Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-164

Method for mapping the French CCAM terminology to the UMLS metathesaurus

Cédric BOUSQUET[ab1], Julien SOUVIGNET[ab], Tayeb MERABTI[d], Eric SADOU[a], Béatrice TROMBERT[ab], Jean-Marie RODRIGUES[abc]

[a]*Department of Public Health, CHU University of Saint Etienne, France*
[b]*INSERM U872, Eq. 20, Paris, France*
[c]*WHO FIC Collaborative Centre for International Classifications in French Language, Paris, France*
[d]*CISMeF, University Hospital, Rouen, France & TIBS, LITIS EA 4108, Institute of Biomedical Research, France*

Abstract. The French coding system of surgical procedures, the *Classification Commune des Actes Médicaux* (CCAM), is used in France for DRG databases and fee for services payment. Mapping between CCAM and other clinical procedures terminologies by the means of UMLS metathesaurus is essential in order to increase semantic interoperability between different healthcare terminologies and between different case mix systems. In a previous work the CISMeF team used an automatic approach to map CCAM descriptors to the French part of the UMLS metathesaurus. In another way for the French funded research project InterSTIS, we performed a mapping using MetaMap based on the top level semantic structure descriptors of anatomy and action of CCAM translated from French to English. This paper aims to present this new approach and to compare the results with the previous one. The combination of both approaches significantly improved the coverage of the mapping to 68 % for both descriptors and 95 % for at least one descriptor.

Keywords. Semantic Interoperability, Mapping, Terminology, Ontology, Coding System, Multilingualism, Clinical procedure.

Introduction

A French coding system of clinical procedures, the *Classification Commune des Actes Médicaux* (CCAM), has been developed between 1996 and 2001 for DRG databases and fee for service payment [1]. CCAM codes are related to descriptors giving relevant information about the procedure, defining the system, tract or anatomy structure, the action performed and the access mode or typical means.

The purpose of NLM's Unified Medical Language System® (UMLS) is to facilitate the integration and interoperability of clinical terminologies in computer systems. The main component of UMLS is the metathesaurus which contains more than 140 biomedical terminologies, classifications, and ontologies and proposes mappings between these terminologies [2].

CCAM is not yet included in the UMLS metathesaurus. Mapping between CCAM and other clinical procedures terminologies by the means of the UMLS metathesaurus is essential in order to increase semantic interoperability between different healthcare

[1]Corresponding author: DSPIM, Bâtiment CIM 42 - Hôpital Nord, Chemin de la Marandière, 42055 Saint Etienne, France; E-mail : cedric.bousquet@chu-st-etienne.fr

terminologies within and across different national languages and between different case mix systems.

In a previous work the CISMeF (*Catalogue et Index des Sites Médicaux de langue Française*) team tried to map the CCAM descriptors to the French terms present in UMLS [3] using the FMTI+ tool [4]. In parallel, we tried a new method where CCAM descriptors are first translated from French to English, and then mapped to UMLS using MetaMap [5]. The objective of this paper is to present this new method for aligning CCAM descriptors with UMLS, and compare the results obtained with this method to those obtained by CISMeF.

1. Material and Methods

1.1. CCAM

CCAM codes have a defined seven characters structure. The first letter designates the system or tract which the action concerns. The second letter provides additional detail such as a function or an organ. The third letter indicates the action performed. The fourth letter indicates the access mode. Finally, three sequential numbers are used to differentiate procedures with four identical letters.

Each code corresponds to multiple descriptors: the first two letters stand for one or more anatomical sites, and the other letters describe one or more action terms and one or more access mode.

Table 1 contains excerpts from the English translation of descriptors for a selected CCAM code: HEQE001 (oesophagogastroduodenoscopy).

Table 1. Excerpt of the English translation of descriptors.

DIGESTIVE SYSTEM	**H**
Lips, tongue, oral cavity as a whole	HA
Teeth, periodontium, gums	HB
…	H_
Oesophagus	**HE**

EXAMINE	**Q**
Descriptors : Analysis, Detection, Study, Examination, Exploration monitoring, Search, Surveillance, -scopy…	

TRANSORIFICE ENDOSCOPIC ACCESS	**E**
Descriptors: Endoscopy, Fibroscopy, Laryngoscopy, Jejunoscopy, Ileoscopy, Colposcopy, Anterograde…	

The first letter *H* is related to *Digestive System* while the first two letters *HE* are related to a more specific site: *oesophagus*.

The third letter *Q* indicates the action: *Examine*. CCAM action descriptors have been defined by grouping them according to action type; each is identified by an action verb (*examine*). The same code is often allocated to several technically similar action verbs (*examine* is in the same group as *measure* or *record*).

The fourth letter *E* indicates the access mode: *Transorifice endoscopic access*. Several descriptors are also allocated for each defined CCAM access mode.

1.2. From CCAM codes to English descriptors

CCAM (v23) was downloaded from the *Assurance Maladie En LIgne* (French National health insurance-AMELI) website [6]. CCAM codes about additional fees were not kept and this study relies on 7,583 CCAM codes. French descriptors [7] have been

translated in English by a German team within the WHO-FIC (WHO Family of International Classifications) network [8].

CCAM includes 194 descriptors for anatomy and 331 descriptors for action. The retrieval of anatomical information was straightforward using the two first letters of the CCAM code. The information on the action using the third letter was more difficult to identify because a single letter was provided in the CCAM code to represent several actions. Therefore we relied on CCAM French labels to identify and retrieve the relevant action verb within the label rather than relying on the third letter. The action term is usually introduced in the initial part of the label, and the access mode and approach are usually at the end of the label. Some rules were applied: keeping the first words in the label as the action or identifying the suffix of the first word (e.g. *-ectomy* in *thyroidectomy*). Access mode descriptors have not been considered in this study.

1.3. Mapping the descriptors to UMLS using MetaMap

MetaMap is a software developed to map biomedical texts to the UMLS. The MetaMap algorithm consists in generating synonyms, derivational variants, acronyms, meaningful combinations, and finally inflectional and spelling variants for a given text. These variants are compared to concepts in the metathesaurus to build a candidate set ordered according to a mapping strength that is a measure of similarity.

We used MetaMap Transfer (MMTx 2009), a Java-Based distributable version of the MetaMap software on CCAM translated descriptors and considered only exact matches with a 100% mapping strength.

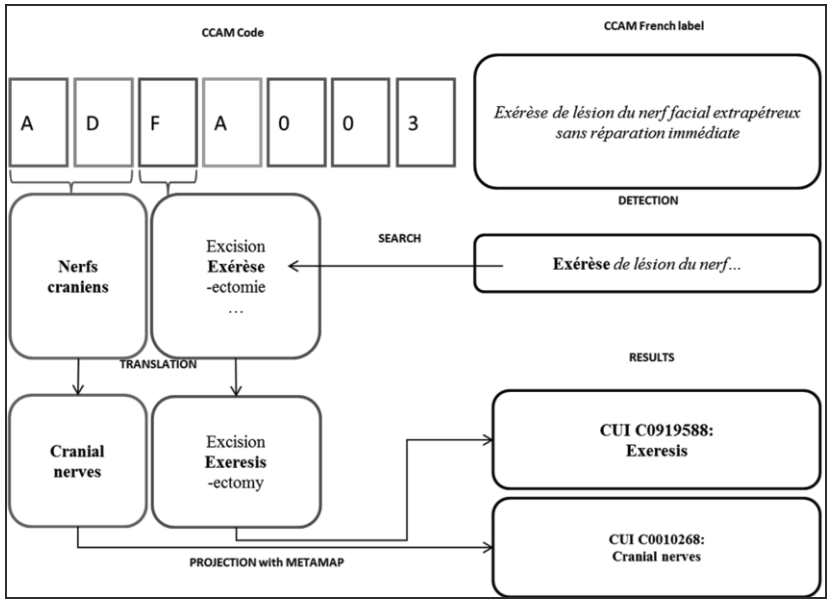

Figure 1. Example using MMTx

Figure 1 shows an example of mapping a CCAM label (Exeresis of lesion of the extrapetrous facial nerve without immediate reparation). The action *exérèse* is detected in the label, and then translated in *exeresis* which is mapped to the *Exeresis* concept

unique identifier (CUI) in the metathesaurus. For the anatomy, the letters AD lead to the descriptor *Nerfs craniens*, which is mapped to *cranial nerves* CUI**.**

1.4. Additional processing of anatomy descriptors

Among the 194 CCAM anatomical descriptors labels, 68 describe several anatomical locations. For example the AB descriptor is "Intracranial ventricles, meninges and cerebrospinal fluid". CCAM codes associated to this descriptor may describe procedures that are located in the intracranial ventricles, meninges, or cerebral fluid. The clinical coders of Saint Etienne University Hospital analyzed 3,455 procedures (45.5%) corresponding to these 68 descriptors in order to associate a unique and main anatomical location to each procedure.

2. Results

Our alignment method of CCAM descriptors to UMLS using MetaMap allows mapping 205 out of 331 (62%) action descriptors and 91 out of 194 (49%) anatomical descriptors. This significantly improves the results of our previous approach where the descriptors were mapped to the French part of the UMLS metathesaurus using FMTI+. Table 2 shows the count of CCAM procedures where descriptors are found for each method.

Table 2. CCAM codes with at least one descriptor mapped and with both descriptors mapped for both approaches (FMTI+ and MM=MetaMap).

	action descriptor			anatomical descriptor			at least one descriptor			both descriptors		
method	MM	FMTI+	All	MM	FMTI+	All	MM	FMTI+	All	MM	FMTI+	All
CCAM codes	5,273	3,253	6,168	4,592	3,668	6,167	6,577	4,878	7,208	3,288	2,043	5,127
%	70%	43%	81%	61%	48%	81%	87%	64%	95%	43%	27%	68%

Results are improved when using the English translation of CCAM descriptors with MetaMap. The coverage rate exceeds 50% on both action and anatomical descriptors: 5,273 codes (70%) have at least the anatomical mapping, and 4,592 (61%) have at least one mapping for action whereas FMTI+ could not map descriptors in more than 50% of CCAM procedures.

Both approaches can be combined to use, for example, the mapping of anatomical descriptor with FMTI+ and the action descriptor identified by MetaMap, or vice versa. This combination significantly improves the coverage of the mapping to 68 % for both descriptors and 95 % for at least one descriptor.

3. Discussion

The English part of UMLS contains more terms than the French part. This may explain why mappings are more numerous using MetaMap rather than FMTI+. The results of each method are quite complementary. This may be related to additional treatment

performed with the FMTI+ tool that includes removing stop words and steaming in French language while we kept only exact matches with MetaMap.

Our method is efficient for describing the action performed but the anatomical descriptor is sometimes more coarse-grained than the exact anatomical site (in the example from figure 1, *cranial nerve* instead of *extrapetrous facial nerve*). Also, some procedures involve several anatomies, but a single descriptor (the main anatomical location) is defined in the CCAM code. This leads to a single mapping where a multiple mapping would be necessary. In both cases, there is a partial loss of information.

Some additional works are needed: first to compare the coherence of the two methods and their degree of agreement. The second work presently ongoing is to take into account CCAM labels that consist of several procedures, e.g. *Scanography of skull and its content and of the thorax, with intravenous injection of contrast product*.

We also need to provide mappings for access mode: It may identify just the approach method, particularly in invasive approaches; in other cases, it designates both approach method and technique.

4. Acknowledgments

We acknowledge the WHO-FIC network and namely Albrecht Zaiss and Susan Hanser for providing the document that provides French-English translations of the descriptors. This work was partially granted by the InterSTIS project funded by the French National Research Agency (ANR-07-TECSAN-010).

References

[1] Trombert-Paviot B, Rector A, Baud R, Zanstra P, Martin C, van der Haring E, Clavel L, Rodrigues JM. The development of CCAM: the new French coding system of clinical procedures. HIM J. 2003; 31(1):1-11.

[2] Humphreys BL, Lindberg DAB, Schoolman HM, Barnett GO. The unified medical language system: An informatics research collaboration. J Am Med Inform Assoc 1998; 5:1-11.

[3] Merabti T, Massari P, Joubert M, Sadou E, Lecroq T, Abdoune H, Rodrigues JM, Darmoni SJ. An automated approach to map a French terminology to UMLS. Stud Health Technol Inform. 2010; 160(Pt 2): 1040-4.

[4] Pereira S, Neveol A, Kerdelhué G, Serrot E, Joubert M, Darmoni SJ. Using multi-terminology indexing for the assignment of MeSH descriptors to health resources in a French online catalogue. AMIA Annu Symp Proc. 2008: 586-90.

[5] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. Proc AMIA Symp 2001: 17–21.

[6] http://www.ameli.fr/accueil-de-la-ccam/telechargement/index.php (Last Accessed June 6th 2012)

[7] Ministère de la santé, de la jeunesse et des sports. Classification commune des actes médicaux. Guide de lecture et de codage. Annexe I de l'arrêté du 27 février 2007. Bulletin officiel No 2007/3 bis. Available from: http://www.sante.gouv.fr/adm/dagpb/bospecia/bos2007/bo0703.pdf (Last Accessed June 6th 2012)

[8] WHO Family of International Classifications) network. Common Classification of Medical Procedures. Coding System. Internal working draft.