

# Classification of noun phrases into concepts or individuals

**Saliha Azzam**

CRIL Ingénierie - CAMS - Paris-Sorbonne  
174, rue de la République 92817 Puteaux France  
e-mail : azzam@cril-ing.fr

## Abstract

We tackle here the problem of discrimination between instances of the language representation and concepts. This procedure is necessary according to the aim of the application that uses the conceptual structures. We propose linguistic rules for doing this discrimination inside natural language texts, and indicate how these rules are combined to build an accurate procedure.

## Introduction

The problem we address here deals with the automatic discrimination inside a natural language (NL) text between instances (i.e., individuals) and concepts. The procedure is integrated into a semantic parser and implemented in the context of COBALT project (CEC/LRE project). The terms that are substituted by concepts are noun phrases (NPs) and the problem is that of classifying the NPs as instances or concepts. The necessity of this classification (of NPs), depends on the aim of the application that uses the conceptual structures. It may be worth noticing that, even if this problem is very general and not restricted to the natural language (NL) processing domain, it is very difficult to find useful suggestions in the literature, at least in a knowledge engineering context. We will only mention here a well-known paper (Brachman *et al.* 1991) about the CLASSIC system, in which the differentiation about concepts and individuals is essential. In this paper, indications about the rules we suggest for a systematic classification are given (Azzam 1993).

## How rules are applied

The rules are ordered and exclusive, i.e., given a rule numbered *i*: if condition *i* then <classification-*i*>, means if (not condition *i*-1 and condition *i*) then <classification-*i*>. If a given rule succeeds on a NP, the latter is suppressed from the list of NPs to be classified and the next NP is considered. The algorithm has an empirical nature and some rules are incremental, i.e., as soon as a given NP is classified, the result is taken into account for the next NP. The procedure is

then reapplied at the end of a rules session, on the unclassified NPs and stops when no rule can be applied.

## The classification rules

The main features taken into account and involved in the rule conditions, are: 1) **The NP type**, e.g. if it corresponds to a proper noun it is an instance. In order to find the location of the concept which subsumes this instance, semantic and syntactic patterns are checked out 2) **The type of the recognized concept**: for example the concepts of 'abstract qualities' sub-hierarchy do not have any instances 3) **The syntactic structure of the NP** 4) **The determiner type of the NP** (possessive, demonstrative, ...) 5) **The mark of punctual situation** in the sentence, as the temporal adverbs that favor instances 6) **The expressions introducing general context** that favor concepts, e.g., "in general" or conditional assertions like "in default of" 7) **The semantic category of the verb** 8) **The syntactic role of other classified NPs**.

## Conclusion

The rules use linguistics knowledge, without involving any additional world knowledge. There are expressed under the form of syntactic and semantic constraints which are tested on the syntactic tree and using the associative knowledge of the representation language. Applied on COBALT corpus extracted from Reuters news, the rules succeed in 90 percent of cases, i.e., rules classify 90 percent of NPs in a correct way. The cases of failure are cases of unclassified NPs and not incorrect classification". Therefore, future works address the extension of rules to process more cases and also, the rules refinement to avoid any hazardous results.

## References

- Azzam, S. 1993. Classification of NPs. Technical report, COBALT report ST/06/93 STEP:Paris.
- Brachman, R.; McGuinness, D.; Patel-Schneider; P.F., R. L.; and Borgida. 1991. Living with classic : When and how to use a kl-one-like language. In Sowa, J., ed., *Principles of Semantic Networks*. San Mateo (CA): Morgan Kaufmann.