

A THEORY OF SEMANTIC RELATIONS FOR LARGE SCALE NATURAL LANGUAGE PROCESSING

Hanne Ruus
University of Copenhagen
Institut for nordisk filologi
and Eurotra-DK
Njalsgade 80
DK-2300 Copenhagen S

Ebbe Spang-Hanssen
University of Copenhagen
Romansk institut
and Eurotra-DK
Njalsgade 80
DK-2300 Copenhagen S

Even a superficial meaning representation of a text requires a system of semantic labels that characterize the relations between the predicates in the text and their arguments. The semantic interpretation of syntactic subjects and objects, of prepositions and subordinate conjunctions has been treated in numerous books and papers with titles including words like DEEP CASES, CASE ROLES, SEMANTIC ROLES and SEMANTIC RELATIONS.

However, no generally accepted set of case roles has emerged from all these investigations and discussions. A comprehensive exposition is given by H.L. Somers (1983).

We see one basic problem with most systems of semantic cases: the labels are not clearly differentiated. Consequently it is not possible to use the labels in a consistent way. We think that the missing differentiation has one principal source: many different aspects of meaning related to predicates and their arguments are treated by the same description: THE ASSIGNMENT OF CASE ROLES.

We propose that at least the following aspects are kept apart:

- feature descriptions of arguments: arguments are classified according to + - concrete, + - animate, + - stuff etc.
- temporal and aspectual features of predicates: predicates, notably verbs and adjectives, are classified according to + - durative, + - action, + - process, + - stative.
- the semantic relations established by the predicates.

In this paper we concentrate on the semantic relations established by predicates: what are they, what are their characteristics, how do they group the predicates.

One example will suffice to show how conflicting criteria make case label assignment impossible:

Suppose we use some wellknown labels to describe the following sentence:

Peter gave John the letter
AGENT RECIPIENT PATIENT

Yet with that kind of labels we get into trouble as soon as we want to describe such a simple sentence as

Peter took the letter from John
?AGENT PATIENT ORIGIN
?RECIPIENT

Obviously, there is a conflict between two criteria: causality: (AGENT PATIENT) and location: (RECIPIENT (GOAL) ORIGIN)

To which of the two criteria should we give the priority? It is not so that the causer (agent) is always origin of some object trajectory as with 'to give'. As a matter of fact, the predicate 'to give' has always been used as a model (cf. the word 'dative'). The causer may well cause that objects move not away from him but towards him, as with 'to take'. In general, we could speak of two different streams or directions, one at the spatial level (the location of objects), another at the causal level (an influence going from one entity towards another entity).

To ensure the self-consistency of a system we appeal to the following two principles:

- we use the mathematical notion of relation: a predicate establishes a relation between pairs of units or one unit is related to itself by a predicate,
- we use few clearly distinguished semantic units.

Some units are relators. They establish relations between other units. The relators are the predicates, the other units are their arguments. When we try to characterize the different kinds of relations, it is important to keep in mind that the arguments are related BY the predicate, not TO it.

We base our system of semantic relations on one semantic unit DIRECTEDNESS. The directedness concerns some object that may be oriented TO and/or FROM some other argument. The directedness may be situated at four semantically different, abstract levels: an influential, a locational, a qualificational

and a pertinential one. At three of these levels there is an additional non-directed relation. In the case of non-directedness the object is related to another argument that is semantically marked in the relevant dimension.

2. PRESENTATION OF THE SYSTEM

2.1. SURVEY

The system we propose looks like this:

influential: object orig.
 qualificational: object orig. goal ref-term
 locational: object orig. goal ref-term
 pertinential: object orig. goal ref-term

The arguments of one predicate can have labels belonging to at most two dimensions: influential and one more dimension. Examples:

 He brought us a new car
 influential: orig. object
 locational: goal object

 The car came from Paris
 locational: object orig.

In qualificational, locational and pertinential, if the dimension is present, there must be at least one argument that is 'object' and at least one argument that has one of the three other values: origin, goal or ref-term. The dimension has no meaning unless there is an 'object' related to at least one of the three reference points. In each of these three dimensions an object may be situated on a directed line, i.e. with respect to an origin or to a goal or to both. Or it may be situated, without any directional idea, with respect to a reference point that is neither origin nor goal, and which we call just ref-term.

 She lives in Paris
 locational: object ref-term

Ref-term is negatively defined as a reference point which is neither origin nor goal. It may be characterized positively, however, since a reference point, if it is neither origin nor goal, will normally be some noticeable object in proximity. This characteristic is important when we try to apply the terms origin, goal and ref-term in the non-locational dimensions.

In influential, the argument that is origin can also be an implicit object, i.e. the influential relation may be reflexive: one argument is related to itself. Examples:

 She turned round
 influential: orig.

 She turned the page
 influential: orig. object

2.2. THE DIMENSIONS

We use the dimension 'influential' to express an influence going from an origin to an object, as in the example just quoted. The concept of 'origin' at the influential level is close to the traditional 'agent', yet somewhat wider. It is not, as the latter, linked to the concept of action, but rather to that of causality. In every sentence there is an 'object' to which something happens. In a much smaller number of sentences there is an indication of what or who makes the thing happen. We need not always, however, fill out 'object' of influential, since it will be the same as 'object' in whatever other dimension is present.

We use 'qualificational' to express relations between arguments some of which are not referential objects:

 He is a teacher
 qualificational: object ref-term

 She became president
 qualificational: object goal

The arguments that are qualifications are easy to distinguish from other arguments by the fact that they cannot be referred to with a pronoun. They are not independent discourse entities, but they are attributes of some discourse entity.

We use 'locational', as already illustrated, to express spatial relationships, and also obvious metaphorical uses of space.

Finally, we use the dimension 'pertinential' to express more abstract relations. We take the Latin word 'pertinens' in the meaning 'concerning, referring to'. Thus the constituting idea of this level is just the idea of referring or relating an object to some other object. Defined in this way, pertinential can serve us as the unmarked dimension, i.e. the dimension we can use whenever the more narrowly defined dimensions do not suffice to describe the relations between the arguments of some predicate. In some cases a pertinential relation may be directional, namely when there is a precedence between the arguments:

 Peter suffers from pneumonia
 pertinential: object orig.

 The gun is used for shooting
 pertinential: object goal

