

Summary

In Linguistics, the semantic relations between words in a sentence are accounted for, *inter alia*, as the assignment of *thematic roles*, e.g. AGENT, INSTRUMENT, etc. As in predicate logic, simple linguistic expressions are decomposed into one predicate (often the verb) and its arguments. The predicate assigns thematic roles to the arguments, so that each sentence has a *thematic grid*, a structure with all thematic roles assigned by the predicate. In order to reveal the thematic grid of a semantically sound sentence, a system called HTRP (*Hybrid Thematic Role Processor*) is proposed, in which the connectionist architecture has, as input, a distributed representation of the words of a sentence, and, as output, its thematic grid. Both a random initial weight version (RIW) and a biased initial weight version (BIW) are proposed to account for systems *without* and *with* initial knowledge, respectively. In BIW, initial connection weights reflect symbolic rules for thematic roles. For both versions, after supervised training, a set of final symbolic rules is extracted, which is consistently correlated to linguistic – symbolic – knowledge. In the case of BIW, this amounts to a revision of the initial rules. In RIW, symbolic rules seem to be induced from the connectionist architecture and training. HTRP system learns how to recognize the correct thematic grid for semantically well-formed Portuguese sentences. Besides this, it leads us to take into account cognitive aspects of the linguistic processing, through the introduced (in RIW) and extracted (from both versions) symbolic rules.

Keywords: Natural Language Processing, Neural Networks, Artificial Intelligence.