Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

PROCEEDINGS OF THE 6th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION

Colette Faucher, Marielle Gonzalez-Gomez, Eugène Chouraqui

DIAM - IUSPIM
Université d'Aix-Marseille III
Domaine universitaire de Saint-Jérôme
Avenue Escadrille Normandie-Niemen
F-13397 Marseille Cedex 20 - France
e-mail : cfaucher@dialup.francenet.fr

1) OBJECTIVES

The topic of this paper is classification reasoning in frame-based representations. In these representations, an object designates either a concept (which is represented by a *class*) or a concrete entity which illustrates one or more classes (represented by an *instance*). Our objective is to define an instance classification method, which gives a more important place to the semantics of the Object To Classify (OTC) than other works in this area. We have studied cognitive psychology works concerning categorization. These works are focused on the identification of the elements which are concerned with the categorization process. We have taken into account certain results of cognitive psychology to define our classification system.

The scheme of this paper is the following : in the section 2, the general problem of classification in the frame-based representations is presented. Then, in section 3, critical review of these representations is proposed. The section 4 presents the main results of the cognitive psychology concerning categorization and the section 5 details the more important points of our classification system. Then, the general approach of our system is given (section 6). Finally, we give the elements of knowledge participating into the classification process (section 7) and the matching step is described, specially important in the defined method (section 8).

2) FRAME-BASED REPRESENTATIONS AND CLASSIFICATION

2.1) Frame-based representations

We focus on frame-based languages which are knowledge representation languages inspired by Minsky's works (Minsky 75). An object is named *frame* in these languages. A frame, class or instance, is a data structure composed of *slots* (which represent the properties of the described object). Slots which introduce simple properties or relations are themselves described by means of *facets*. Facets contribute to the slot description. Facets are divided into two sets: *descriptive facets* (e.g. 'domain', 'value', 'default') and *reflexes* (procedural components of the description, e.g. 'if-needed', 'if-created') which are triggered after the manipulation of the slot value. Slots are characterized either by values (introduced by the *value* facet) or by default values (introduced by the *default* facet). Specifying a default within a class slot means that this default value is generally true for this slot, for the sub-classes or for the instances of the class where the default is declared. Nevertheless, a default can be overridden by an exception. The value within a class slot means that this value is true for this slot, for all the sub-classes or for the instances of the class where the value is declared.

Classes are organized in a hierarchy. Relations, named *links*, connect the frames of the hierarchy. In particular, the link *kind-of* connects the classes in the hierarchy, the link *is-a* allows the connection of an instance to the class it corresponds to. This leads us to specify

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

notions of *inheritance* and *specialization*. The knowledge sharing is accomplished by inheritance properties. A class connected by the link *kind-of* to another class (named super-class) inherits properties from this super-class. A class can specialize its super-class if the specializing class description verifies coherence constraints concerning the partial relation order between these classes. More precisely, these coherence constraints must verify if specializing class description is conformable to the specialized class description.

We have defined semantic classification of instances in the Objog+ language (Faucher 91). Objog+ is a frame-based language. Objog+ has the singularity to be extensible and self-referent. Self-reference implies that all the basic entities of the language (slots, facets, links, ...) are described by means of classes. Self-reference permits reflexivity, i.e. the language is able to adapt its behavior as regards its description. Reflexivity is essentially used when new facets are created. Indeed, the system automatically manages the control structure, i.e. modalities of use of a facet *a posteriori* described. The property of extendibility permits the enrichment of the language expressiveness by defining new elements of the language, which makes it adaptable to a particular domain of knowledge.

2.2) Classification in frame-based representations

The usual mechanism of reasoning in these representations is classification reasoning. Indeed, the knowledge base is organized into a hierarchy of frames based upon a partial relation order, the *specialization relation*. The goal of classification of classes or instances is to find the more appropriate localization of the object to be inserted into the knowledge base.

• Classification of classes

The classification of a class C (Finin and Silverman 86) consists in finding the set of classes which are more general than C, noted sup(C), and the set of classes which are more specific than C, noted inf(C), as :

$$\sup(C) = \{C' \in \mathfrak{C}, C \leq C'\} \quad \text{and} \quad \inf(C) = \{C'' \in \mathfrak{C}, C'' \leq C\}$$

where $\mathfrak{C}$ is the set of classes of the knowledge base and $\leq$ is the partial relation order upon $\mathfrak{C}$, i.e. the specialization relation. This relation implies inclusion constraints between domains of possible values of slots of two related classes.

• Classification of instances

Our work is concerned with the problem of instance classification which consists in finding the most specific membership class for which the OTC satisfies all the constraints. These constraints concern membership constraint between the domain of possible values of a slot within a class and the value(s) of this slot within the OTC.

The nature of slots describing classes is an important point in most classification systems in frame-based representations. Are they slots expressing necessary conditions, sufficient conditions or necessary and sufficient conditions to be an instance of a class? According to the interpretation given to the slots describing a class, the classification process will be more or less efficient. The definitions for each category of slots are the following (Fikes and Kehler 85):

- When a class is described by necessary conditions (NC), the NC only allow the conclusion that an object (instance) is not a member of a class, but they do not allow a conclusion about the membership of an object to be in a class. A system in which there are only NC does not realize the attachment of an object to a class automatically. The attachment must be explicitly given by the user.
- When a class is described by sufficient conditions (SC), the satisfaction of SC by an OTC permits a conclusion about an object's membership in a class. On the contrary, when SC are not satisfied, the system can not conclude that the object is not a member of the class.
- When a class is described by necessary and sufficient conditions (NSC), the satisfaction of NSC permits a conclusion about an object's membership in a class, and when the NSC are not satisfied, the considered class is rejected.

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

3) CRITICAL REVIEW

Most classification systems utilize necessary and sufficient conditions or, at least, necessary conditions to describe a class. In these representations, a class is the description of a concept. However, for many years cognitive psychology researchers have underlined the difficulty of specifying a concept with necessary and sufficient conditions (Murphy and Medin 85).

Another point is that, in these systems, the notion of typicality is completely forgotten. As we will present it, typicality is an important notion in the categorization[1] process.

A lot of studies in cognitive psychology try to know how two things are considered as similar. The determination of similarity is a step in the categorization process. Matching is also a step in a classification system. Nevertheless, in most classification systems matching is syntactical and concerns type constraints. Generally, the classification process is not understood as a cognitive activity. Our approach is based on the similarity considerations in the matching step. A similarity determination is possible thanks to a flexible framework representation.

In the following section, we discuss some works about cognitive psychology which are interested in the classification task. We present the most important points. All these points have been used to define our classification system.

4) COGNITIVE PSYCHOLOGY CONTRIBUTIONS

4.1) Categorization in cognitive psychology

Some works of cognitive psychology study the categorization process and try to answer the following question : *how can we say that X, which is an entity (an object of the world), is a member of the category Y?* This question is fundamental when a classification system has to be defined.

To answer the previous question, researchers in cognitive psychology try to understand the notion of concept. The second question they ask is *what is the structure of a concept?* There are three acceptions of the notion of concept that are more or less accepted in the psychological community. These three acceptions imply three different representations; the classical view, the probabilistic or prototypical view and, finally, the exemplar view.

In the classical view, a concept is defined by a set of necessary and sufficient conditions. A concept is represented by a category and all the elements of the category share common properties. The membership is defined by the "all or none" rule. The limits of this view have been underlined by numerous works (Rosch and Mervis 75), (Cohen and Murphy 84), (Murphy 93).

The probabilistic view (or prototypical view) (Rosch and Mervis 75) refuses to see a concept defined by necessary and sufficient conditions. In this view, a concept is described by an entity which is representative or typical. The membership in a category is graded, the membership level depending on the typicality evaluation between the representative entity of the category, named the prototype, and the considered entity.

Finally, the exemplar view does not see a concept defined by necessary and sufficient conditions. In this view, categories are defined by their members. The membership of an entity in a category depends on the distance evaluation between this entity and one or more members

[1] The term "categorization" is used here in the sense of Rosch (Rosch and Mervis 75). Categorization is a process which consists in putting an entity into a category. In this sense, the term categorization is equivalent to the term classification, which is more frequently used in artificial intelligence. We will use the term "category" in the following. A category designates the set of the entities concerned with a concept (in other words, a category is the extension of a concept). A concept is the representation, by intension, of a category (i.e. a concept is an abstraction).

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

of the category (Medin and Heit 94).

It is obvious that the concept structure constrains models of category processing. Among the previous conceptual structures, the classical view is rejected, except to handle mathematical concepts. Although there are criticisms to prototypical view, we think that it is a very interesting way. So, we describe it with more details in the following. The exemplar view proposes a point of view very innovating in the artificial intelligence domain, but we do not develop this.

4.2) The prototypical model

In the 70's, Rosch proposed an alternative to the classical model, because of the difficulties in identifying NSC of a category (Rosch and Mervis 75). A category is represented by a prototype, i.e. a representative member of the category. The prototype can be an entity (e.g. we consider that the most typical fruit is "apple") or it can be an "idealization" built with typical properties that can be identified in different members of the category (e.g., we consider that fruit is described by the following properties: sweet taste, orange color, to have a peel, to have seeds, etc.).

The membership of a new entity in a category is based on the similarity shared by the entity and the prototype of the category. The entity will be considered typical for a category if it possesses numerous common properties with the prototype and if it shares fewer common properties with the prototypes of other categories. Rosch defines a theory, named "family resemblance", to evaluate the membership relation between an entity and a category. In this approach, some members are more typical members than others, that contrasts with classical view where all members of a category are equally representative or typical.

Rosch's model asserts that the categorization process is based on the typicality evaluation. One major characteristic of the model is the consideration of the properties describing a concept in an additive and independent manner. That is to say, the global typicality of an entity is the sum of the elementary typicalities measured at the property level.

Malt and Smith (84) criticize this typicality approach. For the authors, typicality is highly influenced by the presence of correlated properties. The notion of correlated properties is defined as follows: two (or more) properties are correlated if the presence of one predicts the presence of the other, and the absence of one predicts the absence of the other, or alternatively, if the presence of one predicts the absence of the other and vice versa.

To illustrate this definition, suppose that small birds sing and large birds do not. Suppose further that the properties (height, small) and (expression mode, sings) are more typical than the properties (height, large) and (expression mode, squawks). A small, singing bird will therefore be more typical than a large, squawking bird. Consider, now, a large bird that sings. It is more typical when we do not take into account the correlations than is a large, squawking bird, although it is precisely an atypical bird. By consequence, when correlated properties are not considered, important variations modify the typicality of a category member because there are equalizing phenomena between elementary typicalities.

Hampton (93) suggests that if the prototypical model was able to consider correlated properties, it would be a good representation model. In the prototypical view, an entity is typical for a category if it is similar to the prototype, the representative of the category. So we are led to study the similarity judgment.

4.3) The similarity judgment

Tversky's works (particularly Tversky 77), are important when we focus on the similarity judgment. Tversky defined the "contrast model" in which similarity is a weighted function of common properties and distinctive properties of the two considered entities (Tversky and Gati 78), (Gati and Tversky 84).

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

An entity a is characterized by a set of properties. The similarity between a and b is a function based on the common and the distinctive properties of the entities. Three elements are involved in the similarity judgment :

- the common properties of a and b,
- the properties of a, which are not the properties of b and
- the properties of b, which are not the properties of a.

Then, similarity is a linear combination of measures evaluating the common properties and the distinctive properties.

Tversky underlines that common properties in the similarity judgment are always privileged relative to the distinctive properties, and more specially in the categorization task (Gati and Tversky 84).

In the similarity evaluation, context is an element which is fundamental. To take the context into account in the similarity evaluation implies that two entities are not always similar, but two entities are similar relative to a given context. Suppose that, for instance, we want to compare France and Switzerland. Among a set of African countries, France and Switzerland are more similar than among a set of European countries. Other authors have also insisted upon the need of taking into account the context in the similarity judgment (Murphy and Medin 85), (Barsalou, 89), (Medin and Goldstonne 93), (Medin and Heit 94).

4.4) Similarity and categorization

Rosch's model establishes a direct relation between similarity judgment and the categorization process. Nevertheless, some authors disagree with the importance associated with the role of similarity in the categorization process.

Objections to Rosch's model led Murphy and Medin (85) to define another kind of constraint associated with the notion of similarity in the categorization process. It deals with the "theory of concept". The idea is that background knowledge can be used to explain why an entity is a member of a category. This "theory" occurs in the description of the concept. The theory of a concept can concern people's goals, needs and interests to categorize an entity. The following example, from Murphy and Medin (85) provides a good example of this notion. "The category *apple-or-prime-number*, does not appear to be a very current concept [...]. One could develop a scenario, however, in which this category might make sense. For example, suppose one of our colleagues in the math department has only two interests: prime numbers and apple farming. We might, then, form the concept *apple-or-prime number*, which is explained as the topics of conversation of our mathematician colleagues". The theory provides an explaining framework which gives coherence to the concept. It is based on the intentions and the goals of the categorization. It deals with contextual elements which give information about the knowledge area in which the categorization process occurs. The notion of "theory of concept" concerns, for instance, background knowledge.

5) OUR SUGGESTIONS

An important number of elements presented in the previous section are utilized for defining our instance classification system. In this section, the elements of cognitive psychology upon which our work is based are specified. Besides, we defined in Objlog+ a framework to represent approximate knowledge, i.e. fuzzy, uncertain and default knowledge. Approximate matching mechanisms have been defined to handle approximate knowledge.

5.1) Elements from cognitive psychology

- **Considering typicality**

Typicality takes an important place in the classification process. The defined model allows us to

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

take into account two elements: elementary typicality of a property and typicality of correlated property[2]. Taking into account co-occurrences in our model allows us to bypass the limitations of the prototypical model in which expression of correlations is not possible.

- **Importance degree associated with a slot**

The slots in the defined model are characterized by an importance degree which quantifies the weight of the slot within the global description of an object. The weighting associated with a slot is a means to express the importance of the slot in the global description. The weighings are not contextual. These weightings are used by matching, giving more importance to the properties which are the most meaningful.

- **Defining a partial matching**

The matching process we defined is the key point of the classification. First, it is a partial matching which provides graded membership of an instance in a class. In other words, we are able to say that an instance is more or less representative than another one for a given class.

Matching can be characterized by the following points:

- It is based upon resemblance measures (the compatibility measure and the typicality measure) and a dissimilarity measure (the incompatibility measure).
- An important place is given to the resemblance measures because, according to the results obtained by Tversky, the search for similarity is more important in the categorization process than the identification of dissimilarities.
- Finally, the compatibility measure we define is not a symmetrical measure. In our approach, the comparison is oriented, the class being the "model" and the OTC being the "data".

- **Considering background knowledge**

Another important point is the consideration of background knowledge in the classification process. Domain theory occurs during the matching between the OTC and a class of the hierarchy. It concerns, first, the specification of equivalences between valued slots that we will present later and, secondly, the specification of correlations between properties.

Taking into account all these elements necessitates the utilization of an expressive support language. An expressive model of representation is the necessary condition for defining a "powerful" matching which integrates background knowledge (Medin and Ortony 89). The Objog+ language which is an extensible frame-based language gives a conceptual model sufficiently powerful to express and handle all the previous elements,

5.2) Representing and handling approximate knowledge

We defined a model for representation of approximate knowledge in the support language Objog+ (Faucher, Chouraqui, Gonzalez 95), (Gonzalez, Faucher, Chouraqui 95a), (Gonzalez, Faucher, Chouraqui 95a). In particular, in our model the default values are characterized by a linguistic quantifier (Yager 94) which expresses the scope (the representativeness) of the default. This quantifier permits us to know the typicality (or the atypicality, or the exceptional character) of the default for the considered slot. The notion of *fuzzy term* is also defined. A fuzzy term introduces, within a slot, a vague data (Dubois and Prade 88), (Zadeh 85). More precisely, a fuzzy term designates either a continuous fuzzy subset (*continuous fuzzy term*) or a discrete fuzzy subset, this subset containing only one element (*discrete fuzzy term*). This element can be a partial membet to the subset. It is associated with a membership degreecan be a partial membet to the subset. It is associated with a membership degree. We designate by *precise term* a numerical or symbolical value which is well known. The introduction of fuzzy terms within the slot description is not detailed, but new facets within the language are defined.

---

[2] A property is seen as a pair (attribute, value).

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

6) GLOBAL PRESENTATION OF OUR CLASSIFICATION METHOD

The main characteristics of the proposed model are (i) the definition of an approximate pattern matching, including adapted pattern matching methods to handle approximate knowledge and semantic pattern matching method; (ii) the definition of heuristics to perform classifying and to improve the pattern matching by completing the OTC description through the use of additive knowledge (background knowledge); (iii) the consideration of default values describing slots of the class C, which express typicality for the comparison between OTC and C. Taking into account default values in the matching is singular because most works of classification, in frame-based representations, do not distinguish their particular nature.

Our classification method is composed of three steps. The first one tries to perform the classifying of OTC by researching properties within OTC which are identified as sufficient properties to be attached to a class of the hierarchy. That is to say, the presence of such properties within the OTC description is sufficient to make it a member of a given class. This heuristic approach, whose goal is to perform classification process, does not success every time. In this case, usual search within the hierarchy and matching methods are used.

The second step realizes the matching between OTC and a class C. Before comparing OTC and C, the system tries to complete the OTC description using knowledge about the concept represented by C, knowledge informing about the existence of equivalences between some properties of OTC and C.

Matching provides three measures to evaluate more or less similarity between OTC and C: (i) a measure of compatibility taking into account the common properties of OTC and C; (ii) a measure of incompatibility, considering the properties of C which distinguish it from OTC (properties only existing within the description of C); and (iii) a measure of typicality computed from the common slots of C and OTC, and which are described in C by default values.

Based upon the two first measures, classes are identified as solutions for the OTC classification or are rejected. The last measure permits the method to choose between two solution classes that are judjed equivalent based upon the two previous measures. These classes are related to each other by a specialization relation, and the chosen class constitutes the best solution for the OTC. In the following (section 8), we will detail the second step of the classification process.

At the end of the search in the hierarchy, OTC can be linked to one or more classes (third step, named OTC classification), i.e. our algorithm is able to generate multi-instantiation for OTC.

Finally, inferences are realized after the classification. The values or default values of properties of classes, solutions for OTC, which are not initial OTC properties are inferred into the OTC description. Computed degrees of uncertainty are associated with all the values propagated into OTC to underline the hypothetical character of these inferences.

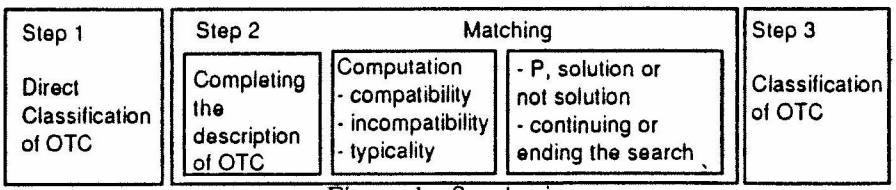

Figure 1 - Synthesis

7) DEFINITION AND REPRESENTATION OF KNOWLEDGE ELEMENTS OCCURRING IN THE CLASSIFICATION PROCESS

In most frame-based languages, a concept is represented by a class. This class possesses slots whose expressive and operational function is related to the modeled universe of discourse. We introduce "additive" knowledge elements whose operational function is independent of the universe of discourse. These elements of knowledge are used in the classification process.

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

They are represented by pre-defined slots. These slots are defined in the frame PROTOTYPE, which is the root of the hierarchy of classes in Objlog+. They are present in all the classes which represent concepts. Some of these elements are detailed in the following sections.

7.1) Equivalence between valued slots

In some cases, the matching between a valued slot noted (sl,v) of a class C and a valued slot (sl',v') of OTC leads to a failure of matching because of the different slot identifiers. Nevertheless, the meaning of the two valued slots can be close. The description of equivalences between valued slots renders the matching more flexible. Porter, Bareiss, Holte (90) designates these elements of knowledge as "matching knowledge", which underlines the operational dimension of this information.

- **Definition**

Matching knowledge specifies relations between valued slots, for a given concept, defining an equivalence between these slots. Two valued slots are said to be equivalent from the point of view of the classification if the presence of one or the other slot in the OTC description (all other slots being identical) suggests the same classification for OTC. We note:

$$[\ (\text{(sl, v), OTC}) \Leftrightarrow_{CL} (\text{(sl', v'), OTC})\ ] \text{ iff } \{\ [\ \text{(OTC, D)} \rightarrow \text{is-a(OTC, C)}\ ] \Leftrightarrow [\ \text{(OTC, D')} \rightarrow \text{is-a(OTC, C)}\ ]\ \}$$

where $D=\{(sl_i,v_i)\}$[3] is the description of OTC and D'= transf(D, (sl, v), (sl', v')) is another description of OTC obtained by transforming D by replacing (sl,v) by (sl',v').

- **Example**

We consider the class C, Safari-Animal, described by the valued slot (class-animal, mammal) where the OTC is described by the valued slot (possesses-teats, yes). The specification of an equivalence, from the classification point of view, between these valued slots (i) permits enrichment of the OTC description by adding the pair (class-animal, mammal) in its description and (ii) permits matching between the C and the OTC descriptions relative to these slots. Concretely, we can notice that the defined equivalence between two pairs of valued slots build a link between a slot expressing an abstract property and a slot expressing perceptible property. Generally, the abstract property appears in the class and the perceptible property occurs in the OTC[4]. These equivalences are given by the expert of the studied domain.

- **Representation**

We define the slot <equivalent-slots> in the PROTOTYPE frame. This slot possesses for value a set composed of two pairs (slot,value(s)). Its cardinality (number of values) is between 0 and infinity. The slot values can be inherited, i.e. the equivalences defined in the class C are true for a class C' which is related to C by the specialization relation. In the previous example, we have:

Safari-Animal

<kind-of>	value : {animal}
<class-animal>	value : {mammal}
<equivalent-slots>	value : {((class-animal, mammal), (possesses-teats, yes))}

7.2) Co-occurrence of slots

Cognitive psychology have shown that typicality plays an important part in the evaluation of similarity (see section 4.2). In the model, it is possible to specify co-occurrences of slots.

---

[3] v_i is either a value, or a set of values, or the empty set.

[4] We must note that the OTC description is given by a non-specialist, so it is mainly specified by perceptible properties.

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

• **Definition**

We take once again the definition presented by Malt and Smith. Two slots described by default values are correlated if the presence (or the absence) of one predicts the presence (or the absence) of the other. We note:

(sl,v) is correlated to (sl',v'), within the class C, if and only if there exists q_c (quantification of the frequency of the co-occurrence) so as $((sl, v) \wedge (sl', v'), q_c)$, where q_c is a linguistic quantifier, for example frequently, rarely, ... According to the values of q_c, we deal with strong or weak correlation.

• **Representation**

We define the slot <correlation> in the PROTOTYPE frame. This multi-valued slot is valued by a set of couples, whose first element is a set of pairs (slot, default(s)) and the second is a linguistic quantifier. Its values cannot be inherited, i.e. the correlations defined within a particular class are private (for this class). We take once again the example given in section 4.2,

Bird

<kind-of>	value : {animal}
<height>	quantified-default : {(small, most)}
<expression-mode>	quantified-default : {(sings, most)}
<correlation>	value : {(((height, small), (expression-mode, sings)), frequently), (((height, large), (expression-mode, sings)), rarely) }

7.3) Definition of taxonomy of slots and approximate matching mechanisms

• **Taxonomy of slots**

We define a taxonomy of slots (figure 2) whose goal is operational. In the taxonomy, the most appropriate matching method is associated with each kind of slot in order to bypass syntactical matching limits. When syntactical matching fails, the system utilizes approximate matching.

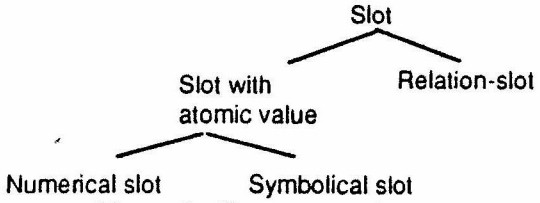

Figure 2 - Taxonomy of slots

According to the kind of slot which is considered and the nature of the values of this slot (fuzzy or precise terms), we will realize:

- a fuzzy matching (Dubois et al. 88),
- a semantic matching between numerical terms,
- a semantic matching between symbolical terms (string),
- a matching upon relation-slots (which is not detailed here).

• **Fuzzy matching**

Fuzzy matching consists in computing either (i) an inclusion degree between two continuous fuzzy terms (the inclusion degree corresponding to a measure of necessity); or (ii) a degree of membership between a precise term and a continuous fuzzy term. Discrete fuzzy terms are handled by semantic matching between symbolical terms.

Let t be a precise term and ft1 and ft2, two continuous fuzzy terms. The inclusion degree, ft1 given in the slot sl of the class C, and ft2 given in the slot sl of the OTC, is computed according to formula (1). The membership degree between ft1, given in the slot sl of C, and t, given in

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

the slot sl of the OTC, is computed according to formula (2):

$$d_{\subset}(ft2, ft1) = N(ft1 / ft2) = Inf_{x \in \Omega} \max(\mu_{ft1}(x), 1 - \mu_{ft2}(x)) \quad (1)$$

$$d_{\in}(t, ft1) = \mu_{ft1}(t) \quad (2)$$

- **Semantic matching between numerical terms**

Semantic matching between numerical terms evaluates approximate equality between two numbers by computing the intersection degree (measure of possibility) between the fuzzy numbers defined from the initial numerical terms (figure 3). The definition of a fuzzy number from a precise number necessitates providing an order of magnitude (noted Ô).

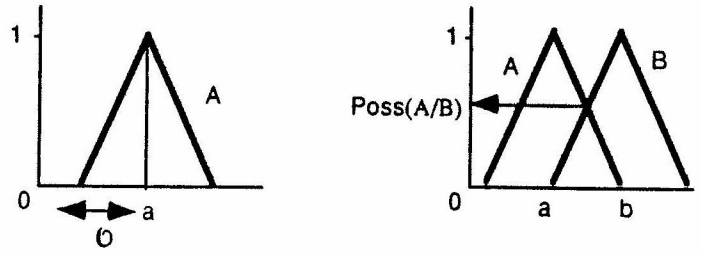

Figure 3 - Approximate equality between numerical numbers

- **Semantic matching between symbolical terms**

Semantic matching between symbolical terms evaluates the approximate equality between two symbolical terms. A method is proposed which searches for the existence of "hyperonymical" and "meronymical" relations between two terms. We define these relations by: (i) two terms are hyperonymically related if one of the terms is the hyperonyme (i.e. ascendant) of the other; (ii) two terms are meronymically related if there exists an additional term (i.e. ancestor) which is the hyperonyme of the two considered terms.

The search for hyperonymical and meronymical relations is possible because of the arborescent organization of terms characterizing the domain of possible values of the considered slot (figure 4). The approximate equality between two terms is not symmetrical (Tversky 77). If the hyperonyme term occurs in the slot of the class or in the slot of the OTC the approximate equality is different. Our goal is to compare the values v_{otc} and v_c, respectively, in the slot sl of OTC and C. Three cases are possible:

(i) v_c is an hyperonyme of v_{otc},
(ii) v_{otc} is an hyperonyme of v_c,
(iii) v_c and v_{otc} are meronymically related.

In the case (i), the approximate equality is 1.
In the case (ii), the approximate equality is (1/ (nb + (1 - 1/r))), where nb is the number of terms which have the same depth as v_c for which v_{otc} is an hyperonyme and r is the relative difference between the respective depths of v_c and v_{otc}.
In the case (iii), the approximate equality is (1 - 1/n), where n is a depth of the first common ancestor of v_{otc} and v_c. Note that the root depth of the arborescence is 1.

Let hyp(x, y) be a predicate, true if x is an hyperonyme of y. We obtain the approximate equality:

$$d_{\simeq}(v_{otc}, v_c) = hyp(v_c, v_{otc}) + hyp(v_{otc}, v_c) \times (1/(nb + (1-1/r))) + (1 - \max(hyp(v_c, v_{otc}), hyp(v_{otc}, v_c)) \times (1 - 1/n)$$

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

For example, for the Color slot described by the domain of enumerated values in the arborescence, "White" and "Whitish" are hyperonymically related (approximate equality is 1 if "White" is in OTC and "Whitish" is in C; in the opposite case, it is equal to 1/3. We obtain $d_{\simeq}$(White, Whitish) =1/3 because of the following elements : hyp(White, Whitish)=1 and nb = 3 and r = 1) and "Grey" and "Green" are meronymically related (approximate equality is $d_{\simeq}$(Grey, Green) =1/2 because of the following elements: hyp(Grey, Green)=0 and hyp(Green,Grey)=0 and n=2).

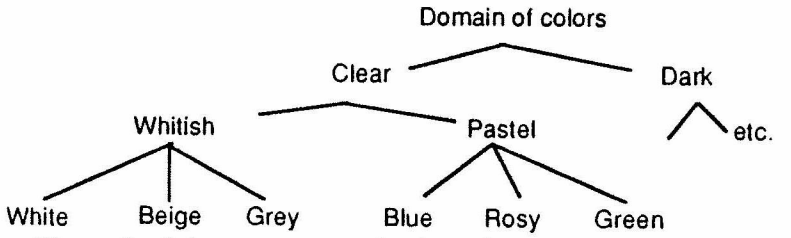

Figure 4 - Arborescent organization of the Color slot values

- **Representation**

The taxonomy of slots is defined in the language. In Objlog+, the slots are represented by instances of the class SLOT (self-reference). The following sub-classes SLOT WITH ATOMIC VALUE, RELATION-SLOT, NUMERICAL SLOT, and SYMBOLICAL SLOT, are defined. Each new sub-class possesses the similarity method which defines appropriate matching process. Any particular slot which occurs in a concept description is automatically classified according to its type (the classification is based on the elements defining the domain values of the slot). The order of magnitude associated with numerical slot values and the arborescent organization of symbolical slot possible values are introduced respectively in the sub-classes NUMERICAL SLOT and SYMBOLICAL SLOT by means of slots created for this task (slots <order-of-magnitude> and <domain-organization>, a particular arborescence being represented by an instance of the class TREE, which is a pre-defined class of the language). The domain expert must specify all these elements of knowledge.

Finally, in this model, each slot is characterized by a threshold (noted α) which specifies the minimal level of matching which is authorized between values of this slot in the class and values of this slot in the instances linked to the mentioned class (because our model permits a partial matching).

8) DESCRIPTION OF THE MATCHING STEP

The matching step utilizes the elements of knowledge presented in section 7. We describe in section 8.1 how the OTC description is completed. Then, we present how measures of compatibility (section 8.2), incompatibility (section 8.3) and typicality (section 8.4) are computed. These measures occur in the three following points:

- **Definition of the notion of solution**

C is a solution for the classification of OTC if it satisfies the two following constraints:

$comp(C, OTC) \geq minimal\_compatibility\_threshold$ (threshold1)
$incomp(C, OTC) < maximal\_incompatibility\_threshold$ (threshold2)

where comp(C, OTC) and incomp(C, OTC) are respectively the compatibility and incompatibility degrees between C and OTC and threshold1 and threshold2 are non-contextual parameters of the classification.

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

• Definition of the notion of best solution

Let C' be a direct or indirect sub-class of C. C' is a better solution than C for OTC if C' is a solution and satisfies

$$comp(C', OTC) > comp(C, OTC) \quad (3)$$
$$incomp(C', OTC) < incomp(C, OTC) \quad (4)$$

If for (3) and (4) an equality is found, the best solution is the one which has the best measure of typicality[5]. Then the chosen solution is the most typical and not necessarily the most specific. If C' satisfies (3) but not (4) and if typic(C', OTC) > typic(C, OTC), then we choose C' because we privilege the common slots in the evaluation of similarity.

• Ending of the search in a branch of the hierarchy

If C is not a solution and comp(C, OTC) = 0, we do not search for the sub-hierarchy of root C. Otherwise, we continue the search in the sub-hierarchy of root C even if C is not a solution because the previous measures are not monotonous.

8.1) Consideration of equivalence between valued slots

Before comparing the descriptions of OTC and the considered class C, we try to complete the description of OTC to perform the comparison. When a valued slot (sl, v) of OTC is not a slot of C, we look for the presence of (sl, v) in one of the values of the <equivalent-slots> slot defined in C. If (sl, v) is equivalent to (sl', v'), which is a slot of C, (sl', v') is added into the OTC description.

8.2) Measure of compatibility

The measure of compatibility evaluates the similarity between the class C and OTC. It is computed from common slots of OTC and C. For evaluating similarity, an elementary compatibility measure is defined and is applied to a slot.

• Elementary compatibility

Let (sl, v_{otc}) be a valued slot of OTC. If a slot within C is only defined by a domain of values, the elementary compatibility, for this slot, is total if the values of the same slot within OTC are included in the domain of values; otherwise, the elementary compatibility is null. There are two cases: (i) either the slot sl of C is not valued and the elementary compatibility is computed from type constraints; (ii) or the slot of C is valued (sl, v_c) and a matching is realized between the values v_c and v_{otc}. First, we syntactically match these values; if it succeeds, the elementary compatibility is 1. Otherwise the approximate matching is used, where different kinds of matching are possible according to the nature of the concerned slot: fuzzy matching or semantic matching. Approximate matching provides a degree between 0 and 1. This result, noted $match(v_c, v_{otc})$, is compared with the threshold α. If $match(v_c, v_{otc}) < \alpha$, then $comp_{sl} = 0$, otherwise $comp_{sl} = match(v_c, v_{otc})$ where $comp_{sl}$ is the elementary compatibility of sl.

• Global compatibility

For all the common slots of C and OTC, the elementary compatibility computation is repeated. The global compatibility is 0 if there is at least one elementary compatibility equal to 0, otherwise, the global compatibility is the following weighted sum:

$$comp(C, OTC) = \left(\sum_{i=1}^{n} comp_{sli} \times rel_{sli} \right) / \left(\sum_{i=1}^{n} rel_{sli} \right)$$

[5] The measure of typicality, noted typic(C, OTC), is not systematically computed.

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

where rel_{sli} is the relevance degree of the slot sl in the class C.

8.3) Incompatibility measure

The incompatibility measure evaluates the dissimilarity between C and OTC. Its computation is based upon the slots which distinguish C from OTC, i.e. only slots which are present in C. In fact, this measure evaluates incompleteness of OTC much more than the incompatibility between these two objects (this measure is a kind of penalty).

- **Elementary incompatibility**

Two elements occur in this calculus: the relevance degree associated with a slot sl of C and the typicality of the default values within the considered slot. Three cases are possible: (i) sl is only described by a domain and a cardinality: the penalty of this absence of sl within OTC is null; (ii) sl is valued: the penalty is equal to the pertinence degree of the slot (the stronger the relevance of the slot, the higher the penalty is); (iii) sl is described by means of one or more default values. The typicality of each default, computed from the strength (noted $o(q)$) of the linguistic quantifier it is associated with, occurs in the computation of the penalty, where $\frac{\sum o(q_i)}{m} \times rel_{sl}$, if sl is described by m defaults. In other words, the presence of defaults (and not of values) decreases the penalty due to the absence of sl within OTC. In the three cases, $incomp_{sl}$ is obtained.

- **Global incompatibility**

The global incompatibility is the normalized sum of elementary incompatibilities,

$$incomp(C, OTC) = \left(\sum_{i=1}^{n} incomp_{sli} \right) / n$$

8.4) Typicality measure

The typicality measure evaluates the typicality between C and OTC. It is computed from the common slots of C and OTC which are described within C by default values.

- **Elementary typicality**

Let (sl, v_{otc}) be a slot of OTC and $(sl, (d, q))$ the same slot within C described by a default value (we consider, here, that sl is mono-valued, to simplify), where d is the default, and q is the linguistic quantifier associated with the default. In a first step, v_{otc} and d are matched, using the most appropriate method and $match(d, v_{otc})$ is the result of the matching, which belongs to the interval [0, 1]. In a second step, the typicality of the default d is evaluated by computing $o(q)$. The elementary typicality is $typic_{sl} = min (match(d, v_{otc}), o(q))$.

- **Global typicality**

Two elements occur in the calculus: the elementary typicality of concerned slots and the typicality of the correlated slots. We focus on the calculus of the correlation typicality. Let $(sl, (d, q))$ and $(sl', (d', q'))$ be correlated slots respectively described by the defaults d and d'. The co-occurrence frequency is characterized by q_c, a linguistic quantifier. These slots within OTC are (sl, v_{otc}) and (sl', v_{otc}'). $match(d, v_{otc})$ and $match(d', v_{otc}')$ and $o(q_c)$ are computed. Then we obtain $typic_{correl} = min (match(d, v_{otc}), match(d', v_{otc}'), o(q_c))$. The typicality of a correlation is more meaningful than an elementary typicality of a slot occurring in the correlation. With n the number of slots for which an elementary typicality is computed and n' the number of correlations, the global typicality is computed as

$$typic(C, OTC) = 1/2 \left(\frac{\sum typic_{sli}}{n} + \frac{\sum typic_{correli}}{n'} \right)$$

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

CONCLUSION

In this paper, we have presented a model that attempts to take into account the notions of typicality and similarity presented in the psychological literature to perform the task of classification in a frame-based language. Our model proposes to consider particular semantic of default values to evaluate the typicality between a class and the OTC. In this way, the possibilities of typicality representation proposed by frame-based languages are developped. To find the most appropriate solution we do not systematically choose the most specific solution to classify OTC but we retain the most typical solution for OTC. Representation and manipulation of approximate knowledge are integrated in the frame-based language Objog+. Approximate knowledge is used to make the matching more flexible. Additional knowledge is also provided which leads to the enrichment of classification reasoning. Additional knowledge concerns heuristics and elements of cognitive psychology which are taken into account in the instance classification task.

REFERENCES

(Barsalou 89) Barsalou L.W., Intraconcept similarity and its implications for interconcept similarity, Similarity and Analogical Reasoning, Vosniadou S.and Ortony A.(eds), Cambridge Press University, pp76-121, 1989.

(Cohen and Murphy 84) Cohen B., Murphy G., Models of Concepts, Cognitive Science, 8, pp27-58, 1984.

(Dubois and Prade 87) Dubois D., Prade H., A theory of possibility, Plenum Publishing, 1987.

(Dubois et al. 88) Dubois D. Prade H., Testemale C., Weighted Fuzzy Pattern Matching, Fuzzy Sets and Sysytems 28, 313-331, 1988.

(Faucher 91) Faucher C., Elaboration d'un langage extensible fondé sur les schémas. Le langage Objog+, doctoral dissertation : université d'Aix-Marseille III, 1991.

(Faucher, Chouraqui, Gonzalez 95) Faucher C., Chouraqui E., Gonzalez-Gomez M., Fuzzy Extension of the Frame-based Language Objog+, Procceeding of SCSC'95 (Summer Computer Simulation Conference) - SCS, Ottawa (Canada), to appear, 5 pages, July 1995.

(Fikes and Kehler 85) Fikes R., Kehler T., The Role of Frame-based Representation in Reasoning, Communications of the ACM, vol 28, n°9, 1985.

(Finin and Silverman 86) Finin T., Silverman D., Interactive Classification as a Knowledge Acquisition Tool, Expert Database Systems, pp79-90, 1986.

(Gati and Tversky 84) Gati I., Tversky A., Weighting Commun and Distinctive Features in Perceptual and Conceptuel Judgements, Cognitive Psychology, 16, pp341-370, 1984.

(Gonzalez, Faucher, Chouraqui 95a) Gonzalez-Gomez M., Faucher C., Chouraqui E., Approximate Knowledge Modelling in a Frame-Based Language, in Proceedings of FLAIRS'95 (Florida Artificial Intelligence Research Symposium, April 26-29) pp52-56, 1995.

(Gonzalez, Faucher, Chouraqui 95b) Gonzalez-Gomez M., Faucher C., Chouraqui E., Modelling Fuzzy Frame Hierarchy to Represent Approximate Knowledge, in Proceedings of ISFL'95 (International Symposium on Fuzzy Logic, Zurich, May 26-27), ppA18-A25, 1995.

(Hampton 93) Hampton J., Prototype Models of Concept Representation, in Categories and Concepts, Theoretical Views and Inductive Data Analysis, Mechelen I.V., Hampton J., Michalski R.S., Theuns P. eds, pp67-95, 1993.

(Malt and Smith 84) Malt B., Smith E., Correlated Properties in Natural Categories, Journal of Verbal Learning and Verbal Behavior 23, 250-269, 1984.

(Medin and Ortony 89) Medin D., Ortony A., Psychological essentialism, Similarity and Analogical Reasoning, Vosniadou S.and Ortony A.(eds), Cambridge Press University, pp179-195, 1989.

(Medin and Goldstone 93) Medin D.L., Goldstone R.L., Gentner D., Respects for Similarity, Psychological Review, vol 100, n°2, pp254-278, 1993.

(Medin and Heit 94) Medin D.L., Heit E., Categorization, Handbook of Cognition and

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

Perception : Cognitive Science, D.E. Rumelhart and B.O. Martin (eds), San Diego, Academic Press, 1994.

(Minsky 75) Minsky M., A Framework for Representing Knowledge, in P. Winston eds, The Psychology of Computer Vision, pp211-281, McGraw-Hill, New York 1975.

(Murphy and Medin 85) Murphy G., Medin D., The Role of Theories in Conceptual Coherence, Psychological Review 92, 289-316, 1985.

(Murphy 93) Murphy G., A Rational Theory of Concepts, The Psychology of Learning and Motivation, vol 29, 327-359, 1993.

(Porter et al. 90) Porter B., Bareiss R., Holte R., Concept Learning and Heuristic Classification in Weak-Theory Domains, Artificial Intelligence 45, 229-263, 1990.

(Rosch and Mervis 75) Rosch E., C. B. Mervis, Family Resemblances : Studies in the Internal Structure of Categories, Cognive Psychology, 7, pp573-605, 1975.

(Tversky 77) Tversky A., Features of Similarity, Psychological Review 84, 327-352, 1977.

(Tversky and Gati 78) Tversky A., Gati I., Studies of Similarity, in Cognition and Categorization, E. Rosch and B. Lloyd (eds), Hillstade, NJ : Erlbaum, pp80-99, 1978.

(Yager 94) Yager R.R., Interpreting Linguistically Quantified Propositions, International Journal of Intelligent Systems, vol 9, n°6, 542-569, 1994,.

(Zadeh 85) Zadeh L.A., Syllogistic Reasoning in Fuzzy Logic and its Application to Usuality and Reasoning with Dispositions, IEEE 6, 754-763, 1985.

ISSN: 2324-9773

Faucher, C., Gonzalez-Gomez, M., Chouraqui, E., Chouraqui, E., & Chouraqui, E. (1995). SEMANTIC CLASSIFICATION OF INSTANCES IN A FRAME-BASED REPRESENTATION. 6th ASIS SIG/CR Classification Research Workshop, 21-36. doi:10.7152/acro.v6i1.12659

ISSN: 2324-9773