Connecting Medical Informatics and Bio-Informatics
R. Engelbrecht et al. (Eds.)
ENMI, 2005

Indexing UMLS Semantic Types for Medical Question-Answering

Thierry Delbecque[a], Pierre Jacquemart[a], Pierre Zweigenbaum[a,b,c]

[a] *INSERM, U729, Paris, France*
[b] *Assistance Publique – Hôpitaux de Paris, STIM/DSI, Paris, France*
[c] *INALCO, CRIM, Paris, France*

Abstract

Open-domain Question-Answering (QA) systems heavily rely on named entities, a set of general-purpose semantic types which generally cover names of persons, organizations and locations, dates and amounts, etc. If we are to build medical QA systems, a set of medically relevant named entities must be used. In this paper, we explore the use of the UMLS (Unified Medical Language System) Semantic Network semantic types for this purpose. We present an experiment where the French part of the UMLS Metathesaurus, together with the associated semantic types, is used as a resource for a medically-specific named entity tagger. We also explore the detection of Semantic Network relations for answering specific types of medical questions. We present results and evaluations on a corpus of French-language medical documents that was used in the EQueR Question-Answering evaluation forum. We show, using statistical studies, that strategies for using these new tags in a QA context are to take in account the individual origin of documents.

Keywords:
Natural Language Processing; Information Retrieval; Language; France; UMLS; Funding, Non-US Government

1. Introduction

An increasing volume of good-quality textual knowledge bases is available in the medical domain, mostly through the development of the Web. Whereas the largest part of these documents is in English, an increasing proportion uses other languages. For instance, high-quality medical documents in French are available through the CISMeF catalog of French medical web sites [1]. The interest of Medical Question-Answering (henceforth QA) systems, which aim to bring precise, short answers to medical questions [2], has been highlighted recently [3]. These systems should benefit from these large document collections.

QA systems generally rely on an indexing of *named entities* [4] (names of persons and locations, dates and amounts, etc.), which are the focus of most questions in open-domain QA. In medical QA, we consider that named entities (henceforth NE) should be adapted to the domain [3]. An appropriate set of classes of entities is therefore necessary: we propose to use the UMLS semantic types for this purpose.

In this paper, we describe an experiment in indexing a French-language medical text collection, obtained from health-oriented Web sites, with UMLS (Unified Medical

Language System) concepts and semantic types. We evaluate the quality of these indexes both intrinsically (appropriateness of UMLS tags *wrt* the corresponding text) and in a goal-oriented task (relevance to help answer a specific question). We show that strategies for using these new tags in a QA context are to take in account the individual origin of documents.

2. Material and Methods

2.1 Material

Our corpus is the text collection prepared for the EQueR 2004 evaluation of French QA systems [5] (http://www.technolangue.net/article61.html). It consists of the documents indexed by the CISMeF catalog for nine sites (see table 1), extended with the documents to which they point (through HTML links) on the same sites. In the UMLS, we used the French terms and their associated concepts from the Metathesaurus 2002AA, together with their semantic types in the Semantic Network.

Table 1: Web sites from which the corpus was collected

[CANCER]	www.fnclcc.fr & www.fnclcc.com : French national federation of anti-cancer centers
[DOCFRA]	www.ladocfrancaise.gouv.fr & www.ladocumentationfrancaise.fr : French official documentation institute
[AFSSAPS]	afssaps.sante.fr : French agency for health safety of health products (former Drug Agency)
[ANAES]	www.anaes.fr : National agency for accreditation and evaluation in health
[ORPHA]	www.orpha.net : Information on rare diseases
[SENAT]	www.senat.fr : official site of the French Senate
[CHUROUEN]	www.chu-rouen.fr : Rouen School of Medicine
[UROUEN]	www.univ-rouen.fr : Rouen University, medical branch
[CANADA]	www.hc-sc.gc.ca : bilingual site of Health Canada, federal department of information on health

2.2 Tagging UMLS Semantic Types in French Texts

The first step of our method consists in locating occurrences of UMLS semantic types inside the corpus. For the English language, there exist UMLS linguistic companion tools such as the Specialist Lexicon [6], which can be used for instance in UMLS-oriented indexing tasks [7]. In contrast, the French language cannot benefit today from such tools (this situation should improve, nevertheless, thanks to the recent UMLF initiative [8]). Therefore, we resort to a mixture of general-purpose tools and specifically-developed methods.

We first tag the corpus with part-of-speech (POS) information (noun, verb, etc.) with Tree-Tagger (http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html). We then use POS patterns to locate noun phrases [9]. Then, for each located noun phrase, we look for all the French strings in the UMLS Metathesaurus that contain at least one common word with the phrase. This builds a lattice of terms, from which only the maximal elements are retained: those that share the maximum number of words with a Metathesaurus term. The phrase is then tagged with the concepts and the semantic types associated with these maximal elements. This method allows for example to locate the UMLS concept for *oedème papillaire* (CUI = C0030353, *Papillary edema*) inside the phrase “oedème bilatéral papillaire” (bilateral papillary edema), though the latter is not an UMLS string.

After noun-phrase processing, our system tries to extract *clauses* from the corpus, a *clause* being a segment of text that is roughly structured as **[Subject part][Verb part][Complement part]**. Relying on punctuation-based segmentation was not possible

due to a lot of noise in the corpus. Therefore the system uses, here again, part-of-speech patterns. Clauses aim to define regions within which we want to detect the cooccurrence of previously located semantic types. If such a cooccurrence occurs, and the two semantic types are linked by a possible semantic relation in the UMLS Semantic Network, the clause is tagged with this possible semantic relation. This completes the tagging step.

2.3 Quantitative Evaluations of the Tagging Process

French language is poorly represented in the UMLS (this situation should also evolve, thanks to the VuMEF initiative [10]). This may cause the above tagging procedure to remain silent in some cases (missing tags). In the same time, a lot of general terms are associated with UMLS concepts, which may cause false tagging. We have evaluated these two aspects of the tagging process by evaluating, on the one hand, for missing tags, the proportion of:

- **false negatives**: phrases that were not tagged, but should have been, given their medical meaning meaning (*e.g.*, "curage ganglionnaire" – *lymph node dissection*);
- **true negatives**: phrases that were rightly not tagged ("cadre stratégique national" for instance);
- **indecision**: ambiguous cases, for which we preferred not to decide ("enfermement familial",or"entourage ultraviolent" for instance).

and on the other hand, for false tagging, the proportion of:

- **correct**: phrases that were tagged according to their meanings;
- **incomplete**: phrases for which not all aspects of the meaning were tagged; *e.g.*, tagging "atteinte gastrointestinale" with T047 (*disease or syndrome*) is incomplete since "atteinte" is tagged, but not "gastrointestinal";
- **incorrect**: phrases for which the meaning of the semantic type is clearly different from the intended meaning of the phrase.

2.4 Selection of a Semantic Relation as a Named Entity, and Performance Analysis

Our experiment aimed at evaluating the performance of UMLS semantic relations as named entities in a QA context. As 53 semantic relations was a too big number for an exhaustive study in a preliminary work, we decided to focus our efforts on a specifically selected relation. The goodness of fit of the corpus tagging with the semantic relation was the main criterion of our selection.

In order to assess this goodness of fit, we assume that the meaning of a clause is closely related to the meaning of its verbal part. Exploratory Data Analysis techniques are natural tools to reveal relationships between the verbs and the projected semantic relations. We tried correspondence analysis and clustering techniques [11, 12], and we finally kept the latter as a selection tool.

In a QA context, the above-chosen semantic relation is used as a criterion to select clauses as answers to medical questions, here related to suitable treatments given some symptoms. The last step of our study therefore consisted in evaluating the precision of answers obtained through this semantic relation.

3. Results

3.1 Quantitative Measures of the Tagging

Table 2a gives the overall density of tags in the corpus. It shows that about 37 % of the noun phrases have been attached to at least one semantic type. The proportion of *missing tags* was evaluated by manually inspecting a random sample of untagged noun phrases. In the same way, the *false tagging* proportion was evaluated by examining a random sample of tagged noun phrases. In each case, the size of the sample was 300 phrases. The results can be read in tables 2, and can roughly been summarized by saying that about half of the phrases were correctly tagged.

Table 2: Evaluation of the tagging

Item	***Number***
Noun phrases occurrences	4101404
Distinct noun phrases	393966
... tagged	147007
... not tagged	246959

(a) Density of tags in the corpus

Item	***Rate***
True negatives	45 %
False negatives	25 %
Indecision	30 %

(b) Missing tags

Item	***Rate***
Correct	52 %
Incomplete	32 %
Incorrect	16 %

(c) Tags correctness

3.2 Semantic Relations and Clause Verbs

An Ascending Hierarchical Classification of the semantic relations, according to the verbs of the clauses in which these relations are suspected to occur (due to the cooccurrence of semantic types) allowed us to investigate the match between the meaning of the relations and the meaning of the clauses. The overall aspect of the obtained dendrogram showed us the most relevant clusters of relations. For each of these groups, we looked for the verbs that contributed the most to the eccentricity of the group. We have postulated that the match was good when the meanings of these verbs were clearly related to the meanings of the relations inside the group.

This analysis led us to focus on the cluster made up by relations *treats* and *prevent*. Table 3 shows the most contributive verbs for that group. French readers can convince themselves of the fit between the medical usage of these verbs and the meaning of the two relations. Finally, we kept only the *treats* relation to pursue our study.

Table 3: Most contributive verbs for {treats, prevents}

Verb	***Eccentricity fraction***	***Cumulated eccent. Frac.***
Envisager (to envisage)	0.0599	0.0599
Traiter (to treat)	0.0465	0.1064
Justifier (to justify)	0.0360	0.1424
Discuter (to discuss)	0.0331	0.1755
Donner (to give)	0.0295	0.2020

Verb	***Eccentricity fraction***	***Cumulated eccent. Frac.***
Modifier (to modify)	0.0275	0.2325
Rédiger (to write)	0.0264	0.2589
Proposer (to propose)	0.0241	0.2830
Recommander (to recommend)	0.0209	0.3039

3.3 The treats *Relation as a Named Entity*

On a random sample, we estimated the precision of the tags obtained thank to the *treats* relation, when it is used alone to extract answers to questions such as “what is the treatment for ...” (which is one of the question types that usually arise during a consultation [2]). We consider that an extraction is successful when the extracted clause contains both the treatment and the symptom or the pathology, and that it fails when the extracted clause is not related to the question, or only contains a truncated answer (for example because of anaphoric references).

A separate sample was drawn from each source making up the corpus. The source by source results are shown in table 4.

Table 4: Precision of treats in extracting treatment-sign associations, ordered by increasing success rate

Source	***Noun phrases number***	***treats-tagged noun phrases number***	***Sampling***	***Successful (%)***	***(%)***
SENAT	199372	1265 (0.6 %)	200	10	2.1
CANADA	90986	2743 (3.0 %)	200	16	2.6
CHUROUEN	10232	230 (2.2 %)	200	19	2.8
UROUEN	14799	621 (4.2 %)	200	20	2.8
AFSSAPS	5187	202 (3.9 %)	202	20	0.0
ANAES	125659	4174 (3.3 %)	200	22	2.9
ORPHA	1460	25 (1.7 %)	25	27	0.0
CANCER	47356	2325 (4.9 %)	200	32	3.3

We can see that tagging with the *treats* relation behaves quite differently depending on the original source to which it is applied. Extreme cases are **[SENAT]** (low tagging rate, and low success rate) and **[CANCER]** (higher tagging rate, and good precision).

The natures of the original sources are quite different from each other: **[SENAT]** is rather legislation-oriented, **[CANADA]** is medically oriented, and aimed to the general public. At the opposite end, **[CANCER]** is a very focused site, intended for specialists. Table 4 shows that the performance of *treats* increases as the intended skills of the users grow higher. Correspondence Analysis, computed on the tagged corpus, has given us more insight of this phenomena, by revealing the structure of the tagged corpus [9].

It is interesting to notice that the differences we have measured among the distinct sources do not come from differences in the density of tagging, but really from differences in the level of specialization of the language used in the documents of each specific source.

4. Discussion and Conclusion

We have presented an experiment in tagging a French corpus with UMLS concepts and, to some extent, with semantic relations, in the context of a Question-Answering system. The amount of incorrect tags (noise) or of missing tags (silence) was evaluated, but a more useful evaluation is linked to an actual task, here QA: the precision of answers generated for *treats*-type questions. A conclusion of this study is that when using the *treats* relation, one is to take into account the individual origin of the documents in which the search is done. Search strategies including this fact now have to be worked out.

Given the status of French in the version of the UMLS that we used, UMLS concept-tagging is very close to MeSH-tagging. The techniques used in the current English- or French-MeSH taggers (see, *e.g.*, [7, 13]) would therefore also be useful in the present task. Besides, the inclusion of an increasing number of French terms and synonyms in the UMLS should enhance the content coverage of French texts; UMLS version 2004AC contains 57,571 French strings, which can be further extended with terms from ICD-10 and from the French translation of SNOMED International [10].

Further work indeed still needs to be done to improve the tagging process, by using more elaborate linguistic tools. This should allow us (i) to use syntactic structures that could be more complex than the approximate clause structure we have used here; (ii) to be able to assert more precisely the presence of a semantic relation in the text, not only by using cooccurrences between semantic types, but also by using dependency analysis of the text; and (iii) to try to take into account anaphora, when this is tractable.

Besides, all statistical aspects of the tagged corpus have not been investigated yet; the dendrogram of the Ascending Hierarchical Classification shows other potentially interesting candidates to study, such as for example the *causes* and *diagnoses* relations. Entities based on these relations may be used to find answers to questions relative to diagnosis.

5. References

[1] Darmoni SJ, Leroy JP, Thirion B, et al. CISMeF: a structured health resource guide. Methods Inf Med 2000;39(1):30–5.

[2] Alper BS, Stevermer JJ, White DS, and Ewigman BG. Answering family physicians' clinical questions using electronic medical databases. J Fam Pract 2001;50(11):960–5. Available at http://www.jfponline.com/content/2001/11/jfp110109600.asp.

[3] Zweigenbaum P. Question answering in biomedicine. In: de Rijke M and Webber B, eds, Proceedings Workshop on Natural Language Processing for Question Answering, EACL 2003, Budapest. ACL, 2003:1–4.

[4] Harabagiu S and Moldovan D. Tutorial on open-domain textual question answering. In: Proceedings of the 19th COLING, Taipei, Taiwan. 24 August–1 September 2002.

[5] Grau B. EQueR: Évaluation de systèmes de Questions Réponses. Project definition, LIMSI-CNRS, Orsay, France, 2002. Part of the EVALDA evaluation initiative, ELDA, Paris, France.

[6] McCray AT and Nelson SJ. The semantics of the UMLS knowledge sources. Methods Inf Med 1995;34(1/2).

[7] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. J Am Med Inform Assoc 2001;8(suppl).

[8] Zweigenbaum P, Baud R, Burgun A, et al. A unified medical lexicon for French. International Journal of Medical Informatics 2004. To appear.

[9] Delbecque T. Structuration de corpus médicaux par l'UMLS. Utilisabilité comme source d'entités nommées pour les systèmes de questions-réponses. Rapport de DEA, Informatique Médicale, Université Paris 5, 2004.

[10] Darmoni SJ, Jarrousse E, Zweigenbaum P, et al. Extending the French part of the UMLS. In: Musen M, ed, Proceedings AMIA Annual Fall Symposium 2003, Washington, DC. AMIA, November 2003:824–. (poster).

[11] Benzécri JP. Correspondances, (vol2). Dunod, 1979.

[12] Benzécri JP. La taxinomie, (vol1). Dunod, 1984.

[13] Ruch P, Baud R, and Geissbühler A. Learning-free text categorization. In: Dojat M, ed, 9th Conference on Artificial Intelligence in Medicine Europe, Cyprus. 2003:199–204.

Address for correspondence

Pierre Zweigenbaum, Mission de recherche en Sciences et Technologies de l'Information Médicale,
STIM/DSI, Hôpital Broussais, 96, rue Didot, 75674 Paris Cedex 14, France
E-mail: pz@biomath.jussieu.fr Url: http://www.biomath.jussieu.fr/~pz/