

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

Tayeb Merabti<sup>a</sup>, Julien Grosjean<sup>a</sup>, Jean-Marie Rodrigues<sup>b,c</sup>, Stefan Jacques Darmoni<sup>a,b</sup>

<sup>a</sup> CISMef & TIBS, LITIS EA 4108, Rouen University Hospital, Rouen, France

<sup>b</sup> INSERM, U1142, LIMICS, F-75006, Paris, France

<sup>c</sup> Department of Public Health, Saint Etienne University Hospital, France

## Abstract

To translate the 11<sup>th</sup> edition of the International Classification of Diseases (ICD-11) into French, we proposed a lexical-based approach based on 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, algorithm translated 164 (49%) into at least one French term.

## Keywords:

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

## Introduction

The 11<sup>th</sup> 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 translated into French nor included in the Unified Medical Language System (UMLS). Thus, we have attempted to apply a lexical-based approach to translate a sample of ICD-11 terms into at least one French term.

## Materials and Methods

To translate ICD-11 terms, we propose a lexical-based approach based on 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, HeTOP includes 349,311 French terms<sup>1</sup> from BMTO included in UMLS, whereas the latter includes 164,071 French terms.

By this approach [3], all terms in English from all bilingual BMTO (English and French) included in HeTOP were normalized. An algorithm was developed to find the most lexically similar terms to target BMTO. When a correspondance was found, the translation of the English target term was proposed as one possible translation of the ICD-11 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.

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”

|                                 |   |
|---------------------------------|---|
| Remove genitives                | Hereditary cerebral hemorrhage with amyloidosis, Dutch type |
| Replace punctuation with spaces | Hereditary cerebral hemorrhage with amyloidosis Dutch type  |
| Remove Stop words               | Hereditary cerebral hemorrhage amyloidosis Dutch            |
| Lowercase                       | hereditary cerebral hemorrhage amyloidosis dutch type       |
| Uninflect each word             | hereditari cerebr hemorrhag amyloidosi dutch type           |
| Word order sort                 | amyloidosi;cerebr;dutch;hemorrhag;hereditari;type;          |

## Results & Discussion

According to our lexical-based approach, of the 336 ICD-11 terms selected, 164 (49%) terms were translated into at least one French term mapped from HeTOP. It is noteworthy that 194 (57%) terms were mapped to at least one English term. Fifteen of these 164 translated ICD-11 terms (9%) were obtained exclusively from added 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.

<sup>1</sup> Statistics calculated on preferred terms.