

« Une zone inflammatoire *séquellaire* d'oesophagite peptique » : Functional analysis of the expression of cause/effect patterns in medical records

The objective of this research is to identify causal patterns specific to medical corpora, built from patient records written by doctors. We try to highlight the different functional roles of each pattern and to formalise the pattern according to these criteria.

This research is based on the idea that some external parameters, such as the type of the analysed text (Patient Discharge Summary, Radiology Reports, etc.) and the medical services from which the elements of the corpus come (such as MRI, gastroenterology, bariatric and abdominal surgery), have an influence on the formulation of the cause/effect relations [1, 2]. Based on this premise, we evaluate to what extent these principles affect the corpus. This anonymised corpus is made of 226106 medical texts (around 27500000 words) that were collected in six services of a Brussels hospital, partner in the iMediate project [3].

The conducted exploratory work that we conducted previously led to the discovery of four linguistic methods to express these relations: “classical” cause and consequence markers (e.g. “donc”, “parce que”), distorted other related markers (e.g. “mais”, “et”), specific lexicon (e.g. “se solder par”, “séquellaire”) and implicate with lexical signs (“in conclusion”, syntactic punctuants). After a study of their occurrences, we have discovered that each of these four devices adopts a function of their own that depends on the frequency and context of use. To analyse our corpus, we have constructed Unitex grammars (search patterns) [4] that extract occurrences of each method used to express causality.

On this poster, we will show the results of the qualitative and quantitative analysis that we led. We will distinguish different causal patterns that we found in the corpus and explain the differences amongst them and between our comprehension and interpretations of these methods. We will also present a typology of functions that we identified and draw the distribution of each pattern and types depending on hospital departments, for instance. These results will be backed up by statistics and occurrences in order for them to be compared with their hypothetical functions.

Eventually, we will compare these results with a field survey by conducting an experiment in which medical cases are submitted to ten physicians (five GPs and five specialists). This experiment will be conducted to understand how the physicians formalise causes and consequences in typical circumstances.

This is a cognitive approach of cause-consequence relations highlights the specific and intrinsic functions of the different types of expressions and emphasises the role of an accurate choice in the development and understanding of a speech [5].

Bibliography :

- [1] Khoo, C., Chan S. & Niu, Y. Extracting Causal Knowledge from a Medical Database Using Graphical Patterns. In : *Proceedings of ACL*, Hong Kong, 2000.
- [2] Serban, R., ten Teije, A., van Harmelen, F., Marcos, M. & Polo-Conde, C., Extracting and use of linguistic patterns for modelling medical guidelines. In : *Artificial intelligence in Medicine*, 2007 (39), 137-149.
- [3] *iMediate : Interoperability of Medical Data through Information extraction and Term Encoding*. Projet sous la direction de C. Fairon, 2014-2016.
- [4] <http://www-igm.univ-mlv.fr/~unitex/>
- [5] Keenan, J., Baillet, S. & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. In : *Journal of Verbal Learning and Verbal Behavior*, 1984 (23/2), 115–126.