

Information Structure Representation and Extraction from a corpus of Patient Data, using an Ontology.

Christian Cote
ERSICOM
Jean Moulin University
Lyon, France
cote@univ-lyon3.fr

Abstract

We propose a methodology to model the information structure for its extraction from any medical text. We experiment this extraction in a corpus that represents the information system of a specific professional activity in the hospital pharmacy.

The model of the information structure presupposes we know what are the constraints of the information system on the symbolic entities (in a way to distinguish the information structure to any sentence description). In a way to determine these constraints, we propose to model the information process by the information flow: we represent in this way how any fact in the body of the patient is symbolised, conveyed and represented into a text.

The information flow characterizes only the constraints of the information on the linguistic entities and structures. But the information is linguistically a referential semantic object: it's the representation at distance of a new fact in the world in the frame of a text that accepts this information. Then the model of the information structure represent how the meaning of a sentence is specified by the constraints of the information flow.

But the information structure is systematically recognized and interpreted in the context of a text: it's also the last object of the information system. Then we consider the text as a contextual frame to model the recognition and the extraction of the information structure.

A text can't be considered only as a linguistic object in a professional and information context: it's an implemented (or externalised following situated and distributed cognition) ontology. The updated text articulates the ontology of the patient body and the referential dimension of the information.

At last, The information flow allows the articulation of an ontology and a semantic precisely on the question of the information structure. We unify the model of the information structure by the definition of five primitives. A sign representation allows both the characterisation of the structure and of each of its components.