From Rhetorical Structures to Document Structure: Shallow Pragmatic Analysis for Document Engineering

Gersende Georg[1], Hugo Hernault[2], Marc Cavazza[3], Helmut Prendinger[4], Mitsuru Ishizuka[2]
[1]Haute Autorité de Santé, 2 Avenue du Stade de France, 93218 Saint-Denis La Plaine Cedex, France;
[2] Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan;
[3]School of Computing, University of Teesside TS1 3BA Middlesbrough, United Kingdom
[4]National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
g.georg@has-sante.fr ; hugo@mi.ci.i.u-tokyo.ac.jp ; m.o.cavazza@tees.ac.uk ;
Helmut@nii.ac.jp ; ishizuka@i.u-tokyo.ac.jp

ABSTRACT
In this paper, we extend previous work on the automatic structuring of medical documents using content analysis. Our long-term objective is to take advantage of specific rhetoric markers encountered in specialized medical documents (clinical guidelines) to automatically structure free text according to its role in the document. This should enable to generate multiple views of the same document depending on the target audience, generate document summaries, as well as facilitating knowledge extraction from text. We have established in previous work that the structure of clinical guidelines could be refined through the identification of a limited set of deontic operators. We now propose to extend this approach by analyzing the text delimited by these operators using Rhetorical Structure Theory. The emphasis on causality and time in RST proves a powerful complement to the recognition of deontic structures while retaining the same philosophy of high-level recognition of sentence structure, which can be converted into application-specific mark-ups. Throughout the paper, we illustrate our findings through results produced by the automatic processing of English guidelines for the management of hypertension and Alzheimer disease.

Categories and Subject Descriptors
J.3 [**Life and Medical Sciences**]: Medical information systems; I.7.2 [**Document and Text Processing**]: Document Preparation - *Markup languages - Hypertext/hypermedia*

General Terms
Algorithms, Documentation, Languages.

Keywords
Medical document processing, natural language processing.

DocEng'09, September 16–18, 2009, Munich, Germany.

1. INTRODUCTION
Traditionally, document engineering has incorporated Natural Language Processing (NLP) techniques principally to recognise keywords or keyphrases to support indexing and retrieval [11]. With the development of marking up techniques as a means to categorise text segments [20], it became possible to use NLP as an aid to document structuring.

We have shown in previous work [8] how the automatic identification of specific linguistic markers in medical documents (more specifically clinical recommendations), known as deontic operators, could be used to provide a first level of structuring. The main outcome of this approach has actually been a software to analyse the structure of clinical guidelines during their development process, and it has been in use over the past years at one national guideline production agency. In this paper, we describe an extension of our previous approach in which the actual contents of recommendations can be further structured using a medically-relevant subset of rhetorical relations.

After summarising our previous approach based on the recognition of recommendations through the identification of deontic operators, we introduce the potential of Rhetorical Structure Theory (RST) [12] to analyse Clinical Guidelines. We then describe a processing pipeline through which to integrate deontic recognition and RST analysis.

This article is organized as follows: the next section introduces our approach, which is based on the structuring of clinical guidelines by rhetorical analysis. The 'Clinical guidelines and their analysis' section describes the G-DEE document engineering platform that incorporates shallow parsing techniques and their use in the French National Authority for Health (HAS). The Recommendations' processing pipeline is then described in three parts: the recognition of recommendations by G-DEE, RST analysis of recommendations and the structuring of recommendations by marking-up the RST functions. Finally we present results on two clinical guidelines.

2. APPROACH
Our objective is to develop methods for the content-based automatic structuring of clinical guidelines to be used to support the authoring process and the document life cycle.

Clinical guidelines are produced by committees through a complex process of consensus building, and often go through extensive rewriting and restructuring by multiple authors. This process is not without consequences on the actual style, clarity and readability of guidelines. Some studies have focused on the influence of data presentation [14] on cognitive decision mechanisms. This is why assisting the authoring of medical documents is essential [17] to guarantee a proper presentation of underlying data. We are convinced that intervention during guidelines authoring is one of the most useful and important steps for their improvement. One main challenge associated with their production is to be able to anticipate the impact on the target readership of the specific recommendations they contain, which depends on the style adopted. Style analysis can be based on Natural Language Processing (NLP) techniques analyzing the specific expression of recommendations as previously described [9].

One research direction consists in standardising guidelines' writing or even recurring to controlled languages [5]. Whilst their automatic processing is beyond the state-of-the-art of NLP techniques, we have shown recently that much benefit could be gained from the recognition of key expressions which would structure portions of text according to the document's logic as shown below. This is a case where local or shallow processing can be used to structure free text segments.

<FrontScope> Antihypertensive therapy </FrontScope>
<DeontOp> should be started </DeontOp>
<BackScope> with a low dose of a single drug, particularly after the age of 80 </BackScope>.

Figure 1. Recommendation analysis by G-DEE: structuring recommendation with <*FrontScope*>, <*DeontOp*> and <*BackScope*>

G-DEE (for *Guidelines Document Engineering Environment*), is a text analysis environment dedicated to the study of clinical guidelines [9]. It supports multiple document processing functions including the automatic recognition of recommendations using shallow NLP techniques (such as Finite-State Automata, FSA) to recognise deontic expressions involving verbs such as "authorise", "forbid", "ought to", which are most likely to correspond to recommendations.

Using specific grammars for deontic expressions, G-DEE parses the original document, identifying deontic operators and the text segments they apply to, which are called scopes (Figure 1). G-DEE integrates FSA that generate mark-ups corresponding to deontic operators and their scopes. Recommendations are thus structured by the marking-up of the deontic operator (<*DeontOp*>). A scope that precedes a deontic operator is called front-scope (<*FrontScope*>), whereas the back-scope (<*BackScope*>) corresponds to the scope which follows the deontic operator.

This environment supports various document analysis features, some dealing with specific text display. The analysed expressions are structured using specific mark-ups, which supports hypermedia presentation of the recommendations. G-DEE was originally conceived as a research environment but has since been adopted as an experimental tool for assisting the guideline authoring process. The same mechanisms that evaluate a guideline's structure can provide an on-line help at various stages of the authoring process (production of individual recommendation, consensus amongst the working party in charge of guideline authoring).

We evaluated the global performance of G-DEE in terms of automatic recognition of recommendations. To that effect, we asked four senior experts involved in the development of clinical guidelines to evaluate recommendations identification by G-DEE on two entire guidelines (103 to 142 sentences), using a scoring sheet similar to these used in the evaluation of Information Extraction systems. The work of each expert consists to check that each sentence is correctly analyzed for the occurrence of recommendations. For this evaluation, we measured consensus between experts using the kappa coefficient, which varied between 0.70 and 0.87. We observed that the precision of deontic marker identification varied between 88% and 98% and the recall between 81% and 99%.

2.1 Using G-DEE to Assist the Authoring of Clinical Guidelines

Since 2007, G-DEE has been integrated into the process of clinical guidelines authoring at the French National Authority for Health (Haute Autorité de Santé, HAS), which is in charge of the elaboration of all official guidelines in France. The elaboration of clinical guidelines is normally dependent on four working group meetings. During the second and the third meeting, recommendations are elaborated based on the rationale. This document constitutes the clinical guidelines themselves. The fourth meeting is dedicated to the analysis of the reading group's comments. Clinical guidelines are thus subject to multiple modifications during their writing phase. G-DEE has been used from the first draft of recommendations to the final version of the document, to provide an independent analysis of guidelines structure. The complete analysis of the different versions analyzed by G-DEE has been discussed with the respective HAS project leader. For the set of 18 guidelines processed during the past year, recommendations represent between 40 and 80.5% (mean 57.7%) of the whole text. This percentage is a preliminary quantitative indicator providing a heuristic judgment of their complexity (due to the fact that recommendations are the essence of guidelines). Whilst the absolute value may be difficult to interpret, significant variations during the authoring process of a single guideline may signal radical modifications which deserve specific attention.

We also use the marking-up of recommendations to study the distribution as well as the structure of clinical guidelines. Each sentence is analyzed and some recommendations which are often implicit or ambiguous can be re-formulated. In terms of structure, sections that do not contain any recommendation have often been re-assessed. The example (#1 - Figure 2) below illustrates another indicator for re-formulating recommendations, i.e. the respective proportion of front-scope (yellow color) and back-scope (blue color). An imbalance between scopes often highlights potential problems with the identification of actions and/or their justifications. Another example concerns the position of recommendations within the overall text. The "best" structure was often found achieved when placing recommendations either at the beginning or at the end of a section.

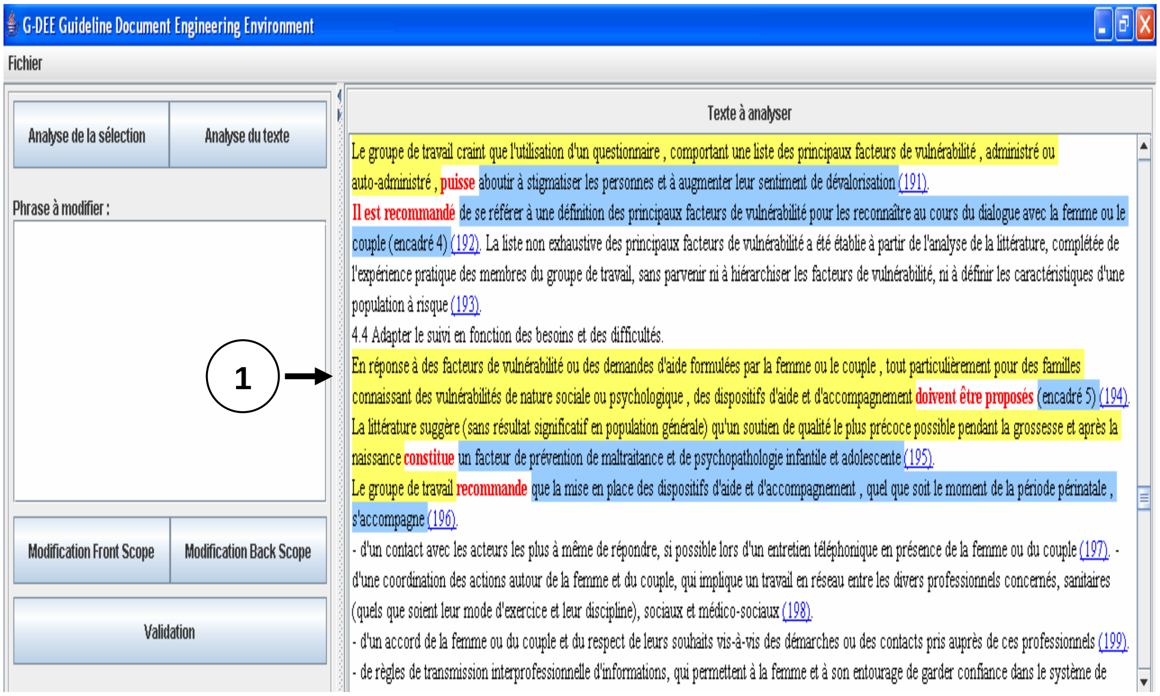

Figure 2. Example of marked-up recommendations after processing by the G-DEE analyzer.

Figure 3 shows the adoption of G-DEE by HAS project leaders. Since its introduction two years ago, the number of requests for authoring by HAS project leaders with G-DEE has steadily increased (44 cumulative number of requests considering that one guideline can be analyzed from 1 to 3 by G-DEE) indicating a growing interest amongst users (its use has not been made compulsory).

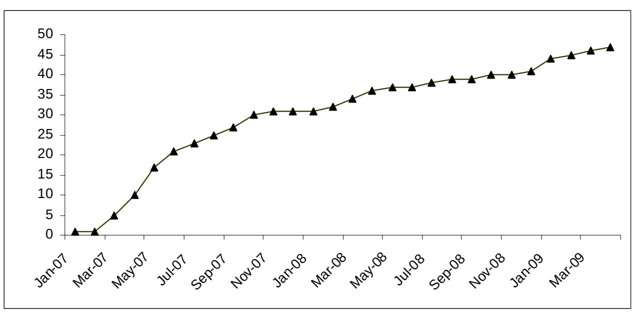

Figure 3. Cumulative number of requests for guideline analysis with G-DEE at HAS since its introduction.

3. COMPLEMENTARITY BETWEEN DEONTIC AND RHETORIC STRUCTURES

In addition to its use to support the guidelines' authoring process, we've shown in previous work [7] that structuring guidelines with deontic operators can help identifying important clinical actions that can be matched to underlying protocols. However, extending the automatic processing of guidelines to the actual contents of individual recommendations, i.e. processing the free text content of deontic operators' scopes, remains a challenging task from an NLP perspective. Our FSA-based recognition of deontic operators already required significant linguistic resources, despite being focused on specific linguistic descriptions. It thus seems difficult to adapt similar principles for the analysis of scopes which exhibit much greater syntactic variability and semantic coverage.

Ideally, we would seek a method which reconciles broad linguistic coverage, shallow NLP, and the ability to further structure the contents of recommendations' scopes. All this points towards discourse-processing methods, and led us to explore the potential use of Rhetorical Structure Theory (RST) [12]. Although some authors have suggested that legal texts were not amenable to RST formalization [19] and we have shown the proximity between legal texts and clinical guidelines in their use of deontic structures [9], we were comforted in our hypothesis by the many previous references applying RST to medical NLP and medical language generation [1 ; 2 ; 10 ; 15]. Fontan and Saint-Dizier [4] have also proposed to use RST for document structuring in the case of procedural texts.

4. RST PARSING OF RECOMMENDATIONS

In order to uncover the rhetorical structure of clinical guidelines, we used a RST discourse parser based on Support Vector Machines (SVM). The parser uses a rich set of shallow lexical, syntactic and structural features from the text, and processes its input in two steps.

Firstly, a discourse segmenter cuts the text into "elementary discourse units" (EDUs), the atomic units of discourse which are the terminal nodes of the rhetorical structure tree. This component is implemented as a binary SVM classifier trained on the RST Discourse Treebank (RST-DT) corpus [3]. Each word and its context are represented by a feature vector. Features used are the word itself (encoded in a dictionary), its part-of-speech (POS) tag, the previous/following words and their POS tags. Other features include the current word's highest ancestor possessing a lexical head equal to the word itself, as well as distances (absolute and relative position of the word in the current sentence, paragraph, and whole text). Finally, we also consider whether the next word is the start of a cue phrase. In our case we used a dictionary of 228 cue phrases in total, such as "despite that", "moreover", "nevertheless"... A feature vector is labeled as belonging to class 0 if there is no EDU boundary after its corresponding word; to class 1 otherwise.

In a second step, the calculated EDUs are passed to the tree-building component, which creates the full RST tree. This component is implemented using two SVM classifiers, also trained on the RST-DT: a binary classifier which, given two EDUs/spans (a sequence of EDUs connected by RST relations), indicates whether these elements are connected by a RST relation. A second, multiclass classifier indicates the label of the most likely RST relation linking two EDUs/spans given as input. Here a rich set of features is used, with among others:

- Textual organization cues, such as the presence of sentence/paragraph boundaries, measurements of the size and positioning of spans.
- Lexical clues, such as cue phrases, punctuation.
- Syntactic clues, such as the POS tags of a fixed-length prefix and suffix of each span.
- Dominance sets, a notion used to describe the logical nesting order of rhetorical relations between different spans, using associated syntax trees [18]. This is encoded using a set of syntactic, lexical and structural features (distance to the root of the syntax tree, dominating node's POS tag and lexical head, etc.).

By combining those two classifiers and using a simple bottom-up tree-construction algorithm, the final discourse tree can be built in linear time.

The RST Discourse Treebank was annotated with an extensive set of 78 rhetorical relations. However, in order to improve the computational properties of the classification problem and ensure a good separability between the label classes, we used the reduced set of 18 relations defined in [3] and used by [18]. In this set, the original relations are partitioned into 16 classes according to their rhetorical meaning similarity. For instance, *Problem-Solution*, *Question-Answer*, *Statement-Response*, *Topic-Comment* and *Comment-Topic* are all grouped under the *Topic-Comment* relation. Furthermore, two structuring relations are kept as originally described: *Textual-Organization* and *Same-Unit*. While we have observed a natural and quite efficient complementarity between deontic recognition and the RST analysis of recommendations' scopes, it would still be appropriate to investigate whether RST parsing could be used as a sole principle for guidelines' structuring and recommendations analysis. On the theoretical side, examples described by Gallardo [6] suggest that RST functions fail to capture key elements of recommendations. Direct RST analysis of recommendations mostly produces structures based on conditions and elaborations. When conditions are explicitly represented as part of the recommendation, RST analysis correctly identifies part of the recommendation, although it fails to provide a complete segmentation along the lines of those produced by G-DEE, with proper identification of scopes. On the other hand, when recommendations do not include conditionals, their deontic operators tend to be embedded in an underspecified relation (Figure 4).

(**Joint [N] [N]**
  (The treatment decision and strategy should be based on BP value)
  (and global CVR as assessed from the patient 's and their family 's history, clinical examination and further tests.)) [1]

Figure 4. Example of a RST analysis of coordination in the context of a recommendation's back-scope.

Although it some cases, RST parsing correctly isolates the deontic operator, this remains an exceptional situation. The extension of RST parsing to incorporate the recognition of deontic structures would require a corresponding extension to the parsing algorithm to ensure that all occurrences of deontic (irrespective of surface form, active/passive voice) are recognised, which would move away from the rather shallow parsing associated to RST analysis (see above) and would somehow *de facto* incorporate a 'deontic plug-in' parser not dissimilar to the one used in our first-step processing. All this eventually supports a pipeline model of processing, which will be described in the next section. Overall, the performance of RST analysis is lower than that of deontic operators recognition, which is consistent with the broader coverage imposed on RST parsing. Observed performance is in line with generic scores described for the RST parser when tested on non-medical texts (some loss of performance can be explained by difficulties with the processing of itemised lists, which occur more frequently in guidelines). A more detailed discussion of performance in the context of the integrated system is given in section 5 below.

5. THE RECOMMENDATIONS' PROCESSING PIPELINE

Since our deontic parser has been validated through user experiments and through real-world deployment within a guidelines production agency (HAS), we naturally thought of extending G-DEE by a further step of RST analysis, targeting the recommendations' scopes. However, a preliminary RST analysis limited to recommendations' scopes failed to produce usable results, as RST parsing requires well-formed sentences rather than isolated propositions. This is why we adopted a processing model based on the fusion of outputs from G-DEE and the RST parser, each presented as XML structures (Figure 5).

We took advantage of the aforementioned processing pipeline of the G-DEE system, which allows a multi-step analysis in which previously marked-up structures serve to identity segments for further analysis. For instance, recommendations' scopes which have been marked-up but are not analysed any further by G-DEE are the intended targets for substitution by corresponding RST structures.

[1] Throughout this paper, when illustrating results from the RST parser we shall display an indented formula, rather than the marked-up version, for the sake of readability. Only section 5 will use marked-up text since it deals with the unification of such structures.

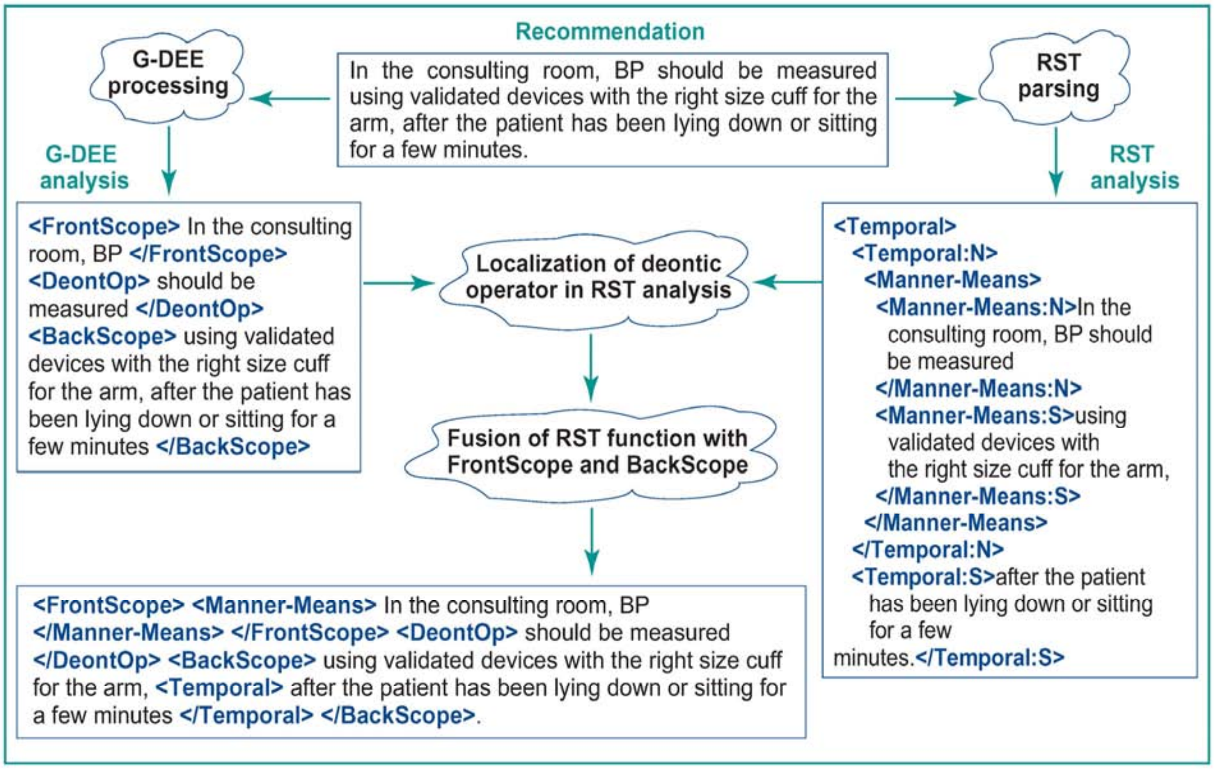

Figure 5. Refining Recommendations' Structure by Merging Deontic and Rhetorical Mark-ups.

From our perspective of content-based structuring, this leads to a further structuring of the front-scope, and the back-scope, based on the results of the RST analysis. We have developed a module operating in two steps: (i) localization of a deontic operator within the RST structure; (ii) fusion of the RST structure with the front-scope and the back-scope (resulting in these scopes being structured by RST functions).

This can be illustrated on the following recommendation:

In the consulting room, BP should be measured using validated devices with the right size cuff for the arm, after the patient has been lying down or sitting for a few minutes.

A pre-processing step consists of analysing the guideline using G-DEE to determine sentences that correspond to recommendations. An RST analysis is then performed on the file containing the recommendations identified by G-DEE. Both G-DEE and the RST parser generate XML files, which contain respectively mark-ups for RST functions (e.g. *Manner-Means*, *Temporal*, *Condition*)[2], and mark-ups for recommendations as described above (deontic operators and their scopes).

The G-DEE processor scans the sentence and extracts the deontic operator using the specific mark-up <*DeontOp*>. The next step consists of localizing the same deontic operator in the XML RST file, using the G-DEE processor that proceeds through a standard finite-state processing algorithm. The successful match leads us to determine whether the deontic operator is a part of the nucleus (N) or the satellite (S) by the recognition of the RST function (Figure 6).

Two configurations are tested in the algorithm for subsequent processing:

(i) the deontic operator is part of the nucleus: the RST function corresponds to the function of the front-scope, and the satellite corresponds to the back-scope.

(ii) the deontic operator is part of the satellite: the RST function corresponds to the function of the back-scope, and the nucleus corresponds to the front-scope.

The G-DEE processor then scans the sentence from the RST file, and extracts the function corresponding to the front-scope (either the nucleus or the satellite previously recorded information). It then uses a dedicated FSA to mark-up the corresponding front-scope with an appropriate tag (<Manner-Means> in the example presented in Figure 6).

In a similar way, the function corresponding to the back-scope is recorded and the G-DEE processor tags the back-scope accordingly (<Temporal> in the example presented in Figure 6).

[2] We have used our simplified XML notation for RST functions rather than previously described mark-up languages such as Reitteer and Stede [16].

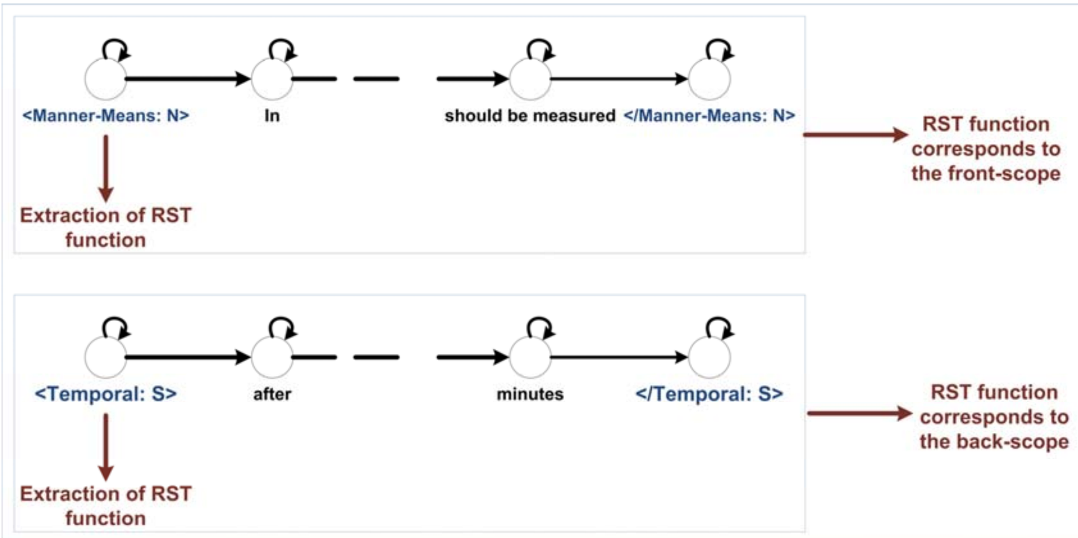

Figure 6. Parsing the RST structure to localise the deontic operator and determine the corresponding scopes.

The recommendation final marked-up structure becomes as shown by Figure 7.

<FrontScope>

<Manner-Means> In the consulting room, BP

</Manner-Means>

</FrontScope>

<DeontOp> should be measured **</DeontOp>**

<BackScope> using validated devices with the right size cuff for

the arm,

<Temporal> after the patient has been lying down

or sitting for a few minutes

</Temporal>

</BackScope>.

Figure 7. Final marking-up of a recommendation resulting from the fusion of G-DEE and RST marking-up.

6. RESULTS AND DISCUSSION

Hypertension Guidelines

We have extracted recommendations from the 2005 Hypertension Guidelines (in English, “Management of adults with essential hypertension”), obtaining a test set of 79 recommendations from 26 pages (approximately 920 lines) of guidelines’ text. As per the pipeline processing described above, all individual recommendations were processed by G-DEE and separately by the RST parser, both producing their own marking-up (the fact that RST parsing was applied to individual recommendations also kept RST structures manageable).

G-DEE processes documents offline and analyses an entire clinical guideline in an average of 300 seconds for a 26-pages text. The time required for RST parsing and fusion of representations adds a further 200 seconds to the processing pipeline.

We discuss the added value of RST parsing in terms of structuring recommendations. Overall, RST processing with basic functions had a very significant contribution for approximately 25% of recommendations. This means that not only it did refine the recommendations’ structure, but the new relations were directly meaningful. The most useful RST functions detected on these Guidelines are: *Condition* (10 occurrences), *Manner-Means* (5), *Temporal* (4), and *Enablement* (3). Despite their generic nature, these are the functions most closely related to clinical knowledge, since they describe processes or mechanisms. This would suggest that it is not necessary to create ‘specific’ RST functions adapted to the Medical domain. These functions are close to those that Gallardo [6] has identified as being used by experts (rather than journalists popularizing medical subjects). This can plausibly be explained by the fact that experts tend to resort to mechanistic explanations and underlying process descriptions (up to the mention of pathophysiological knowledge).

RST parsing also contributed to an improved structure with generic functions, through the *Elaboration* function, for 14% of recommendations: this includes isolating the grade of the recommendation or some specific target from within (generally back-) scopes.

Other phenomena were qualitatively relevant but contributed less quantitatively (5% overall). The *Background* function was able to identify very specific information and could have been considered as part of the main RST relations, were it not for its few occurrences (Figure 8).

(**Elaboration [N] [S]**

(**Elaboration [N] [S]**

(They may also be useful :)

(- in patients with refractory hypertension,))

(**Background [N] [S]** (-) (when assessing treatment efficacy.)))

Figure 8. RST analysis of a recommendation: the RST generic `Background` function can be used to identify context in recommendations.

Co-ordination can be detected by the *Joint* relation, often distinguishing between nouns and phrase co-ordination (as per the example presented in Figure 9). This can play a useful role in refining recommendations' structure and improving readability.

(**Condition [S] [N]**
(If hypertension is confirmed,)
(**Joint [N] [N]**
(lifestyle and dietary measures should be introduced)
(**Elaboration [N] [S]** (and the patient reassessed) ((Table 3).))))

Figure 9. RST analysis of a recommendation's condition: this type of structuring is common to the deontic step and the RST step and can be used for fusion representation.

In terms of detected processing errors, *Conditional* relations have been overridden by *Background* or by *Temporal* (sometimes even by *Manner-Means* or *Enablement*), or poorly detected within deeply nested rhetorical structures, leading to a lower detection score. A better joint recognition of functions could achieve substantial improvements of the *Conditional* relations.

Finally, RST analysis was unproductive for 20% of the recommendations, and failed to properly attribute functional relations, preventing the fusion between deontic and RST marked-up representations.

Alzheimer Guidelines

In a similar fashion, we have extracted recommendations from the 2008 Guidelines on Alzheimer's disease, obtaining a test set of 167 recommendations from 27 pages of text (Figures 10 – 11). The time performance for the analysis is similar to the one described above for hypertension guidelines.

RST processing with basic functions had a very significant contribution for approximately 29% of recommendations.

The most useful RST functions detected on these guidelines are: *Joint* (25 occurrences), *Condition* (16 occurrences), *Background* (15 occurrences), *Contrast* (7 occurrences), *Enablement* (3 occurrences), *Cause* (3 occurrences), *Explanation* (1 occurrence) and *Explanation* (1 occurrence). In a similar fashion to Hypertension guidelines, RST processing of scopes makes it possible to further structure recommendations using generic functions. These will play a role in the XSL-based transformations of guidelines to generate summaries or to customise guidelines to a specific audience. For instance, *background* may be removed from summaries, whilst *conditions*, through the additional level of structuring they offer can be used for explanatory purposes.

Conversely, RST analysis was unproductive for 25% of the recommendations. For example:

A risk/benefit analysis must always be conducted before introducing treatment.

Drugs must be prescribed for a short period and at the lowest effective dose.

This would be consistent with the use of RST advocated in this paper, which consists mainly in refining the structuring of recommendations via the rhetorical analysis of scopes, as the above recommendations do not allow for extensive RST analysis.

<FrontScope>
<Condition> If the initial assessment confirms a cognitive decline,
</Condition>
</FrontScope>
<DeontOp> it is recommended **</DeontOp>**
<BackScope> that diagnosis and care be undertaken jointly by the patient's regular doctor and a consultant. **</BackScope>**

Figure 10. Marking-up of a recommendation evidencing its condition part.

<FrontScope>
<Background> Genetic consultation **</Background>**
</FrontScope>
<DeontOp> may be necessary **</DeontOp>**
<BackScope>
<Background> as it often runs in families. **</Background>**
</Backscope>

Figure 11. Marking-up of a recommendation evidencing its background.

7. CONCLUSION

The analysis of recommendations' scopes using RST can successfully extend our previous approach, improving automatic structuring for 44% of recommendations, which increases significantly the quality of the automatic processing, even more so considering that documents tend to be analysed several times during their authoring cycle. Further, it remains compatible with our philosophy of document processing, which is to structure text segments using discourse markers, specific (e.g. deontic), or not. This type of automatic analysis tends to be well-accepted by guidelines' developers as it is designed as a human-in-the-loop approach.

This is also an interesting test case for medical NLP, where the recognition of discourse structures, rather than of named entities or actions (for instance through Information Extraction or terminological processing) can support the identification of clinically relevant information over an entire text.

This approach should also be portable to other application areas than medical texts: a condition for portability is not so much the existence of sublanguages as the document genre. For instance, our deontic approach was originally inspired by work from Moulin [13] on legal texts, which share many similarities with clinical guidelines in their use of deontic structures, such similarity deriving mainly from the document's genre and purpose. Another condition would be to establish that the relationships between RST functions and deontic structures remain the same across domains.

We have seen that deontic aspects were poorly covered by generic RST functions alone: however, in other technical domains other argumentative structures than deontics may play a major role; in the absence of background work on these, a potential research direction would be to consider specific extensions to the RST formalism itself, intending to capture genre-specific rhetorical structures.

ACKNOWLEDGMENTS
Dave du Verle developed the first version of the RST parser used in these experiments. Gersende Georg and Marc Cavazza have been funded by NII for a summer visit in 2008 during which this collaboration was established.

8. REFERENCES
[1] Afantenos S, Karkaletsis V, Stamatopoulos P. Summarization from medical documents: a survey. *Artificial Intelligence in Medicine* 33 (2), pp. 157-177 (2005).

[2] Androutsopoulos I, Spiliotopoulos D, Stamatakis K, Dimitromanolaki A, Karkaletsis V, Spyropoulos CD. Symbolic authoring for multilingual natural language generation. In: *Methods and Applications of Artificial Intelligence 2308*, pp. 131-142 (2002).

[3] Carlson L, Marcu D, Okurowski ME. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In: *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue. Annual Meeting of the ACL*. Association for Computational Linguistics 16 pp. 1-10 (2001).

[4] Fontan L, Saint-Dizier P. Constructing a know-how repository of advices and warnings from procedural texts. *Proceedings of the 2008 ACM Symposium on Document Engineering.* ACM Press, pp. 249-252 (2008).

[5] Fuchs NE, Kaljurand K, Schneider G. Attempto Controlled English Meets the Challenges of Knowledge Representation, Reasoning, *Interoperability and User Interfaces. FLAIRS Conference*, pp. 664-669 (2006).

[6] Gallardo S. Pragmatic support of medical recommendations in popularized texts. *Journal of Pragmatics 37* (6), pp.813-835 (2005).

[7] Georg G, Cavazza M. Integrating Document-based and Knowledge-based Models for Clinical Guideline Analysis. In: R. Bellazzi, A. Abu-Hanna, J. Hunter (eds). *Proceedings of 11th Conference on Artificial Intelligence in Medicine in Europe*. LNAI 4594, Springer-Verlag Berlin Heideberg, pp.421-430 (2007).

[8] Georg G, Jaulent M-C. A Document Engineering Environment for Clinical Guidelines. In: Peter R. King & Steven J. Simske (eds). *Proceedings of the 2007 ACM Symposium on Document Engineering*. ACM Press, New York NY, USA, pp. 69-78 (2007).

[9] Georg G, Jaulent M-C. An Environment for Document Engineering of Clinical Guidelines. In: *Proceedings AMIA Symposium*, pp.276-280 (2005).

[10] Grasso F. Rhetorical coding of health promotion dialogues. In M. Dojat, E. Keravnou and P. Barahona (eds). *Proceedings of the 9th Conference on Artificial Intelligence in Medecine in Europe*. Berlin: Springer, pp. 179-188 (2003).

[11] Kumar N, and Srinathan K, 2008. Automatic keyphrase extraction from scientific documents using N-gram filtration technique. *Proceedings of the 2008 ACM Symposium on Document Engineering*. ACM Press, pp. 199-208 (2008).

[12] Mann WC, Thompson S. A. Rhetorical Structure Theory: Toward a functional theory of text organisation. *Text 8*, no. 3, pp. 243-281 (1988).

[13] Moulin, B. and Rousseau, D., Knowledge acquisition from prescriptive texts. In *Proceedings of the 3rd international conference on Industrial and engineering applications of artificial intelligence and expert systems*, (Charleston, South Carolina, United States, 1990), pp. 1112-1121 (1990).

[14] Patel V, Arocha J, Diermeier M, How J, Mottur-Pilson C. Cognitive psychological studies of representation and use of clinical practice guidelines. *International Journal of Medical Informatics* 63(3):147-167 (2001).

[15] Piwek P, Hernault H, Prendinger H, Ishizuka M. T2D: Generating Dialogues Between Virtual Agents Automatically from Text. *Intelligent Virtual Agents*, pp. 161-174 (2007).

[16] Reitteer D, Stede M. Step by Step: Underspecified markup in incremental rhetorical analysis. In: *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora* (LINC-03) at the EACL, Budapest (2003).

[17] Shiffman R, Karras B, Agrawal A, Chen R, Marenco L, Nath S. GEM: A proposal for a more comprehensive guideline document model using XML. *J Am Med Informatics Assoc* 7, pp.488-498 (2000).

[18] Soricut R, Marcu D. Sentence level discourse parsing using syntactic and lexical information. In: *Proceedings of the 2003 Conference of the North American Chapter of the Association For Computational Linguistics on Human Language Technology* 1, pp.149-156 (2003).

[19] Taboada M, Mann WC. Applications of Rhetorical Structure Theory. *Discourse Studies* 8 (4), pp. 567-588 (2006).

[20] Zerida N, Lucas N, and Crémilleux B. Combining linguistic and structural descriptors for mining biomedical literature. *Proceedings of the 2006 ACM Symposium on Document Engineering*. ACM Press, pp. 62-64 (2006).