

## Semantic annotation of electronic health records in a multilingual environment

Luís Campos<sup>1</sup>, Vasco Pedro<sup>2</sup>, Francisco M. Couto<sup>1</sup>

<sup>1</sup>Faculty of Sciences, University of Lisbon

<sup>2</sup>Unbabel

Beyond the useful uses they already have, Radiology reports still have the potential of being an useful source of information for the improvement of health-related practices. They usually are better structured than other types of Electronic Health Records, making them more prone to be analyzed by Natural Language Processing (NLP) tools. But most of these tools assume that the reports were written in English, which is not always true. One obvious solution is to translate the text in the native language to English, before applying the NLP techniques. But what kind of translation should be used? Machine Translation (MT) is one option, being way cheaper than the other obvious alternative, Human Translation (HT), but has the low of having worse quality.

This work aims studying how MT compares with HT on the simple task of Named-entity recognition, using RadLex terms, on biomedical texts related to Radiology. The selected corpus for this work consists of research papers related to Radiology that are available both in Portuguese and English. The Portuguese versions of the papers were machine translated to English and compared with the original English translation.

Through our analysis we found that, as expected, the terms identified in texts resulting from MT are different than ones identified in texts resulting from HT, but surprisingly, only part of these differences are explained by bad translation from MT. They can also be explained, for example, by deficiencies on the RadLex terminology and on the system used to identify the terms. Correction of these deficiencies will improve the identification of terms in MT text and the analysis done during this work will be used as the basis for that.

**Preference for presentation:** Talk/Poster

**Location:** University of Minho

**Author for Correspondence:** [lcampos@lasige.di.fc.ul.pt](mailto:lcampos@lasige.di.fc.ul.pt)