

Methodology for using the UMLS as a background knowledge for the description of surgical procedures

A. Burgun^{a,*}, G. Botti^b, O. Bodenreider^c, D. Delamarre^a, J.M. Levêque^d, B. Lukacs^e,
D. Mayeux^c, M. Bremond^f, F. Kohler^c, M. Fieschi^b, P. Le Beux^a

^aLaboratoire d'Informatique Médicale, C.H.U., rue H. Le Guilloux, F-35033 Rennes, France

^bService de l'Information Médicale, Hôpital de la Timone, bd Jean Moulin, F-13385 Marseille, France

^cDépartement d'Information Médicale, C.H.U., av du Mal de Lattre de Tassigny, F-54035 Nancy, France

^dClinique de la Sagesse, place Saint-Guénolé, F-35013 Rennes, France

^eService Urologie, Hôpital Tenon, 4 rue de la Chine, F-75020 Paris, France

^fGroupe Image, Ecole Nationale de la Santé Publique, 14 rue du Val d'Osne, F-94410 St Maurice, France

Received 4 August 1995; accepted 6 June 1996

Abstract

The Unified Medical Language System (UMLS) contains and organizes a large number of terms from a variety of biomedical terminology systems. This study examines the relevance of the UMLS content and structures to the specific purpose of the conceptual representation of medical procedures. The MAOUSSC modelling is a compositional formalism with a description of elementary procedures in terms of elementary concept entities and combinations of such descriptions into more complex ones. The UMLS knowledge base is expected to provide semantically categorized medical concepts and interconcept relations. A method to reuse the UMLS has been developed. Quantitative and qualitative results are presented. Some difficulties in reusing the UMLS as a background knowledge are related to the preeminence of some terminology sources and to the instantiation of interconcept links. Other ones suggest that purpose-independence in categorization cannot be achieved.

Keywords: Surgical procedures; Semantic description; Knowledge representation; Unified medical language system (UMLS); Sharability and reusability

* Corresponding author. E-mail: anita.burgun@univ-rennes1.fr

1. Introduction

Diverse medical nomenclatures have been designed: MeSH [1], SNOMED [2], ICD [3], CDAM [4] or CPT-4 [5], each of them for a restricted purpose. Therefore, communication between several systems remains inefficient [6]. For example, in France two coding systems of surgical procedures are in use, one for the hospital DRG system [7], the other one for the National Medical Insurance Services. The MAOUSSC project was initiated in 1992 with the main objective of building a unified representation of medical procedures [8]. A conceptual structure is the most adequate representation to support use of information from different viewpoints [9]. We have implemented a model in which medical procedure terms are represented in a standard format based on a semantic description. The National Library of Medicine has developed the Unified Medical Language System (UMLS) [10] as a compilation of information from a variety of biomedical terminology systems. The UMLS knowledge base transcends its individual sources by reorganizing knowledge into a semantic network. The content and the structure of the UMLS Knowledge Base make it potentially useful for building a core vocabulary for computerized developments applied to specific medical domains. There have been several attempts to reuse the UMLS knowledge base in a variety of contexts [11–16]. We have experimented with the UMLS knowledge base for providing classes, concepts, terms and interconcept links for the MAOUSSC modelling. The idea of reusable lexicons and ontologies in medicine has been extensively discussed by Musen [17]. The notion of knowledge reuse refers to reutilisation of parts of a knowledge base in a significantly different context. This paper describes some results about the relevance of the UMLS content and structures to the specific purpose of the conceptual representation of medical procedures.

2. Background

The UMLS knowledge base includes [18]: (i) a metathesaurus of terms and concepts from many different biomedical vocabularies and classifications. Each main concept is linked to its synonyms, lexical variants and to other concepts which are children or parents or related terms. Each concept in the metathesaurus is assigned to one or more semantic types [19]; (ii) a semantic network of relationships among the semantic categories to which concepts of metathesaurus are assigned [20]; (iii) an information sources map provides information about the content, scope and access conditions for information sources [21,22].

We have been experimenting with the UMLS as the foundation for a concept base through the MAOUSSC project, the goal of which is to implement a formal description of medico-surgical procedures. More precisely, our objectives in the MAOUSSC project are:

(1) The representation of the dimensions that describe a procedure. MAOUSSC modelling is a compositional formalism with a description of elementary procedures in terms of elementary concept entities and relations that may be used to combine such descriptions into more complex ones. Three kinds of information are used to describe medico-surgical procedures. The first one is the set of the basic information that is necessary to describe an elementary procedure. This set can be partitioned into eight categories. Four of them are mandatory: the nature of the action (what is done, e.g. removal, insertion), the topography (the body part or organ the action is applied to), the instrumentation (the equipment used to perform the action, e.g. surgical instruments), the approach (how the anatomic site involved is reached). Four complementary categories can be required according to the value of the nature axis. The

axis matter/device corresponds to the material, organic or not which is moved, it may be a pathologic substance, an anatomical structure, a tissue or a medical device, e.g. a prosthesis; it fits with the direct object for ectomy or insertion verbs. An additional topography is required for a shunt or a pexy for example. It may be used to give precisions in location, for skin for example. A body process is mentioned when the action focuses on a physiologic process. The axis disease is instantiated only for imprecise descriptions of procedures such as treatment of urinary incontinence.

Second, it is obviously impossible to describe some complex procedures with only one value in each axis. As a procedure may be either a single action or a sequence of actions that are associated and performed at the same time, another kind of information is the expression of the links between elementary procedures that make up a complex one. The third kind of information may be named modifiers as they are used in the CPT4. Modifiers are related to the achievement of the procedure, e.g. patient characteristics or emergency.

This modelling is compatible with the European CEN pre-standard [23]. The conceptual structure of the CEN model includes the action, the anatomy, the device or instrument and the pathology. Each description must contain an action and a direct object. Each description must contain the anatomy concept either as a direct object or an indirect object. The pathology concept has to figure only when it modifies the sense of the description. The MAOUSSC model is therefore compatible with the CEN pre-standard. The four basic categories in CEN pre-standard exist in the MAOUSSC structure. Both models also include modifiers and some specific rules.

(2) The formalisation of the rules and the patterns that are associated with this model. We apply a set of rules to construct a formal

description of procedure terms. Moreover, the value of the nature category, which is mandatory, determines which axes are relevant or not. It looks like a verb in a well-formed sentence. For example, biopsy, excision, extraction, aspiration and several other verbs constitute an homogeneous group in which the matter/device axis is always pertinent and has to be instantiated. That group corresponds to the ECTOMY profile. In our model, 18 profiles allow the representation of procedures in surgery, obstetrics and endoscopy fields. For each profile, specific constraints which are both syntactic and semantic have been designed.

(3) The construction of a controlled vocabulary in which concepts are organized in order to have an operational structure for each axis. Rather than building the concept base from scratch, we have used the UMLS metathesaurus and the semantic network to fill in the model for different medical specialities. Our objectives were to capture from the UMLS the terms which are specifically related to the clinical domain and the terms that may be considered as supertypes or meaningful units used for the description of surgical procedures.

The MAOUSSC modelling was initially supported by a computerized system running on Macintosh™. Since 1995, it has been implemented on a Unix server with a Web interface.

3. Methods

3.1. Principles for the description of medical procedures in the MAOUSSC system

The user of the MAOUSSC system that describes the medical procedure terms is a physician. He has to be an expert on the considered medical domain in order to choose the granularity level required for the

MAOUSSC description, to find inconsistencies in existing labels, to give precisions and to point out ambiguous terms. He must be assisted in his task by the system. We will describe the process with the following example: removal of ureteral calculus, transurethral. First, the main concept of the action must be pointed out: it is removal. The selection of the action concept may be assisted by a computerized matching process. Removal belongs to the ECTOMY group; so the description will be continued in accordance with the ECTOMY group profile. Relevant axes must be valued with MAOUSSC concepts, most of them being UMLS concepts. Therefore, the process creates a kind of 3D matrix where lines are action profiles, columns are semantic axes and the third dimension represents medical domains. In our example, removal of ureteral calculus, transurethral, for the axis Topography, the only concepts that are shown to the user belong to the intersection of the set (body part, organ or organ component UNION body location or region UNION body space or junction UNION embryonic structure) and the set (urology). The same principles are applied for every axis.

3.2. UMLS semantic types and MAOUSSC categories for the description of medical procedures

The UMLS semantic types were initially examined to determine which types were relevant to fill up the MAOUSSC categories. Table 1 shows the correspondence between the UMLS semantic types and the categories that we have delineated for the description of medical procedures. For every kind of action, specific constraints restrict the list of the semantic types that

corresponds to each axis. In that way action profiles are implemented. For example, for any procedure corresponding to an ECTOMY action, the matter/device axis, i.e. the direct object of the ECTOMY action, has to be valued with a concept assigned to one of the following UMLS semantic types: 'body substance', 'body part', 'organ or organ component', 'embryonic structure', 'acquired abnormality', 'congenital abnormality', 'biomedical or dental material', 'tissue', or valued with a concept which is an 'implantable device'. Such a table is available for each nature profile. It can be noted that an additional category, 'implantable device', has been added in order to handle the specifics of medical description. The UMLS semantic type 'medical device' is defined as a manufactured object used primarily in the diagnosis, treatment, or prevention of physiologic or anatomic disorders. Metathesaurus examples include 'hip prosthesis', 'oxygenators', 'syringes and obstetrical forceps'. For our purpose, this category was split up in order to distinguish (a) devices which are equipment or tools and correspond to the MAOUSSC axis instrumentation and (b) the subtype that we have named implantable device that subsumes devices which may be left inside the body such as prosthesis. The subtype implantable device is a part of the matter/device axis.

Terms that describe approaches in surgical or medical procedures are missing from the UMLS knowledge base. Thus, we have created a new category, approaches, that subsumes such concepts. Most of the concepts of this category are MAOUSSC concepts and do not come from the UMLS.

It may be remarked that the semantic type 'therapeutic' or 'preventive procedures' encompasses terms that may be either atomic

Table 1
Semantic categories of description in the MAOUSSC modelling and corresponding UMLS semantic types

MAOUSSC axes	Definition	UMLS semantic types
Nature	What is done (the 'verb')	No reference to the UMLS Knowledge Base
Topography	The part of the body to which the action is applied	Body part, organ, or organ component Body location or region Body space or junction Embryonic structure
Instrumentation	The equipment used to perform the action	Medical device
Approach	How the anatomic site involved is reached	(not defined in the UMLS)
Additional topography	For shunt, -pexy or to give precisions for location (fore-skin for example)	(like Topography)
Matter/device	Material, organic or not, or pathologic substance which is moved	Body part, organ, or organ component Body space or junction Body location or region Embryonic structure Tissue Body substance Chemical Inorganic chemical Element or ion Isotope Pharmacologic substance Biologically active substance Hormone Biomedical or dental material Congenital abnormality Acquired abnormality Implantable device (a subtype of Medical Device)
Body process	Physiologic process	Biologic function Physiologic function Organism function Organ or tissue function Cell function
Disease	(Only for an imprecise description of the procedure, for example, treatment of a urinary incontinence)	Congenital abnormality Acquired abnormality Pathologic function Disease or syndrome Injury or poisoning Sign or symptom

terms or complex entities. For example, ‘therapeutic’ or ‘preventive procedures’ is assigned to ‘excision’, as well as to ‘transurethral cryosurgical removal of prostate (postoperative irrigations and aspiration of sloughing tissue included)’. Because of such situations, the MAOUSSC axis Nature—which contains exclusively atomic terms—was not filled up from the UMLS but was created from scratch by the MAOUSSC team after a review of the French procedure coding systems.

3.3. UMLS metathesaurus and medical domains

Since the volume of the metathesaurus is so huge, the first step in designing the MAOUSSC application is to identify the set of terms on which the computerized application will operate for each medical domain. Although the whole UMLS knowledge base is on line, a specific view is drawn for each clinical domain. We have developed a method to select a specific subset of the UMLS knowledge base relevant to a medical domain. Many medical nomenclatures such as MeSH and ICD are partitioned into several fields which are organized within one or several branching hierarchies, each field being related to a medical discipline. The method that we have developed is based on the scanning of those nomenclatures and the capture of the associated terms within the UMLS base. This method was designed as a contribution to the European AIM A2023 MENELAS project [24]. It has been used and applied in the MAOUSSC project.

The first step consists of selecting the appropriate terms within some of the nomenclatures that have been incorporated in the UMLS sources such as MeSH or SNOMED. At present, this step is realized manually although it may be computer-assisted. The selection has to take into account the structure of the nomenclatures. For a tree-structured

nomenclature, we have to select the nodes that are the roots of the sub-trees specifically related to the domain. If a node belongs to several hierarchies, each sub-tree will be explored in the next phase. We start the program by the location in the metathesaurus of the concepts corresponding to the selected codes of the external nomenclatures. This operation enables one to integrate the information-stemming from the other sources of the UMLS-related to those concepts, and to free oneself from the structure of the original nomenclature in order to exploit the whole corpus of the UMLS.

During the second phase, the initial set of concepts is extended in order to build a consistent subset of the UMLS knowledge base. The different kinds of semantic links within the UMLS network are explored as follows:

- (1) Addition of the children of the concepts. S_0 is the initial set of concepts. The CHD relationships are explored by a recursive process until the resulting set, S_1 , is stable.
- (2) Addition of the ascendants. The PAR relationships of every S_1 concept are explored in order to obtain the set S_2 and the process is repeated from S_2 until no new concept is added to S_2 .
- (3) Addition of the related concepts. From S_2 concepts, the RR and UR relationships are explored and the result is the S_3 set. That process is not reiterated.
- (4) Addition of the ascendants of the related terms. The ascendants of the S_3 concepts are captured.

The whole process is entirely automated and computerized. The programs run on UNIX platforms. They have been implemented as C-Shell procedures that operate on the ASCII files of the CD-ROM UMLS Sources distribution [25]. Fig. 1 illustrates the progressive extension of the database from the initial selection of the term cardiovascular system, the code of which is A7 in the MeSH.

4. Results

The MAOUSSC model has been used for the description of 1331 medical and surgical procedures that occur in the discharge summaries of the national survey of hospital activities that is performed by the French Health Statistics Agency. In this context, the methodology described above has already been applied to several surgical domains: urology, cardiovascular system, gynecology and obstetrics, respiratory system, digestive system, endocrine system. They represent almost 45% of the whole set of distinct procedures that were recorded during the survey.

Subsets of the UMLS knowledge base applied to each domain have been designed. A computerized partition of the subsets according to the MAOUSSC axes as shown in Table 1 is complementary to that automated process.

The analysis was applied to the medical sub-bases captured by this method. It was performed in two phases. The first one tested

the presence of relevant concepts and its results are expressed in terms of noise and silence. Noise was judged through a manual review by the MAOUSSC project team. Within the MAOUSSC project, our aim was to build a knowledge base in order to describe medical procedures in a multi-dimensional pattern; so we considered as relevant items the terms closely related to the medical domain, the atomic concepts such as excision or catheter and the concepts which are kinds of generic terms or supertypes of the former ones. Silence was estimated by the number of the terms that we actually needed for the description of medical procedures and that were missing from the sub-base. The item was then searched in the whole metathesaurus. If the item was found in the metathesaurus, the silence was due either to the algorithm or to the lack of links in the metathesaurus.

The second one tried to evaluate the knowledge itself and especially the concepts and the links between them. Quantitative results and qualitative results on concepts and categories are presented.

4.1. Noise and silence in the databases captured by the automated process

The noise is generated by the capture of non-relevant terms from the metathesaurus. It has been evaluated for six clinical domains. We used the fifth version of the UMLS knowledge base during the first year of the MAOUSSC project. Two domains were explored during that period: urology and cardiovascular surgery. For example, the algorithm described above was applied to the eight initial concepts selected from MeSH presented in Table 2 in order to build a base of 'cardiology concepts'. The final vocabulary in cardiology included 2319 concepts from the 1994 edition of the UMLS knowledge

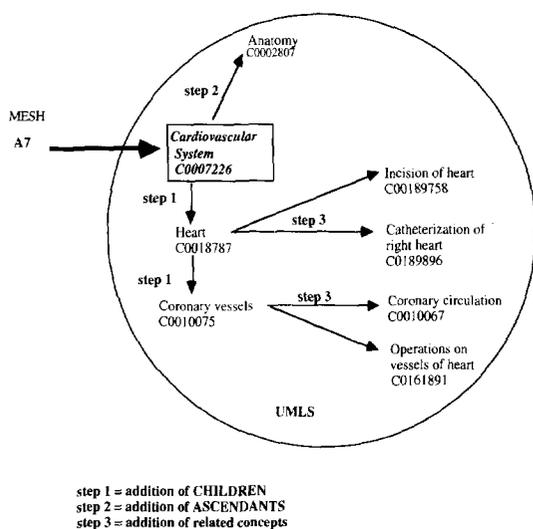


Fig. 1. The progressive extension of the database by the computerized selection process. An example in cardiology.

Table 2

The eight initial concepts of the cardiology vocabulary on which the selection algorithm operated (UMLS 5th edition)

C0004704 S0017695 MSH94 Balloon Dilata-
tion E2.148.102 +
C0007168 S0022395 MSH94 Cardiac Pacing, Arti-
ficial E2.145
C0007226 S0022490 MSH94 Cardiovascular Sys-
tem A7 +
C0007436 S0022958 MSH94 Catheterization, Periph-
eral E2.148.224 +
C0011904 S0032361 MSH94 Diagnosis, Cardiovascu-
lar E1.145 +
C0013778 S0036878 MSH94 Electric Counter-
shock E2.344
C0018795 S0047215 MSH94 Heart Catheteriza-
tion E2.148.442 +
C0038897 S0089726 MSH94 Surgery, Cardiovascu-
lar E4.752 +

sources. Among the 2319 concepts of our cardiology sub-base, 1582 (68%) were found to be relevant to this domain, either they were specific or not. Similar results were obtained for the urology sub-base with 69% of relevant terms. Noise was due to the capture of terms from other clinical domains, in most cases 'close to' the target domain, e.g. gynecology for urology, but sometimes quite distant. Noise may also arise from the fact that no condition about the semantic types was included in the algorithm. This kind of noise is hidden from the user as the MAOUSSC application ignores terms from non relevant semantic types such as 'nucleic acid', 'nucleoside', or 'nucleotide' or 'gene' or 'genome'. When taking into account the exclusion of the concepts that are assigned to non relevant semantic types, almost 16% (293) concepts were still out of the cardiology domain.

A significant improvement occurred with the sixth edition of the UMLS knowledge base. Since the sixth version, the metathesau-

rus has encompassed the complete set of SNOMED International terms. Rather than selecting MeSH terms, we have made the algorithm operate on a set of SNOMED anatomical terms. From our experience in urology and cardiology, we noted an increase in noise during the last step of the algorithm. So the computerized process has been limited to the first three steps: (1) from an initial set of concepts, addition of the descendants, the result is the set S_1 ; (2) from S_1 , addition of the ascendants, the result is S_2 ; (3) from S_2 , addition of the related concepts, the result is the final set for the given medical domain. Because of those modifications that concern the initial selection of nomenclature items and the algorithm itself, the rate of the captured concepts that were out of the domain has decreased to below 5%.

Reducing the silence is a priori a more difficult task. Since the method is based on the exploitation of the links between the concepts within the UMLS, terms will not be represented in the sub-base when relationships are missing from the metathesaurus. That is frequently the case for medical instruments. Although they are present in the metathesaurus, generally coming from the unified medical device nomenclature, some of them are not linked to concepts from a specific medical domain. Therefore, such terms are not captured by the automated process. Nevertheless, the MAOUSSC application is now supported by a relational database and a web interface. As the whole UMLS terminology is on-line, the user can search for the considered term in the UMLS and, if the concept exists in the metathesaurus, he can add it to the relevant medical domains. For example 'sphincterotome', even if it does exist in the metathesaurus, is not found in the initial gastroenterology database because it is not linked to any gastroenterology concept. The user must add the term 'sphincterotome'

to the 'gastroenterology' domain. Except for additional MAOUSSC semantic types, i.e. implantable device, the user does not need to assign sphincterotome to a MAOUSSC category: the belonging to a MAOUSSC axis is inferred for the UMLS semantic type of sphincterotome. Other concepts do not exist in the UMLS knowledge base. That was the case for anatomic precisions like 'vesical neck' or 'upper third of esophagus' until the release of the sixth version of the UMLS knowledge base. In the latest medical domains that have been described, excluding the surgical approach axis and the modifiers, the number of concepts that were really needed and therefore created de novo was about ten.

4.2. Report on the UMLS contents faced with some MAOUSSC requirements

4.2.1. Terms and concepts

For Mac Cray [26], metathesaurus concepts are essentially clusters of synonymous terms. One problem in assessing the reusability of the UMLS terminology is to deal with equivalent terms and synonyms. X is a synonym of Y if any proposition P containing X has equivalent truth-conditions to another proposition P', which is identical except that X is replaced by Y. A broader conception of the notion of synonymy can be developed which is based on a scale of synonymy. Synonymous terms are thus defined by the conjunction of two properties: (1) they manifest a high degree of semantic overlap and (2) they have a low degree of implicit contrastiveness. Synonyms are identical in respect of major, central semantic traits and differ only in respect of minor, peripheral traits. Since they differ in respect of semantic traits, a pair of synonymous terms can be incompatible, compatible, or hyponym and superordinate. Cruse therefore speaks of plesionyms [27]. Plesionyms yield sentences with

different truth-conditions: two utterances which differ only with regard to plesionyms are not mutually entailing, although if the terms are in a hyponymous relation there may well be unilateral entailment. Clinical distinctions regarding diagnoses or procedures are missing from Meta-1. For example 'cryptorchism' and 'ectopic testis' are synonymous in Meta-1. This pair of items bare a special sort of semantic resemblance to one another. It is customary to call such items synonyms for bibliographic purposes. But for a clinical purpose, they just manifest a high degree of semantic overlap and a for a physician there is an incompatibility between the two terms.

Another example is given by the pair of terms 'deformity', 'acquired' and 'traumatic deformity' that are considered synonymous in the metathesaurus. More exactly, they should constitute a plesionymous pair. There is one member of the plesionymous pair which is possible to assert while simultaneously the other member is denied: 'it is an acquired deformity but it is not a traumatic one' whereas 'it is a traumatic deformity' entails 'it is an acquired deformity'. The two concepts bare capital traits but 'traumatic deformity' is an hyponym of the other term.

The characterization of permissible differences between synonyms depends on the final purpose. The above examples illustrate the assertion that the intended use of the UMLS to facilitate information retrieval may erase valuable information about semantic properties of terms.

4.2.2. Ontology

An ontology can be regarded from an implementation viewpoint. Mac Cray said that the UMLS Semantic Network [28] 'represents knowledge about the biomedical domain and may be considered a basic ontology for that domain'. A basic ontology should be orga-

nized around a basic level which may become a sharable core for classifications. According to Rosch [29], a basic category is clearly differentiated from other ones and its category resemblance—which is defined as the sum of all the common features within that category, minus the sum of all the distinctive features that belong to only some members of the category, as well as those belonging to contrasting categories—is maximized at that level of abstraction. In the unified semantic network some semantic types are not basic level categories. For example, a ‘diagnostic procedure’ is a procedure, method, or technique used to determine the nature or identity of a disease or disorder; a ‘therapeutic’ or ‘preventive procedure’ is a procedure, method, or technique designed to prevent a disease or a disorder, or to improve physical function, or used in the process of treating a disease or injury. Those two categories obviously manifest a high degree of definition overlap. They are cognitively subordinate of an intermediary and basic type ‘procedure’. To categorize a concept means to consider it, for the purpose of that categorization not only equivalent to other concepts in the same class but also different from concepts belonging to other categories. The fragility of the UMLS categorization of concepts is more important from an implementation point of view. That may be exemplified by the fact that pleura and peritoneum are both ‘tissues’ while meninges and pericardium are ‘body parts’ or ‘organs’. Though they belong to distinct UMLS Semantic Types, pleura, peritoneum, meninges and pericardium share a lot of properties. Another example is given by the fact that the peritoneal cavity is a ‘body space’ or ‘junction’ while the pleural cavity is a ‘body location’ or ‘region’.

5. Discussion

5.1. A required improvement of the MAOUSSC application: making the description more automated, limits of the UMLS knowledge base

From our experiment in describing 1331 distinct procedures from the set of the 2985 distinct medico-surgical procedures that occur in the 80 000 discharge summaries of the French national survey of hospital activities, it is obvious that a main limitation of the MAOUSSC application is the time required for the description of the medical procedures by the user. At the present time, 55% of the whole set of distinct procedures remain to be described. Our objective is to assist the physician in that tedious task. In order that the UMLS knowledge base can be used to support an automated description of the MAOUSSC elementary procedure terms, two conditions are required: first, the metathesaurus must include both complex entities referring to medical procedures and atomic concepts corresponding to axis values, and second, the interconcept links must be operational. We are going to present the preliminary analysis of those two points as a starting point for a further computer development.

The granularity of the concepts in Meta-1 is the granularity that exists in the source vocabularies. A concept in the metathesaurus may be ‘transurethral cryosurgical removal of prostate (postoperative irrigations and aspiration of sloughing tissue included)’ which refers to a composed notion as well as ‘excision’ which is an atomic concept. Transurethral cryosurgical removal of prostate is a procedure and one of its properties is to be an ‘excision’, which is a value of the MAOUSSC Nature axis, applied to the ‘prostate’, which is a topography value.

Table 3
 Instanciation of the MAOUSSC axes according to the ECTOMY profile for the description of transurethral prostatectomy.

MAOUSSC category	ECTOMY profile	Value
Nature	Mandatory	ECTOMY group
Topography	Mandatory	Prostate
Instrumentation	Mandatory	Resectoscope
Approach	Mandatory	Transurethral
Additional topography	Optional	None
Matter/ Device	Mandatory	Prostate ^a
Body process	Not relevant	None
Disease	Not relevant	None

^aThe default value for matter/device axis for the ECTOMY profile is the value of the topography axis.

The interconcept links are expected to be used for inference. By inference, we mean an automated process that captures the relevant set of elementary concepts which describe a surgical procedure. For example, a transurethral prostatectomy is described by the instanciations of the MAOUSSC axes shown in Table 3. Its context in the metathesaurus may be visualized in Fig. 2.

As the process is based on the computerized exploration of interconcept relationships, its performance will be a function of the

exhaustiveness and the quality of the links. The metathesaurus interconcept links are (1) hierarchical links present in the source nomenclature, (2) links toward terms that are related to the entry in question. Three variants of the second type of relation are implemented in the metathesaurus: they are qualified as broader (B in Fig. 2), narrower (N) and other (O). A first obstacle is the fact that all the MAOUSSC axes do not exist in the UMLS Semantic Network. Thus there is no information in the metathesaurus that could be used to fill the approach axis.

Other difficulties are related to the nature of the UMLS interconcept relationships. Hierarchical links in the UMLS knowledge base are labelled as: 'is a', 'part of', 'manifestation of', 'is surrounded by', etc. In the last cases, the UMLS hierarchical links are inconsistent, they just express the principle of differentiation applied specifically at this node of the original classification instead of referring to well-known structures as taxonomies or meronomies. In addition, some related (RR) terms which belong to the same semantic category are closely related and the link between the concepts should rather be a kind of hyponym/hyperonym relation. A broader term is like an hyperonym and a narrower term is like an hyponym. For example, we should have: 'prostatectomy, transurethral is a prostatectomy'. In the same way, we should have: 'prostatectomy is a excision'. Fig. 2 shows that in both cases, metathesaurus implemented links are RR ones. Because of those distinct labels for the same absolute semantic link, it is difficult to have a standardized exploitation of interconcept links.

Another problem is the distance between the terms, appreciated by the number of nodes between the concepts in the metathesaurus network. In our example, 'transurethral prostatectomy' is linked with 'prostatectomy' which is linked to 'prostate'

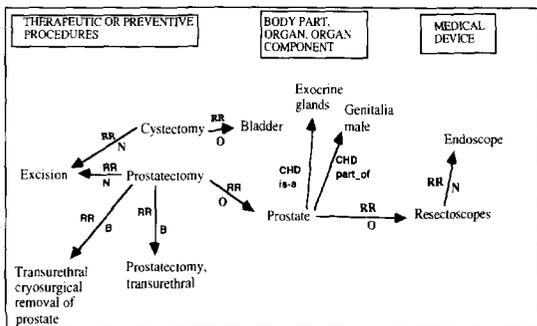


Fig. 2. A sample of the metathesaurus conceptual network.

which is linked to 'resectoscope'. It is therefore impossible to get automated inference from the metathesaurus that a resectoscope is the instrument used to perform a transurethral prostatectomy.

Finally, it may be the case that clinical expertise is required in order to instantiate some of the MAOUSSC axes.

5.2. Reusing the UMLS system versus starting building a new specialized conceptual system

Despite the fact that the UMLS is not yet complete and still has some deficiencies, we consider it a useful tool for terminology applications. Although the metathesaurus is organized by concept, it is built by integration of existing terminologies, classifications and other sources. Thus, a concept is an object which is found in one or more original sources. Close to Wittgenstein's viewpoint [30,31], for Mac Cray, the cluster of terms can be viewed as the meaning of that concept. A concept is given by the union of its instantiations in medical terminologies. This is similar to Wittgenstein's approach where the concept 'game' is the set of the existing games. We think that such an approach that intends to unify existing terminologies rather than trying to build a universal system is a powerful one [32].

From a pragmatic point of view, an automated process for designing a subset applied to a clinical domain has been implemented. It is easily performed and can be repeated on the successive versions of the UMLS knowledge base without adaptation on condition that the structure of the files is identical. The first results were not encouraging but the selection of the initial set of terms on which the algorithm works has got better with experience. Another significant improvement is related to the inclusion of SNOMED III into

the UMLS knowledge base since the most recent releases. That process brings an initial corpus of terms that has been considered by the MAOUSSC team as acceptable in the latest medical domain building. In the recent implementation of the MAOUSSC application based on a relational database and a web interface, the user can access the whole metathesaurus and thus can reuse any term of it.

Nevertheless, it can be noted that for urology and cardiology, only 365 terms were needed by the physicians for the MAOUSSC representation of the surgical procedures in those medical domains. Such a low number of terms may be considered as a factor that reduces the interest of reusing an existing terminology.

Over the building of a corpus of terms, we have to deal with the use of such a set of terms as knowledge. For Mac Cray [26], the value of the metathesaurus lies in the multitude of connections it provides between and among medical concepts. The challenge is to use them in knowledge representation applications. From our experience, despite their large number, the use of UMLS interconcept links as knowledge is limited by their non-standardized labelling. What is operational for information retrieval purposes cannot be used for inferences in knowledge representation.

6. Conclusion

We have developed a model for the description of surgical procedures based on a semantic representation. This modelling has been used to describe the surgical procedures present in the 80 000 discharge summaries of the French national survey of hospital activities and is currently being extended to all the medical fields. We have tried to reuse the

UMLS knowledge base in order to support our terminology. The UMLS is a repository of concepts, classes and relationships which occur in medical information sources. Each object in the UMLS keeps its personal identity: source, parents and children in the original hierarchy. Inadequacies may exist in the source vocabulary and thus entail inadequacies in the UMLS. In our experience, the main limitation in the re-use of the UMLS concerns the interconcept links because of the lack of standardization of the relationships labels. Nevertheless, having a common kernel with UMLS provides an initial corpus of classes and concepts and contributes to the knowledge unifying process.

Acknowledgements

The MAOUSSC project was supported by the SESI (Statistics and Information Systems Department) of the French Ministry of Public Health. The authors thank the National Library of Medicine for having provided the UMLS Knowledge Sources on CD-ROM.

References

- [1] National Library of Medicine: *Medical Subject Headings*, Bethesda, Maryland, 1986.
- [2] Cote RA (Ed.): *Systematized Nomenclature of Medicine*. 3rd edn., College of American Pathologists, Skokie, Illinois, 1994.
- [3] World Health Organization: *International Classification of Diseases*, Genève, 1993.
- [4] Ministère de la Santé et des Affaires Sociales: *Catalogue Des Actes Médicaux*. Bulletin Officiel 93–13 bis, Paris, 1994.
- [5] American Medical Association: *Physicians' Current Procedure Terminology*, AMA, Chicago, 1994.
- [6] Burgun A, Le Beux P, Brémond M, Lenoir P: Translation from one medical nomenclature into another: a categorization of the problems. In: *Proc MEDINFO 92*, (Eds: KC Lun, P Degoulet, TE Piemme and O Rienhoff), North Holland: Amsterdam, 1992, pp 1458–1461.
- [7] Fetter R, Shin YS, Freeman JL, Averill RF and Thompson JD: Case mix definition by diagnoses related groups. *Med Care*, 18 (2) (1980) 1–53.
- [8] Burgun A, Botti G, Lukacs B, Mayeux D, Seka LP, Delamarre D, Bremond M, Kohler F, Fieschi M and Le Beux P: A system that facilitates the orientation within procedure nomenclatures through a semantic approach. *Med Informatics*, 19 (4) (1994) 297–310.
- [9] Rector AL, Nowlan WA and Kay S: Conceptual knowledge: the core of medical information system. In: *Proc MEDINFO 92*, (Eds: KC Lun, P Degoulet, TE Piemme and O Rienhoff), Elsevier Science, DV, 1992, pp. 1420–1426.
- [10] Lindberg DAB, Humphreys BL and Mac Cray AT: The Unified Medical Language System. *Methods Inf Med.*, 32 (1993) 281–291.
- [11] Cimino JJ and Barnett GO: Automated translation between medical terminologies using semantic definitions. *MD Computing*, 7(2) (1990) 104–109.
- [12] Cimino JJ, Johnson SB, Ping Peng and Aguirre A: From ICD9–CM to MeSH using the UMLS: A how-to guide. In: *Proc 17th Ann Symp Comp Appl Med Care*, IEEE Computer Society Press, Washington DC, 1993, pp. 730–734.
- [13] Volot F, Zweigenbaum P, Bachimont B, Ben Said M, Bouaud J, Fieschi M and Boisvieux JF: Structuration and acquisition of medical knowledge: using UMLS in the conceptual graph formalism. In: *Proc 17th Symp Comp Appl Med Care*, IEEE Computer Society Press, Washington DC, 1993.
- [14] Friedman C: The UMLS coverage of clinical radiology. In: *Proceedings of the 17th Symposium Computer Applications in Medical Care*, McGraw Hill, New York, 1993, pp. 307–316.
- [15] Cimino JJ: Representation of clinical laboratory terminology in the Unified Medical Language System. In: *Proc 15th Symp Comp Appl Med Care*, McGraw Hill, New York, 1991, pp. 199–203.
- [16] Cimino JJ: Use of the Unified Medical Language System in Patient Care at the Columbia-Presbyterian Medical Center. *Methods Inf. Med.*, 34 1/2 (1995) 158–64.
- [17] Musen MA: Dimensions of knowledge sharing and reuse. *Comp. Biomed. Res.*, 25 (1992) 435–467.
- [18] National Library of Medicine: *UMLS Knowledge Sources*, 5th Exp. edn. Bethesda, MD, 1994.
- [19] Tuttle MS, Nelson SJ, Fuller LF, Sheretz DD, Erlbaum MS, Sperzel WD, Olson NE and Suarez-Munist ON: The semantic foundations of the

- UMLS Metathesaurus. In: *Proc MEDINFO92* (Eds: KC Lun, P Degoulet, TE Piemme and O Rienhoff), Elsevier Science, DV, 1992, pp. 1506–1511.
- [20] Mac Cray AT: The UMLS Semantic Network. In: *Proc Symp Comp Appl Med Care* (Ed: LC Kingsland), IEEE Computer Society Press, Washington DC, 1989, pp. 503–507.
- [21] Masys DR, Humphreys BL: Structure and function of Information Sources Map. In: *Proceedings of MEDINFO 92* (Eds: Lun, Degoulet, Piemme, Rienhoff), Elsevier Science Publishers, DV, 1992, pp. 1518–1521.
- [22] Miller PL, Wright LW, Frawley SJ, Clyman JJ, Powsner SM: Selecting relevant information resources in a network-based environment; the UMLS information sources map. In: *Proc MEDINFO 92* (Eds: KC Lun, P Degoulet, TE Piemme and O Rienhoff), Elsevier Science, DV, 1992, pp. 1512–1517.
- [23] European Committee for standardization: *Technology and coding systems of medical procedures. Part 1: surgical procedures* TC251 1993, N93–213.
- [24] Consortium MENELAS, Zweigenbaum P: MENELAS: accès à des comptes-rendus d'hospitalisation en langage naturel. In: *Actes des Cinquièmes Journées Francophones d'Informatique Médicale*, 1994, pp. 23–30.
- [25] Burgun A, Delamarre D, Botti G, Lukacs B, Mayeux D, Bremond M, Kohler F, Fieschi M and Le Beux P: Designing a sub-set of the UMLS knowledge base applied to a clinical domain: methods and evaluation. In: *Proc 18th Symp Comp Appl Med Care*, McGraw Hill, New York, 1994, pp. 968.
- [26] Mac Cray AT and Nelson SJ: The semantics of the UMLS knowledge sources, Communication, *International Working Conference: IMIA WG6*, Geneva, 1994.
- [27] Cruse DA: *Lexical Semantics*, Cambridge University Press, Cambridge, 1986.
- [28] Mac Cray AT and Nelson SJ: The representation of meaning in the UMLS. *Methods Inf. Med.*, 34 1/2, (1995) 193–202.
- [29] Rosch E: *Principles of categorization in Cognition and Categorization*, L. Erlbaum Associates, Hillsdale, NJ, 1978, pp. 27–48.
- [30] Wittgenstein L: *Investigations philosophiques*, Gallimard, Paris, 1953.
- [31] Wittgenstein L: *Le Cahier bleu et le Cahier brun*, Gallimard, Paris, 1958.
- [32] Burgun A: Représentation conceptuelle du langage médical appliquée aux nomenclatures internationales, *Thesis*: University of Rennes, 1994.