Structuring Clinical Guidelines through the Recognition of Deontic Operators

Gersende Georg, Isabelle Colombet and Marie-Christine Jaulent

Université Paris Descartes, Faculté de Médecine; INSERM, U729; SPIM, F-75006 Paris, France

Abstract

In this paper, we present a novel approach to structure Clinical Guidelines through the automatic recognition of syntactic expressions called deontic operators. We defined a grammar and a set of Finite-State Transition Networks (FSTN) to automatically recognize deontic operators in Clinical Guidelines. We then implemented a dedicated FSTN parser that identifies deontic operators and marks up their occurrences in the document, thus producing a structured version of the Guideline. We evaluated our approach on a corpus (not used to define the grammar) of 5 Clinical Guidelines. As a result, 95.5% of the occurrences of deontic expressions are correctly marked up. The automatic detection of deontic operators can be a useful step to support Clinical Guidelines encoding.

1. Introduction

Clinical Guidelines are normalized documents which play an important role in the dissemination of standardized medical knowledge and Best Practice. In recent years, significant research has been dedicated to the computerization of Clinical Guidelines in order to facilitate their authoring or to promote their inclusion in Decision Support Systems [1]. An important aspect of that research has been the development of document processing tools that could support the encoding of textual guidelines. The most comprehensive approach is the Guideline Elements Model (GEM) [2], which is a model facilitating the translation of natural language guidelines into a standard, computer interpretable format based on XML markups. In this paper, we introduce a novel approach to the process of structuring textual guidelines, based on their actual linguistic content. Because guidelines are organized around recommendations, they tend to contain specific linguistic expressions which organize the medical knowledge they convey. Our initial hypothesis was that the automatic processing of such linguistic expressions could be used to automatically derive a document structure, while at the same time being tractable due to the limited number of these linguistic forms. This prototype is integrated into our G-DEE (Guideline Document Engineering Environment) environment, a document processing tool dedicated to the study of Clinical Guidelines. We conclude by presenting example results and analyzing system performance.

2. Material and Methods

Our source of inspiration was previous research in document processing, essentially for knowledge extraction from prescriptive texts, such as legal documents. These share many similarities with Clinical Guidelines at the linguistic level, both in terms of style and textual genre. Moulin and Rousseau [3] have described a method to automatically extract knowledge from legal texts based on the concept of "*deontic operators*". They found that the contents of prescriptive statements are specified by "normative propositions", which in the French texts studied by Moulin and Rousseau manifest themselves through such verbs as "*pouvoir*" (to be allowed to or may), "*devoir*" (should or ought to), "*interdire*" (to forbid). Von Wright [4], who developed deontic logic, considered normative statements as "a statement to the effect that something ought to or may or must not be done. Deontic propositions have been found by Kalinowski [5] to be the most characteristic linguistic structures of normative texts. Moulin and Rousseau showed that knowledge extraction from legal texts based on deontic operators is a convenient way to resolve problems of knowledge acquisition from texts, without having to take into account the detailed meaning of recommendations, using them instead to structure the document.

The purpose of this research is to investigate the potential of automatic recognition of deontic operators to structure Clinical Guidelines. The automatic structuring of Clinical Guidelines can support their encoding through markup languages such as GEM, or the generation of IF-THEN rules from their contents [6].

Linguistic Analysis: Lexicometric Studies

To validate the above hypothesis on the role of deontic verbs in Clinical Guidelines we first carried a lexicometric analysis on a corpus of 20 Clinical Guidelines (in French) published by the ANAES[1], the French National Agency for Accreditation and Health Evaluation.

We first studied the frequency of deontic verbs for the set of 20 Clinical Guidelines collected. We used the statistical text analysis software *Tropes*[2] to analyze these documents, particularly words occurrences and lemmatized verbs. The corpus is composed of 280 pages containing 83 997 word occurrences. We counted 1137 occurrences of verbs indicating deontic modalities (for the verbs "*devoir*" (should or ought to), "*pouvoir*" (to be allowed to), "*recommander*" (to recommend) and "*convenir*" (to be appropriate)). We considered "to recommend" ("*recommander"*) as a deontic modality due to the fact that in medical texts it always expresses recommendations. These verbs account for 18 % of all occurrences of verbs in the corpus (7686 occurrences for the set of verbs), while deontic verbs only represent 3% of the verbs in the corpus vocabulary. We also noticed an important number of occurrences of verbs indicating deontic modalities, in individual texts (from 18 to 111 occurrences). Another property that needs to be investigated is the distribution of deontic operators throughout the text. Clinical Guidelines being a set of structured recommendations, one would expect deontic operators to be distributed in a way which is consistent with these documents' style. By analyzing the distribution of the principal verbs constitutive of deontic operators (i.e. in French "*recommander*" (to recommend)) in each guideline, we obtained several distribution patterns. All these patterns share two common features. The first one is the scope of distribution, which spans across the entire text. The second one is the recurrence of groupings of deontic verbs. The latter finding is an indicator of textual structure, namely the repetition of deontic operators within specific sections.

---

[1] http://www.anaes.fr
[2] http://www.acetic.fr/

Defining the Grammar of Deontic Operators

Our next step is to design a tool for the automatic recognition of deontic operators, which feature deontic verbs within specific syntactic structures. Those syntactic structures also relate the deontic operator to the remainder of the text, a property which will be used for automatic text structuring. To identify specific syntactic structures we have studied the occurrences of deontic verbs in context. In order to define a grammar for the recognition of deontic operators, we used a variant of our corpus, which comprises 17 documents, and also includes consensus conferences and medical teaching material (in the field of diabetes, hypertension, asthma, dyslipidemias, epilepsy, renal disease). This enables to collect examples of variability in the authoring of medical documents so as to improve coverage. We used the "*Simple Concordance Program (release 4.07)*[3]" to analyze these documents. This program provides scopes for each word in the corpus, as shown below deontic operators are: "is then recommended" (in French, "*est donc recommandée*"), "are not recommended" ("*ne sont pas recommandées*"), "should be proposed" ("*doit être proposé*"), "may then be advised" ("*peut être alors conseillé*").

*une insulinothérapie **est donc recommandée** lorsque l'HbA1C*
*traitement oral **ne sont pas recommandées**, sauf en cas de*
*traitant les addictions **doit être proposé** aux patients*
*thérapeutique **peut être alors conseillée**. L'interrogatoire*

As previously described by Moulin and Rousseau, the first set of deontic operators is composed of "*pouvoir*" (to be allowed to or may) and "*devoir*" (should or ought to). By analyzing scopes of these deontic operators, we observed that these verbs are mainly followed by a verb in an infinitive form, which generally corresponds to a specification of the deontic operator. We also observed that deontic operators most often occur at the passive voice. In addition, we identified a set of deontic operators that appear to be specific to Clinical Guidelines. For example, in French "*recommander*" (to recommend), "*conseiller*" (to advise), or "*préférer*" (to prefer). By analyzing the scopes of these deontic operators, we observed that these verbs as well occur mostly in the passive voice (i.e. auxiliary followed by the past participle form of the verb). In order to define the coverage of a grammar recognizing these deontic operators from their surface form, we studied their expressions in our corpus, which led to the identification of 56 syntactic patterns for deontic operators corresponding to their various surface forms.

Automatic Recognition of Deontic Operators

The identification of deontic operators is based on a dedicated parser, which analyzes the whole document using our specialized grammar. We chose to use Finite-State Transition Networks (FSTN) [7] as a syntactic formalism. One reason was that they provide a convenient way to define specialized structures, including their morphological variants. Another rationale was their performance on certain Information Extraction (IE) tasks [8] (such as named entity recognition), which are similar to the recognition of deontic structures. Using results from our corpus analysis, we defined a complete set of FSTN to represent the surface forms of deontic operators. From the 56 syntactic patterns identified, we defined 546 FSTN to take into account morphological variants. An example of a FSTN is given in Figure 2 (deontic operators can contain negative statements as well, which means that the negation of a deontic operator does not have to be processed independently). Another aspect that needs to be represented is the relation between the operator and the surrounding text, in other words how the deontic operator actually structures the document. We defined, following Moulin and Rousseau [3], an operator's scope as that part of the

---

[3] Downloadable from: http://www.download.com

sentence to which the modal operator applies. A scope that precedes a modal operator is called *front-scope*, whereas the *back-scope* corresponds to a scope which follows the operator (see Figure 3). The actual parsing process identifies deontic operators in the guideline document and marks them up, together with the relevant text sections which constitute their *back-scope* and *front-scope* (Figure 1). It has been implemented as a two-stage process: the first stage is dedicated to the recognition and marking up of deontic operators in the textual document. The second stage uses the document marked up for deontic operators to specify their associated scopes.

Parsing Strategy

The FSTN grammar for deontic operators is contained in an external file which is loaded at the beginning of the automatic processing step (see Figure 1). Each sentence of the textual guideline is parsed in sequence by determining the relevant FSTN.

During parsing, FSTN are activated from the recognition of key vocabulary: once activated the corresponding text portion is parsed using generic parsing functions applied to the FSTN contents. Upon successful parsing the deontic operator is identified and marked up as such in the document (see Figure 4).
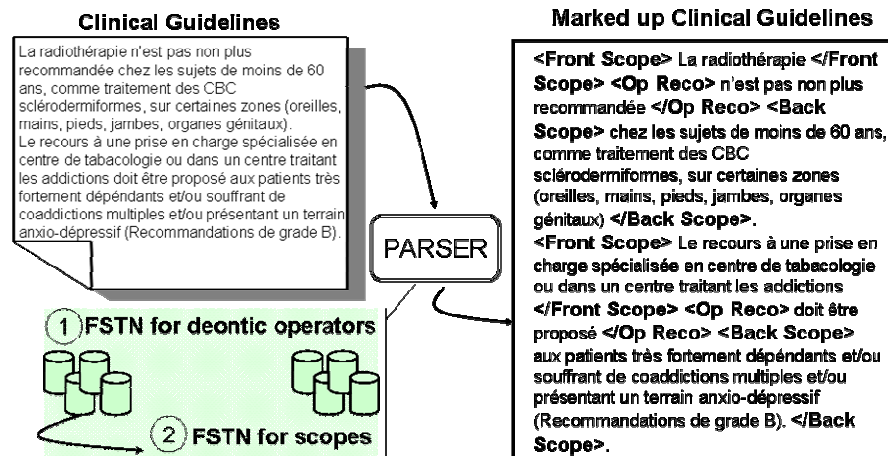

Figure 1 – Structuring of Clinical Guidelines through the recognition of deontic operators and their front- and back-scope.

A conflict resolution step determines the most relevant FSTN when several of them can be activated (for instance, due to shared patterns). By default, the most appropriate pattern is the pattern containing the largest number of concordances hence the more specific one. The application of a given FSTN recognizes a specific occurrence of a deontic operator in the text: it does not give raise to an instantiated representation but to the marking up of the textual occurrence of the specific operator recognized in the text.

Validation

We tested our FSTN parser on 5 randomly selected Clinical Guidelines among the 20 considered in our statistical analysis. None of these 5 texts has been used for the definition of our deontic operators' grammar. For this evaluation, we mainly focused on the correct identification of the following deontic verbs: "*recommander*" (to recommend), "*devoir*" *(should or ought to)*, "*pouvoir*" (to be allowed to or may) and "*convenir*" *(be appropriate)*. For each Clinical Guideline, we counted the number of correctly marked up occurrences of

deontic verbs. As a preliminary result, 95.5% of documents in our test set (which contains 311 deontic operators) are correctly marked up (in terms of operator identification as well as *front-scope* and *back-scope* recognition). The 4.5% of errors identified arise from a few specific syntactic phenomena which had not previously been identified.

3. Example Results

We present two examples of marking up for the operator "*recommander*" (to recommend) and the operator "*devoir*" (should or ought to) (cf. Figure 4). The first occurrence of "*recommander*" (to recommend) illustrates the recognition of that operator expressed in the passive voice and in a negative form, using the FSTN of Figure 2. The second shows the direct recognition of the passive voice for "*devoir*" (should or ought to).
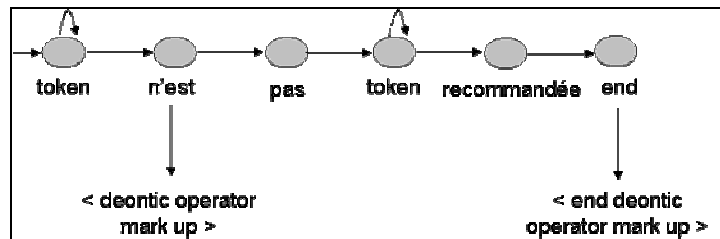

Figure 2 – FSTN for the recognition "is not recommended" deontic operator.

While the first step of document structuring marks up all occurrences of deontic operators, the second step identifies the operator's scopes. It does so by parsing the guideline previously marked up for deontic operators using another dedicated FSTN (Figure 3), which recognizes punctuation signs as well as markups inserted in the previous step.
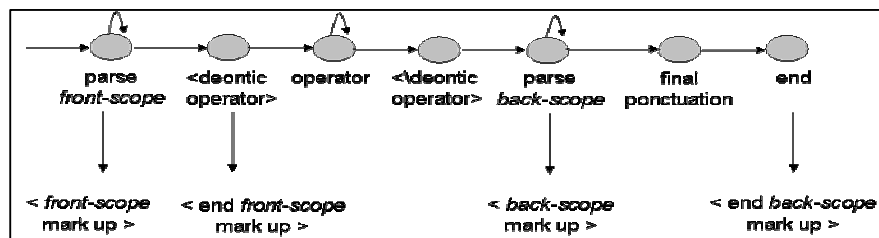

Figure 3 – FSTN for recognizing front- and back-scope for deontic operator previously recognized in Figure 2.

The final result is presented on Figure 4, which shows the kind of structuring produced on a portion of textual guidelines.

\<Front Scope\> La radiothérapie **\</Front Scope\> \<Op Deont\>** n'est pas non plus recommandée **\</Op Deont\> \<Back Scope\>** chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiformes, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux) **\</Back Scope\>**.[4]

\<Front Scope\> Le recours à une prise en charge spécialisée en centre de tabacologie ou dans un centre traitant les addictions **\</Front Scope\> \<Op Deont\>** doit être proposé **\</Op Deont\> \<Back Scope\>** aux patients très fortement dépéndants et/ou

---

[4] **\<Front Scope\>** Radiotherapy **\</Front Scope\> \<Op Deont\>** is not recommended either **\</Op Deont\> \<Back Scope\>** for persons under 60 years of age, as treatment of sclerodermiforms CBC, or on certains body parts (ears, hands, feet, legs, genitals) **\</Back Scope\>**.

souffrant de coaddictions multiples et/ou présentant un terrain anxio-dépressif (Recommandations de grade B) **</Back Scope>**.[5]

Figure 4 – Results of marking up deontic operators.

4. Conclusion

In this study, we discussed the role of deontic expressions in Clinical Guidelines and how these contribute to the textual structure of Clinical Guidelines. We showed that their automatic recognition using a dedicated FSTN parser supports the automatic structuring of the document. Our first prototype integrated into the G-DEE environment correctly marked up 95.5% of the occurrences of deontic expressions on a test set composed of Clinical Guidelines not used in the initial grammar definition. An analysis of recognition errors suggested that simple extensions of the recognition grammar could further improve the recognition score. The automatic detection of deontic operators can be a useful step to support Clinical Guidelines encoding in document-based approaches such as GEM. They can also serve as a pre-processing tool for knowledge acquisition from textual guidelines, where they would assist in the identification of expressions that could be converted into IF-THEN rules. A valuable extension of this approach would consist in the further processing of the textual contents of a deontic operator's *back-scope* and *front-scope*, which would identify relevant expressions, such as those denoting medical treatments. Such processing can be based on terminological recognition or information extraction methods such as named entity recognition.

References

[1] Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes R, et al. Comparing Computer-Interpretable Guideline Models: A Case-Study Approach. *J Am Med Inf Assoc* 2003, 10(1):52-68.

[2] 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* 2000, 7:488-498.

[3] Moulin B and Rousseau D. Knowledge Acquisition from Prescriptive Texts. In: *ACM: 1990*; 1990: 1112-1121.

[4] Von Wright G-H. "The Foundation of Norms and Normative Statements". *Philosophical Papers, Oxford, Blackwell* 1983:67-82.

[5] Kalinowski G. La Logique Déductive. Presses Universitaires de France (in French); 1996.

[6] Georg G, Séroussi B, Bouaud J. Does GEM-Encoding Clinical Practice Guidelines improve the Quality of Knowledge Bases? A Study with the Rule-Based Formalism. In: *Proceedings of the Amer Med Informatics Assoc: 2003; Washington, DC*; 2003: 254-258.

[7] Roche E and Schabes Y. Finite-State Language Processing. MIT Press; 1997.

[8] Pazienza M T. Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, vol. 1299: Springer-Verlag; 1997.

Address for correspondence

Gersende Georg
INSERM, U729; SPIM, 15 rue de l'Ecole de Médecine, F-75006 Paris, France.
Gersende.Georg@spim.jussieu.fr

---

[5] **<Front Scope>** Referral to a specialized center treating addiction to tobacco **</Front Scope>** **<Op Deont>** should be proposed **</Op Deont>** **<Back Scope>** to the patients strongly dependants and/or suffering of multiple coaddictions and/or presenting an anxio-depressive context (Recommendations of grade B) **</Back Scope>**.