MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1036

Translating ICD-11 into French using lexical-based approach: a preliminary study

Tayeb Merabti[a], Julien Grosjean[a], Jean-Marie Rodrigues[b,c], Stefan Jacques Darmoni[a,b]

[a] *CISMeF & TIBS, LITIS EA 4108, Rouen University Hospital, Rouen, France*
[b] *INSERM, U1142, LIMICS, F-75006, Paris, France*
[c] *Department of Public Health, Saint Etienne University Hospital, France*

Abstract

To translate the 11th edition of the International Classification of Diseases (ICD-11) into French, we proposed a lexical approach using Natural Language Processing techniques. This method relies on the 56 biomedical terminologies and ontologies included in the Cross-lingual Health Multiple Terminologies and Ontologies Portal. From a sample of 336 ICD-11 terms, the algorithm translated 164 (49%) terms into at least one French term each.

Keywords:
Coding System, Mapping, Multilingualism, Semantic Interoperability, Terminology as Topic.

Introduction

The 11th edition of the International Classification of Diseases (ICD-11) is ongoing and its publication is not expected until 2017 [1]. Currently, ICD-11 is neither tranlated into French nor included in the Unified Medical Language System (UMLS). Thus, we have attempted to apply a lexical approach to translate each term from a sample of ICD-11 terms into at least one French term.

Materials and Methods

To translate ICD-11 terms, we proposed a lexical approach using Natural Language Processing (NLP) techniques. This method relies on 56 biomedical terminologies and ontologies (BMTO) included in the Health Multiple Terminologies and Ontologies Portal (HeTOP) [2]. Compared to UMLS, which includes 164,071 French terms, HeTOP includes 349,311 French terms[1] from BMTO.

This approach [3] helps to normalize all the English terms from the bilingual BMTO (English and French) included in the HeTOP. An algorithm was developed to find the terms most lexically similar to the target terms in the BMTO. When a correspondance was found, it was proposed as a possible translation of the English ICD-11 target term. The normalization process involved stripping genitive marks, transforming plural forms into singular, replacing punctuation, removing stop words, lower-casing each word, breaking a string into its constituent words, and sorting the words into their alphabetic order (see Figure 1). Examples of two ICD-11 term translations are listed in Table 1.

[1] Statistics calculated on preferred terms.

Table 1 - Examples of translation of two terms using lexical approach

ICD-11 term	English term (BMTO)	French term
Intracerebral hemorrhage	intracerebral hemorrhage (MeSH)	hémorragie cérébrale
Hypertensive encephalopathy	Hypertensive encephalopathy (MedDRA)	Encéphalopathie hypertensive

Figure 1 - Example of Normalization process for the ICD-11 term "Hereditary cerebral hemorrhage with amyloidosis, Dutch type"

Results & Discussion

According to our lexical approach, of the 336 ICD-11 terms selected, 164 (49%) terms were mapped from HeTOP and translated into at least one French term each. It is noteworthy that 194 (57%) terms were mapped to at least one English term. However, fifteen of these 164 translated ICD-11 terms (9%) were obtained exclusively from the translations performed by the CISMeF team [3].

Acknowledgement:

The authors are grateful to Nikki Sabourin-Gibbs, Rouen University Hospital, for editing assistance.

References

[1] http://www.who.int/classifications/icd/revision/en/ [accessed in November 2014].

[2] Grosjean J, Merabti T, Dahamna B, Kergouraly I, Thirion B, Soualmia L, Darmoni SJ. Health multi-terminology portal: a semantic added-value for patient safety. In: PSIP Workshop 2011; pp. 129-138.

[3] Merabti T, Soualmia LF, Grosjean J, Joubert M, Darmoni SJ. Aligning Biomedical Terminologies in French: Towards Semantic Interoperability in Medical Applications. In Book: Medical Informatics 2012; pp. 41-68.