Sa_1,Sa_2)$) replaces T_1 and T_2 with two new subtrees T_1' and T_2' such that:
- T_1''s root has Nu_1 as nucleus children and Sa_2 as satellite children, among which the rhetorical relation RR_2 holds;
- T_2''s root has Nu_2 as nucleus children and Sa_1 as satellite children, among which the rhetorical relation RR_1 holds.

We can now describe the *plan reordering* algorithm. Let $SPAN_1 = \{l1_1, \ldots, l1_n\}$ and $SPAN_2 = \{l2_1, \ldots, l2_m\}$ be two text spans to be exchanged in the tree T.

step 1: let $L_{12} = SPAN_1 \cup SPAN_2$ be the union of the two leaf sets: identify the minimal subtree $MS_{L_{12}}$ generating L_{12}. Let N_{12} be the root node of $MS_{L_{12}}$ and RR_{12} the RR associated with N_{12}.

step 2: Two situations may occur:

case 1: RR_{12} **is mono-nuclear.** Identify MN_{12}, the minimal subtree producing the most nuclear part of the subtree $MS_{L_{12}}$. Let $L_{MN} = \{l_1, \ldots, l_m\}$ be the leaves generated by MN_{12}.

case 1.1: L_{MN} **has empty intersection with** L_{12}. Let T_1 be $MS_{SPAN_1 \cup L_{MN}}$, the minimal tree generating both $SPAN_1$ and

4 In this context a *span* is a collection of adjacent leaves of a text tree.

L_{MN} and T_2 be $MS_{SPAN_2 \cup L_{MN}}$, the minimal tree generating both $SPAN_2$ and L_{MN}. Let RR_1 and RR_2 be the RRs associated with the roots of T_1 and T_2 respectively, and Sa_1 and Sa_2 the sets of satellite children of these roots. Then execute **Exc**(T_1,T_2,Sa_1,Sa_2).

case 1.2: L_{MN} has non-empty intersection with L_{12}. Because of Assumption 3.1., it will then have a non empty intersection with either $SPAN_1$ or $SPAN_2$ (not with both of them). Then, if N_1 and N_2 are, respectively, the sets of children of $MS_{L_{12}}$ generating $SPAN_1$ and $SPAN_2$, execute **Inv**($MS_{L_{12}}$,N_1,N_2).

case 2: RR_{12} is multi-nuclear. Treat as case 1.2.

The reader may verify that, because of the assumptions made, no other case is possible.

Let us illustrate how the algorithm applies to an example taken from Fig. 7, which produces, for the second side effect, the sentence: *The second one is headache; it is serious, it occurs infrequently, in a strong form, in sensitive patients.*

case 1.1: Let $SPAN_1 = \{InformSeverity\}$ and $SPAN_2 = \{InformRiskCategory\}$. In this case, the root of the minimal subtree $MS_{L_{12}}$ is $DescribeSingleSE2$, and the RR associated with it is mono-nuclear (ElabEvCircumstance). The main nucleus is just the leaf $InformSign$, so it has empty intersection with both the spans. The exchange of satellite children will create a node $DescribeSingleSE2'$ associated with a RR *ElabObjAttr* whose nucleus is $DefineSingleSE2$, and whose satellites are $\{InformSeverity, InformFrequency, InformIntensity\}$. The nucleus of this RR is, in turn, replaced by a node $DefineSingleSE1'$, associated with a RR *ElabEvCircum*, whose nucleus is $InformSign$ and whose satellite is $InformRiskCategory$. The sentence produced by the new tree will be *The second one is headache; it occurs in sensitive patients; it is serious, it occurs infrequently and in a strong form.*

case 1.2: Let $SPAN_1 = \{InformSign\}$ and $SPAN_2 = \{InformRiskCategory\}$. The minimal subtree's root is, again, $DescribeSingleSE2$, and the RR associated with it is mono-nuclear. The main nuclus is just the leaf $InformSign$ which has a non empty intersection with $SPAN_1$. In $DescribeSingleSE2$, the child generating $SPAN_1$ is $DefineSingleSE2$, and the child generating $SPAN_2$ is $InformRiskCategory$. These two children are inverted in a new $DescribeSingleSE2'$, which will produce a sentence like: *The second one occurs in sensitive patients: it is headache, it is serious, it occurs infrequently and in a strong form.*

case 2: let $SPAN_1$ be a (sub)set of leaves generated by $DescribeSingleSE1$ and $SPAN_2$ be a (sub)set of leaves generated by $DescribeSingleSE2$ (for instance the two $InformSign$ corresponding to the first and the second side effect). The minimal subtree's root is $ExplainSideEffect$, and the RR associated with it is multinuclear. Then, an inversion between $DescribeSingleSE1$ and $DescribeSingleSE2$ is executed, which produces an inversion in the order of presentation of the two side effects of this drug.

We tested the algorithm on the texts produced by our planner, and its outputs show the coherence required. Nevertheless, some considerations hold.

The only multi-nuclear relations we used assumed an uninfluential order among nuclei (eg. Joint, Sequence, etc): the last case of the algorithm would not work, for instance, in rhetorical relation such as a Temporal Sequence, when a precise order is considered to

be the correct one. However, we cannot think of any situation in which a pre-established order among the nuclei of such a relation needs to be changed, without changing the whole meaning of the text (and then the rhetorical relation itself).

As it was anticipated, the algorithm may move more information than it is necessary: in the example, swapping the side effect's severity and its risk category resulted in moving all the attributes of the side effect together with its severity.

Assumption 3 prevents a span to be chosen as an arbitrarily big set of adjacent tree leaves. A different algorithm would be needed if one or both spans extend across both nucleus and satellite of the root of the minimal subtree that generates them; however, in this case as well, we suspect that for such a situation to occur the whole meaning of the text would need to change, and another rhetorical structure tree would prove to be more appropriate.

Assumption 2 need to be studied more: we cannot be sure that the swapping works well only because of the particular genre of our texts: the performance of the algorithm should be tested on other kinds of text. Moreover, the fact that exchanging administration details and effects of the drug works particularly well is probably due to the couple of RRs involved (see the pattern for the schema Motivation + Enablement in (Mann and Thompson, 1988)).

6.4 Related research

The importance of persuasion in medical explanations has been proved by several authors, who declared that "enlistment of the patient by the physician is promoted by the use of inclusive language, that is, language that explicitly recognizes patients' competence and thus treats them with respect.... such language increases the physician's (and therefore also the prospective system's) chances of having the patient hear the diagnosis and treatment recommendations." (Forsythe, 1995). In her paper about the generation of 'empathetic responses' in a dialogue about medical treatment, Haimovitz (1991) argues that, to produce texts suited to the direct and the indirect users' needs, sentences with a given information content should be generated so as to 'stress favourable information while downplaying or offsetting unfavourable information'. This effect can be obtained by exploiting knowledge about the indirect user's 'concerns and worries', in addition to domain knowledge about the possible 'impact' of information items. A 'very sensitive test' or a 'non invasive treatment' are examples of favourable information, whereas 'a very long hospital stay' or 'a painful procedure' are examples of unfavourable ones. The 'cost' of an utterance is calculated as a function of the Addressee's characteristics (for example: anxiety about pain); the 'accent' of a sentence is then modified by joining sentences with appropriate conjunctions, or by introducing 'detensifier' or 'intensifier' adverbs. Our emphasizing methods are similar to those employed by these authors: we just apply them by introducing some conditions on ATN's arcs.

As far as the new intermediate phase between text planning and linguistic realization is concerned, our modules differ somehow from those suggested in the literature. The main purpose of what the majority of authors call 'sentence planning' is to prepare an input to linguistic realization, by transforming a semantic structure produced by the planner into an adequate syntactic structure (Reiter, 1994). A sentence planner takes the list of propositions produced by the text planner and determines how to express them in natural language (Rambow and Korelsky, 1992). This means determining the sentence scope (how to assemble clauses into larger sentences through embedding, paratactic coordination and so on) and make lexical, grammatical and reference choices by choosing appropriate RR markers. It means, as well, optimizing the shape of sentences by ordering locally related clauses, and avoiding repetitions through aggregation (the most thorough investigation to date on this topic is due to (Dalianis, 1996)).

The majority of sentence planning methods have been developed for texts which are aimed at illustrating a subject to the user. The expected effect of these texts is, once again, to increase the user's level of knowledge, and the role of sentence planning techniques is to increase the audience's ability to 'get the message', by improving its *accessibility* (Scott and Sieckenius de Souza, 1990) and *assimilation* (Fiedler and Huang, 1995). Methods employed aim at reducing the risk of ambiguities and at making the text more *concise* (Zukerman and McConachy, 1995; Carberry and Harvey, 1997) or more *streamlined* (Fiedler and Huang, 1995). Less frequent are systems aimed (like OPADE) at convincing the user to perform some action, by employing persuasion techniques. In this case, as Wilkinson (1995) suggests, reasons such as emphasis might be taken into account for the decision about whether a sentence planning task is appropriate.

As to global plan revision algorithms, to our knowledge there is no similar mechanism in other text generation systems. It can be argued that, especially for the ordering task, a more sophisticated planning technique would make this task unnecessary. Nevertheless, we maintain that, even in this case, the suitability of the plan produced could not be taken for granted. Moreover, it is important to consider the trade off between the cost of implementing a sophisticated planner and the one of "adjusting" the output of a simpler planner by using *a posteriori* techniques.

With these considerations in mind, we think that a multi-step text generation approach is, for the time being, a more practical solution to the kind of problems that we discovered in our evaluation studies. This is especially true in cases like the one we describe in this paper, when the text generator uses a pre-defined (and potentially extremely big) data base about the domain, rather than an *ad hoc* one: in this case, employing a simple top-down planner to collect and order the information content of the explanation, and subsequently refining the plan by applying some persuasion-increasing strategies, is still to be preferred. Our heuristics-based method of plan refinement contributed to making the output of our text generator more effective, though keeping it simple enough. The approach carries all the shortcomings of a meta-rule approach in contrast to a "purer" one, including the fact of being too domain-dependent. But we think that our heuristics can be generalized with little effort to consider argumentative texts in every domain.

7. Current State Of The System

We designed and tested separately the modules described in the previous Section, applying them in turn to the first version of OPADE's output, and therefore cannot include a revised version of Fig. 6, with all heuristics applied.

The main obstacle in coming to an integrated new system is that we have to define how to combine the described techniques, and to verify what happens exactly when the various modules are assembled: it may happen, for instance, that too much emphasizing–de-emphasizing or too much empathy produce a suspicion of deceit. Or, from a more practical point of view, order of application of the heuristic can be crucial: for instance, aggregating before ordering may cause more information to be moved, whereas the opposite may prevent the application of the aggregation heuristics.

Another problem that needs to be studied concerns the choice of the architecture which best integrates our modules: a pipeline (Reiter, 1994) a blackboard (Wanner and Hovy, 1996), a parallel processing (Rambow and Korelsky, 1992) or something else?

Finally, we should evaluate if the new texts are really preferred by Indirect Users: evaluation experiments often reserve surprises; an enlighting study about whether 'naturalness' increases effectiveness of automated telephone dialogues, for instance, proved that "lengthy stretches of interactional language are inappropriate for human-machine dialogues' and that 'a natural dialogue between a human being and a machine should

not (...) attempt to mimic human dialogues, but should take full account of the non-human nature of the machine" (Cheepen and Monaghan, 1997).

7.1 Final remarks

OPADE's generator relies on the corpus of transcripts described in Section 2.1: this corpus includes a limited number of texts, on a limited number of diseases, by a limited number of doctors, and that certainly caused some of the problems revealed by the evaluation studies. A first step towards a possible generalization would be to enrich the plan operator library, by increasing the corpus of transcripts with other scenarios or by analyzing them with other techniques (Walker and Moore, 1997). However, generalization of the considerations in this paper to a larger number of diseases and doctors and to other languages and cultures should be made with care.

The planning and the surface generation module of OPADE are domain-independent; the library of plan operators and of ATNs are not. Evaluation studies gave useful indications of the limitations of texts generated in the drug prescription domain, and defined Local and Global Plan Revision heuristics are a direct consequence of these studies. However, we maintain that similar limitations might be found in any instructional text generated by a two-step procedure, and that similar heuristics might be applied to increase the persuasion strength of these texts. For instance: "if the subtree that explains an action to be performed includes several leaves that are labelled as 'unpleasant', then find out items that are labelled as 'positive for compliance' in the same subtree, and add an introduction to emphasize them", and alike for treating 'relevant for compliance' items. We can say, in general, that the left side of our heuristics (expressing the conditions in which they should be applied) are strictly domain-dependent and should therefore be adapted to the specific application context; algorithms in their right sides are domain-independent or can be made so with minor changes.

To conclude: we share Wanner and Hovy's (1996) opinion that "the generation process is even more complex than originally viewed", and that "this complexity centrally affects the planning procedures". A totally domain-independent improvement in the discourse generation process could only be obtained by reconsidering text planning as a non linear and non monotonic process, in which strong interactions between subgoals exist. This would require representing 'communicative actions' in a more fine-grained way than at present, by specifying their negative as well as their positive main and side effects. In argumentative texts, in particular, graduality in the persuasion process could be obtained by introducing degrees of goal adoption and degrees of belief holding in the mental model of the addressee. With the text planning techniques that are applied at present, we claim that several plan repair tasks are needed to counteract the negative effects of not considering subgoal interactions: sentence planning is only one of them. We acknowledge the importance of style refinement, but we contend that other, more argumentative tasks are important as well, and that knowledge about the domain and about the addressee play a crucial role in all of them. In the medical field, this knowledge concerns not only what addressees presumably know, want or prefer, but also emotional factors such as, for instance, fear about disease prognosis or anxiety about pain.

Acknowledgments

We are indebted with the anonymous referees for the extremely useful comments to the first version of this paper.

References

André, E. and T. Rist. 1995. Generating Coherent Presentations Employing Textual and Visual Material. *Artificial Intelligence Review*, 9:147–165.

Berry, D., T. Gillie, and S. Banbury. 1995. What Do Patients Want to Know: an Empirical Approach to Explanation Generation and Validation. *Expert Systems With Applications*, 8(4):419–428.

Berry, D., T. Gillie, and F. de Rosis. in press. Evaluating Explanations about Drug Prescriptions. *Psychology and Health.*

Berry, D., I. Michas, M. Forster, and T. Gillie. 1997. What Do Patients Want To Know About Their Medicines and What Do Doctors Want To Tell Them?: A Comparative Study. *Psychology and Health*, 12:467–480.

Buchanan, B., J. Moore, D. Forsythe, G. Carenini, S. Ohlsson, and G. Banks. 1995. An Intelligent Interactive System for Delivering Individualized Information to Patients. *Artificial Intelligence in Medicine*, 7(2):117–154.

Carberry, S. and T. Harvey. 1997. Generating Coherent Messages in Real-time Decision Support: Exploiting Discourse Theory for Discourse Practice. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society.*

Cawsey, A. 1992. *Explanation and Interaction: the Computer Generation of Explanatory Dialogues.* MIT Press, Cambridge (Mass.).

Cawsey, A., K. Binsted, and R. Jones. 1995. Personalised Explanations for Patient Education. In *Proceedings of the 5th European Workshop on Natural Language Generation*, pages 59–74.

Cawsey, A., B. Webber, and R. Jones. 1997. Natural Language Generation in Health Care. *Journal of the American Medical Informatics Association*, 4(6).

Cheepen, C. and J. Monaghan. 1997. Naturalness in Automated Dialogues - less is more. In Levy and Wilks (Levy and Wilks, 1997), pages 83–89.

Dale, R., C. Mellish, and M. Zock, editors. 1990. *Current Research in Natural Language Generation.* Academic Press.

Dalianis, H. 1996. *Concise Natural Language Generation from Formal Specifications.* Ph.D. thesis, Royal Institute of Technology/Stockholm University, Department of Computer and System Science.

Dalianis, H. and E. Hovy. 1996. On Lexical Aggregation and Ordering. In *Proceedings of the 8th International Workshop on Natural Language Generation*, pages 29–32. Poster session.

De Carolis, B., F. de Rosis, F. Grasso, A. Rossiello, D. Berry, and T. Gillie. 1996. Generating Recipient-Eentered Explanations About Drug Prescription. *Artificial Intelligence in Medicine*, 8(2):123–145.

de Rosis, F. and F. Grasso. 1997. Simulating Plausible Conflict-Resolution Dialogues. In Levy and Wilks (Levy and Wilks, 1997), pages 55–61.

de Rosis, F., F. Grasso, F. Castelfranchi, and I. Poggi. 1996. Goals and Attitude Change in Generation: a Case Study in Health Education. In *Proceedings of the ECAI-96 Workshop on Modelling conflicts in AI*, pages 23–28.

de Zegher, I., A. Venot, C. Milstein, B. Séné, B. De Carolis, and S. Pizzutilo. 1994. OPADE - Optimization of Drug Prescription Using Advanced Informatics. *Computers Methods and Programs in Biomedicine*, 45(1-2):131–136.

DiMarco, C., G. Hirst, and E. Hovy. 1997. Generation by Selection and Repair as a Method for Adapting Text for the Individual Reader. In *Proceedings of the Workshop on Flexible Hypertext, Eighth ACM International Hypertext Conference.*

Fiedler, A. and X. Huang. 1995. Aggregation in the Generation of Argumentative Texts. In *Proceedings of the 5th European Workshop on Natural Language Generation*, pages 5–9.

Forsythe, D. 1995. Using Ethnography in the Design of an Explanation System. *Expert Systems With Applications*, 8(4):403–417.

Haimowitz, I. 1991. Modeling all Dialogue System Participants to generate Empathetic Responses. *Computer Methods and Programs in Biomedicine*, 35:321–330.

Hammond, K. 1990. Explaining and Repairing Plans that Fail. *Artificial Intelligence*, 45(1-2):173–228.

Hovy, E. 1990. Unresolved Issues in Paragraph Planning. In Dale et al. (Dale, Mellish, and Zock, 1990), chapter 2, pages 17–45.

Hovy, E. 1993. Automated Discourse Generation using Discourse Structure Relations. *Artificial Intelligence*, 63(1-2):341–385.

Hovy, E. and L. Wanner. 1996. Managing Sentence Planning Requirements. In Jokinen et al. (Jokinen et al., 1996), pages 53–58.

Jacobs, P. 1992. Why Text Planning isn't (still) Planning. In A. Ortony, J. Slack, and O. Stock, editors, *Communication from an Artificial Intelligence Perspective.* Springer-Verlag, pages 73–89.

Jokinen, K., M. Maybury, M. Zock, and I. Zukerman, editors. 1996. *Proceedings of the ECAI-96 Workshop on: Gaps and Bridges: New directions in Planning and NLG.*

Kolln, M. 1995. Employing User Attitudes in Text Planning. In *Proceedings of the 5th European Workshop on Natural Language Generation*, pages 163–179.

Korf, R. 1987. Planning as Search: a Quantitative Approach. *Artificial Intelligence*, 33(1):65–88.

Levy, D. and Y. Wilks, editors. 1997. *Proceedings of the First International Workshop on Human-Computer Conversation.*

Mann, W., C. Matthiessen, and S. Thompson. 1989. Rhetorical Structure Theory and Text Analysis. Technical report, USC/Information Sciences Institute.

Mann, W. and S. Thompson. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text*, 8(3):243–281.

Marcu, D. 1996. Building up Rhetorical Structure Trees. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI96)*, volume 2, pages 1069–1074.

Maybury, M. 1992. Communicative Acts for Explanation Generation. *International Journal of Man-Machine Studies*, 37:135–172.

Meteer, M. 1991. Bridging the Generation Gap Between Text Planning and Linguistic Realization. *Computational Intelligence*, 7(4):296–304.

Miller, P. 1984. *A critiquing approach to expert computer advice: ATTENDING.* London : Pitman Advanced.

Moore, J. and C. Paris. 1992. Exploiting User Feedback to Compensate for the Unreliability of User Models. *User Modeling and User-Adapted Interaction*, 2(4).

OPADE. 1993. Specification of the OPADE Kernel. AIM European Community Project A2027. Deliverable n. 5.

Penman. 1989. The Penman Documentation. Technical report, USC/Information Sciences Institute.

Rambow, O. and T. Korelsky. 1992. Applied Text Generation. In *Proceedings of the 3rd Conference on Applied Natural Language Processing*, pages 40–47.

Reed, C., D. Long, and M. Fox. 1996. Planning Persuasive Argument. In Jokinen et al. (Jokinen et al., 1996), pages 47–50.

Reiter, E. 1994. Has a Consensus NL Generation Architecture Appeared, and Is It Psycholinguistically Plausible? In *Proceedings of the 7th International Workshop on Natural Language Generation*, pages 163–170.

Reiter, E. 1995. NLG vs. Templates. In *Proceedings of the 5th European Workshop on Natural Language Generation*, pages 95–105.

Reiter, E. and L. Osman. 1997. Tailored Patient Information: Some Issues and Questions. In *Proceedings of the ACL-1997 Workshop 'From Research to Commercial Applications: Making NLP Technology Work in Practice'*, pages 29–34.

Scott, D. and C. Sieckenius de Souza. 1990. Getting the Message Across RST-based Text Generation. In Dale et al. (Dale, Mellish, and Zock, 1990), chapter 3, pages 47–73.

Séné, B., A. Venot, I. de Zegher, C. Milstein, S. Errore, F. de Rosis, and G. Strauch. 1995. A General Model of Drug Prescription. *Methods of Information in Medicine*, 34(4):310–317.

Vander Linden, K. and J. Martin. 1995. Expressing Rhetorical Relations in Instructional Text: a Case Study of the Purpose Relation. *Computational Linguistics*, 21(1):29–57.

Walker, M. and J. Moore. 1997. Empirical Studies in Discourse. *Computational Linguistics*, 23(1):1–12.

Walker, M. and O. Rambow. 1994. The Role of Cognitive Modeling in Communicative Intentions. In *Proceedings of the 7th International Workshop on Natural Language Generation.*

Wanner, L. and E. Hovy. 1996. The HealthDoc Sentence Planner. In *Proceedings of the 8th International Workshop on Natural Language Generation*, pages 1–10.

Wilkinson, J. 1995. Aggregation in Natural Language Generation: Another Look. Co-op work term report, Department of Computer Science, University of Waterloo.

Woods, W. 1970. Transition Network Grammars for Natural Language Analysis. *Communications of the ACM*, 13(10):591–606.

Zukerman, I. and R. McConachy. 1995. Generating Explanations across Several User Models: Maximizing Belief while Avoiding Boredom and Overload. In *Proceedings of the 5th European Workshop on Natural Language Generation*, pages 143–161.