Stephens, L. M. (1990). The classification of semantic relations based on primitive properties. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 161-170. doi: 10.7152/acro.v1i1.12476

The Classification of Semantic Relations Based on Primitive Properties

Larry M. Stephens

Department of Electrical and Computer Engineering
University of South Carolina, Columbia, SC 29208, USA

ABSTRACT

Although much research has been done on the classification of objects (for example, taxonomic hierarchies in biology), there has been little research to date on the classification of semantic relations among objects. This paper describes a method that uses the definitional properties of semantic relations to classify the relations in a manner similar to that used for the classification of nonrelation objects. The classification schema is being implemented in CYC, a frame-based, knowledge representation system under development at the Microelectronics and Computer Technology Corporation, Austin, Texas [Lenat and Guha, 1990]. CYC researchers are attempting to encode into a computer common-sense knowledge about the real world, along with mechanisms for reasoning about that knowledge. CYC currently contains more than 5,100 relations (called *slots*), which are classified loosely according to several different schemes. This paper presents a uniform classification scheme that is being used to reorganize these slots. The limitations of the scheme are also discussed.

INTRODUCTION

In a frame-structured knowledge base, objects are represented by a collection of slot-value pairs. In the CYC system, the slots are also encoded as frames--a representation that allows the properties of the slots themselves to be expressed declaratively. When a user represents a new object in the knowledge base, he must use a subset of the existing collection of slots. If suitable slots are not available, the user must create them. In a sense, the set of currently available slots acts as an "instruction set" for knowledge entry. As more and more knowledge is added to CYC, the number of slots increases and so does the probability that needed slots are already represented; however, *finding* the slots becomes more difficult unless a principled scheme for their classification is devised.

Knowledge in CYC consists of slots, classes, and instances, each of which is an instance of some object that represents a class. The knowledge is represented by frame-like structures, which can be partitioned into those that represent slots and those that do not. In the simplified example frames below, the "instruction set" for knowledge entry consists of the slot set *{age, hasChild, domain, range}*.

Nonslot Frame: Mike	**Slot Frame:** hasChild
age: 42 hasChild: Hilary	domain: AdultPerson range: Person

ISSN: 2324-9773

Stephens, L. M. (1990). The classification of semantic relations based on primitive properties. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 161-170. doi: 10.7152/acro.v1i1.12476

This paper concentrates on *binary* slots, each of which may be viewed as a mathematical binary relation: a set of ordered pairs formed by specifying, either *intensionally* or *extensionally*, a subset of the Cartesian product of the domain and range of the relation [Stanat and McAllister, 1977]. Thus, in the example slot above, the slot values for the domain (*AdultPerson*) and the range (*Person*) must represent classes (*sets* of objects).

The frame representing a slot may be considered as the *intensional* specification for that slot; the slot's frame describes the properties of the relation and provides some information about specifying the mapping from the domain to the range. However, the frame for a slot does not explicitly list the set of ordered pairs comprising the relation. An *extensional* specification for a slot may be obtained from the knowledge base by finding all frames that use the slot and forming ordered pairs consisting of those frames and the values for that slot in those frames. For example, the extensional specification for *hasChild* would be obtained by finding all frames that use the slot *hasChild*. If the frame *Mike* above is the only one found to use the slot *hasChild*, then the *hasChild* relation is explicitly the set {*<Mike, Hilary>*}. If we update the knowledge base with frames *Hilary* and *John*, shown below, to the knowledge base, the relation *hasChild* becomes the set {*<Mike,Hilary>,<Hilary,Jason>,<John,Jane>,<John,Mary>*}.

Hilary

age: 21
hasChild: Jason

John

age: 33
hasChild: (Jane Mary)

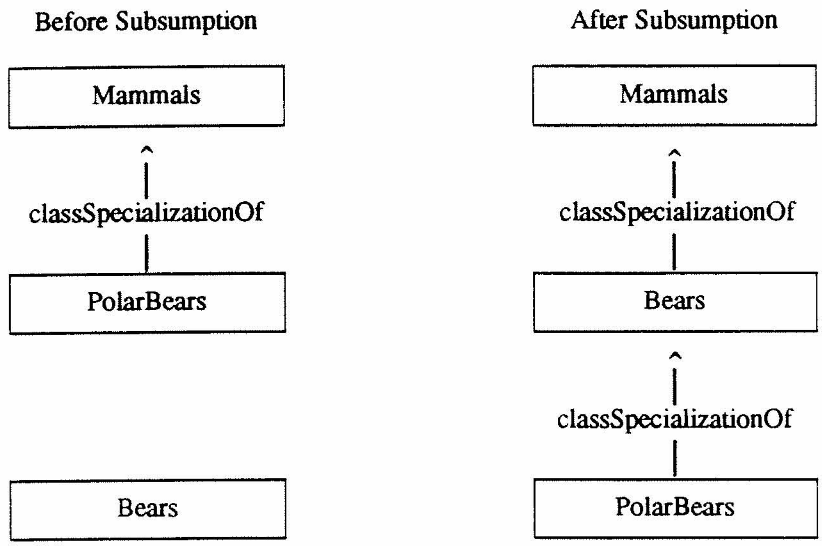
CLASSIFYING KNOWLEDGE

The need for automatic classification of knowledge in artificial intelligence systems has been widely recognized, and algorithms have been implemented to classify objects in a number of systems [Brachman and Schmolze, 1985], [Abrett and Burstein, 1987], [Finin, 1986]. In these systems, the primary task of the classifier is to compare a new frame with existing frames and compute *subsumption*. For one frame to subsume another, the full set of defining features of the subsuming frame must be a subset of the features of the definition being subsumed. The subsuming frame is, therefore, more general than the frame being subsumed. Important knowledge base functions achieved by subsumption-based classification include maintaining consistency, detecting redundancy, and simplifying algorithms for accessing frames based on explicit hierarchical paths going from subsumer (the more general) to subsumee (the more specific).

Consider a particular application of updating a knowledge base with a new frame representing a class. As seen in the following rather typical example, a new frame *Bears*, representing the set of all bears, is subsumed by *Mammals*; *Bears* in turn subsumes the existing frame *PolarBears*:

ISSN: 2324-9773

Stephens, L. M. (1990). The classification of semantic relations based on primitive properties. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 161-170. doi: 10.7152/acro.v1i1.12476



Figures 1 and 2 show the frame representation for these classes. In CYC, the properties of the *instances* of a class are encoded as indented slots and values under the slot *allInstancesHave* for the class. For example, the frame for mammals is interpreted to mean that all instances of mammals bear their young alive. The values of the slots *classSpecialization* and *classGeneralization* in these examples are computed by the subsumption algorithm based on the slots and slot values specified for the instances of the classes.

Mammals

```
allInstancesHave
    reproductionMethod: (LiveBirth)
classSpecialization: (PolarBears)
```

PolarBears

```
allInstancesHave
    reproductionMethod: (LiveBirth)
    hasBodyCovering: (Fur)
    hasColor: (White)
classGeneralization: (Mammals)
```

Figure 1. Knowledge base before new frame is added.

ISSN: 2324-9773

Mammals

allInstancesHave
    reproductionMethod: (LiveBirth)
classSpecialization: (Bears)

Bears

allInstancesHave
    reproductionMethod: (LiveBirth)
    hasBodyCovering: (Fur)
    hasColor: (Brown Black White)
classGeneralization: (Mammals)
classSpecialization: (PolarBears)

PolarBears

allInstancesHave
    reproductionMethod: (LiveBirth)
    hasBodyCovering: (Fur)
    hasColor: (White)
classGeneralization: (Bears)

Figure 2. Knowledge base after new frame is added.

CLASSIFYING RELATIONS

In this paper, the domain and range are considered as *definitional properties* for slots. In addition, certain primitive properties, called *relation primitives*, are also considered definitional. If slots are represented declaratively using their definitional properties, then slots may be classified using the same principles that are used to classify nonslot frames. The slot taxonomy can be based on the following definition of slot specialization [Huhns and Stephens, 1988]:

> *A slot S_1 is a* **slot specialization** *of a slot S_2 if and only if S_1 has all of the definitional properties of S_2, plus at least one of the following: 1) an additional definitional property, 2) a more restricted value of a definitional property, and 3) a more specialized definitional property. In particular, slot S_1 is a specialization of slot S_2 if the domain of S_1 is a class specialization of the domain of S_2 or the range of S_1 is a class specialization of the range of S_2 or both. The inverse of slot specialization is* **slot generalization**.

This definition is *not* based on a requirement that the relation S_1 (mathematically, a set of ordered pairs) be a subset of the relation S_2.

ISSN: 2324-9773

Stephens, L. M. (1990). The classification of semantic relations based on primitive properties. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 161-170. doi: 10.7152/acro.v1i1.12476

As an example of how one slot may be a specialization of another, consider two semantic relations that have identical defining properties except for range values. Suppose that the domain of the relations *hasChildren* and *hasDaughter* is the class *AdultPerson*, the range of *hasChildren* is *Child*, and the range of *hasDaughter* is the class *Girl*, a specialization of the class *Child*. The slot *hasDaughter* then becomes a specialization of the slot *hasChildren*, regardless of any occurrences of these slots in the knowledge base. These specializations are illustrated in Figure 3.

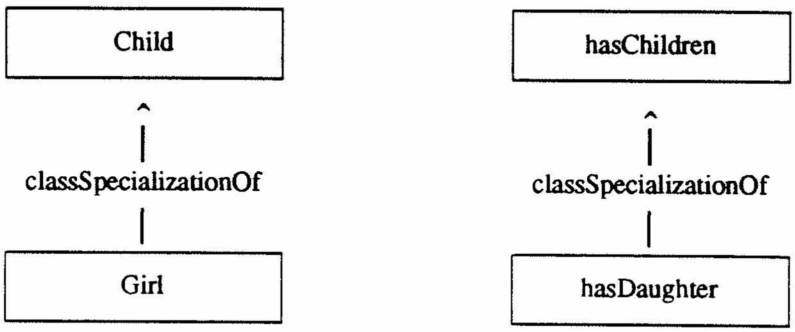


Figure 3. Class and slot specializations.

The next section presents relation primitives as definitional properties and discusses their usefulness for classifying slots.

RELATION PRIMITIVES

[Huhns and Stephens, 1989] identified a group of ten relation primitives that can be used to predict plausible inferences. Each relation primitive is a fundamental property that holds between an element of the domain and element of the range of the relation. These primitives were derived from a literature survey [Cohen and Loiselle, 1988], [Winston *et al.*, 1987] and an analysis of numerous semantic relations in the CYC knowledge base [Lenat and Guha, 1990]. These primitives are independently determinable for each relation and relatively self-explanatory. They specify a relationship between an element of the domain and an element of the range of the semantic relation being described.

These primitives may be divided into groups according to the values they assume. One group takes on values from the set {+, -}, where + indicates that the relationship holds and - that it does not. (In [Huhns and Stephens, 1989] a value of 0 was used to signify that the primitive did not apply; this is equivalent to the absence of that primitive.) Selected primitives from this first group are described below. In the following discussion, the notation *a.R.b* is used to indicate that the tuple <*a*, *b*> is an element of the relation *R*.

ISSN: 2324-9773

Stephens, L. M. (1990). The classification of semantic relations based on primitive properties. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 161-170. doi: 10.7152/acro.v1i1.12476

Functional: The domain of a Functional relation is in a specific spatial or temporal position with respect to the range of the relation. For example, in an instance of the *componentOf* relation, such as *Wheel.componentOf.Car*, the *Wheel* is in a specific spatial position with respect to the *Car*. This property does not hold for *Juror.memberOf.Jury*.

Homeomerous: In each instance of a Homeomerous relation, the element of the domain must be the same kind of thing as the element of the range. For example, in *PieSlice.pieceOf.Pie*, the slice is the same stuff as the pie.

Separable: The domain of a Separable relation can be temporally or spatially separated from the range, and can thus exist independently of the range. For the above *componentOf* example, the *Wheel* can be separated from the *Car* and can exist independently. For *Aluminum.constituentOf.Wheel*, the *Aluminum* cannot be separated from the *Wheel* if the *Wheel* is still to exist as an object.

Near: The domain of a relation with property Near is physically or temporally close to the range.

Connected: The domain of a relation with property Connected is physically or temporally connected to the range. A connection, which may be indirect, is indicated by +; no connection is denoted by -.

The primitives Structural and Intangible, described below, comprise a second group and take on values from the set {*Higher, Lower, Neutral*}. (In [Huhns and Stephens, 1989] these values were denoted {+, -, 0}. The new values are more accurate conceptually.) These primitives characterize relations in which a hierarchy exists between an element of the domain and an element of the range. If the element of the domain is *higher* in the hierarchy than the element of the range, then the primitive takes on the value *Higher*. If the element of the domain is *lower* in the hierarchy than the element of the range, then the primitive takes on the value *Lower*.

Structural: The domain and range of a Structural relation have a hierarchical relationship in terms of a physical structure. For example, in the relation tuple *Wheel.componentOf.Car*, the hierarchical structure is from part to whole, and the Structural property of *componentOf* has a value of *Lower*, indicating that *Wheel* is subordinate in the hierarchy. The converse relation, *hasComponent*, has *Higher* as its Structural value.

Intangible: The domain and range of an Intangible relation have a hierarchical relationship in terms of ownership or mental inclusion. As an example, the relation *ownedBy* has a value of *Lower* for Intangible, because the element owned is intangibly subordinate to the owner's sphere of influence.

ISSN: 2324-9773

Stephens, L. M. (1990). The classification of semantic relations based on primitive properties. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 161-170. doi: 10.7152/acro.v1i1.12476

These primitives impose restrictions on some relations and their converses. If the tuple *<a,b>* is an element of binary relation *R* (*a.R.b* in the alternative notation), then the tuple *<b,a>* is an element of the *converse* binary relation R^C ($b.R^C.a$). As noted above, a slot and its converse must have *opposite* values for the primitives Structural and Intangible. In CYC, some slots may be their own converses; for example, the slot *spouse* is its own converse. Such slots are mathematically *symmetric*. If a symmetric binary slot possesses either the primitives Structural or Intangible, then that slot must necessarily have a value of *Neutral* (its own opposite) for the primitive.

As an example of how these primitives may be used to classify slots, consider the slot *physicalPartOf*, a very general semantic relation that has the set {+ -} as the value of its Functional primitive, and the more specific relations *componentOf* and *constituentOf*, which have values of + and - respectively for their Functional primitives. If all other defining properties (domain value, range value, and other primitives and values) of these slots are identical, then both *componentOf* and *constituentOf* are slot specializations of *physicalPartOf*. The slot specialization links exist because the value sets (+) and (-) are more specific than (subsets of) the value set {+ -}. The following slot frames illustrate this example. In these frames, the slots *specializedSlots* and *generalizedSlots* are *nondefinitional, taxonomic* slots that record the results of slot classification.

physicalPartOf

```
domain: IndividualObject
range: IndividualObject
functional: (+ -)
specializedSlots: (componentOf constituentOf)
```

componentOf

```
domain: individualObject
range: individualObject
functional: (+)
generalizedSlot: physicalPartOf
```

constituentOf

```
domain: IndividualObject
range: IndividualObject
functional: (-)
generalizedSlot: physicalPartOf
```

As a further example, if the slot *componentOf* were to have as its domain or range value a more general class than *IndividualObject*, then *componentOf* would not be a slot specialization of *physicalPartOf*. Finally, it should be noted that a more general relation might not have some of the primitives of the corresponding specialized relations. This is consistent with the definition of generalization for nonslot objects.

ISSN: 2324-9773

Stephens, L. M. (1990). The classification of semantic relations based on primitive properties. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 161-170. doi: 10.7152/acro.v1i1.12476

DISCUSSION AND FUTURE RESEARCH

CYC now supports several methods for searching through the set of available slots, which now number more than of 5,100 slots. For example, the taxonomy of slot classes may be easily examined. In addition, CYC can list all the allowable slots for instances of a given class. (These are simply the slots for which the domain value is the given class.) A knowledge enterer can browse this list for promising candidate slots and display each to examine other properties (such as range) in more detail. But this can be time consuming. For example, the class *Person* may have any of 971 slots. Often novice users abandon the search and create new slots unnecessarily--making the number of slots even larger and further complicating searches later.

Given these problems, our research goal is to discover ways to organize slots so that they are easy to find when entering new knowledge. The expectation is that if a set of defining properties can be found to describe slots, then these can be used to help in searching the knowledge base for slots. However, as the following list demonstrates, there are many unanswered questions:

- Are the proposed primitives adequate for representing slots? Is each necessary? Is the set of primitives sufficient?
- Once a knowledge enterer selects a primitive, how does he decide on its value?
- Will the proposed scheme be more useful than the slot-searching techniques already available in CYC? Are we trading one set of problems for another?
- What metrics can be used to measure the utility of the slots classification algorithm.
- How do we represent the primitives? After all, they themselves are slots and may have a frame-based representation. Can the primitives be defined in terms of themselves, or must we find yet another set of primitives for them; and so on recursively?

We have partial answers for some of these questions:

- The primitives were represented in terms of themselves, but when composed with each other (as in [Huhns and Stephens, 1989]), were found not to be orthogonal. Nevertheless, the search for primitives of primitives is not being pursued for the moment.
- The primitives and their values were selected as being as unambiguous and easy-to-understand as possible. But we recognize that value assignment is a subjective process and that different users may disagree on the values chosen.
- There is evidence that the primitives chosen do represent fundamental properties of slots [Chaffin and Herrmann, 1984, 1987, 1988], [Winston *et al.*, 1987]. However, we do not know if we have a complete list or if these primitives are suitable for slot classification. The research for slot primitives continues.

A research plan was been devised to address the above questions. The plan includes an analysis of the current CYC slot taxonomy, a measurement of its complexity and inferential power, a test of the algorithm on a subset of slot space, and an evaluation of the algorithm's effectiveness. At this time we have only preliminary results and look forward to the classification workshop for suggestions and comments from the research community.

ISSN: 2324-9773

Stephens, L. M. (1990). The classification of semantic relations based on primitive properties. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 161-170. doi: 10.7152/acro.v1i1.12476

REFERENCES

[Abrett and Burstein, 1987] G. Abrett and M. H. Burstein, "The KREME Knowledge Editing Environment," *International Journal of Man-Machine Studies*, vol. 5, 1987, pp. 103-126.

[Brachman and Schmolze, 1985] R. J. Brachman and J. G. Schmolze, "An Overview of the KL-ONE Knowledge Representation System," *Cognitive Science*, vol. 9, 1985, pp. 171-216.

[Chaffin and Herrmann, 1984] R. Chaffin and D. J. Herrmann, "The Similarity and Diversity of Semantic Relations," *Memory & Cognition*, vol. 12, no. 2, 1984, pp. 134-141.

[Chaffin and Herrmann, 1987] R. Chaffin and D. Herrmann, "Relation Element Theory: A New Account of the Representation and Processing of Semantic Relations," in D. Gorfein and R. Hoffman, (eds.), *Memory and Learning: The Ebbinghaus Centennial Conference*, Lawrence Erlbaum, Hillsdale, NJ, 1987.

[Chaffin and Herrmann, 1988] R. Chaffin and D. Herrmann, "The nature of semantic relations: a comparison of two approaches," in *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*, M. W. Evens, ed., Cambridge University Press, Cambridge, England, 1988, pp. 289-334.

[Cohen and Loiselle, 1988] P. R. Cohen and C. L. Loiselle, "Beyond ISA: Structures for Plausible Inference in Semantic Networks," *Proceedings AAAI-88, St. Paul, MN, August 1988*, pp. 415-420.

[Finin, 1986] T. W. Finin, "Interactive Classification: A Technique for Acquiring and Maintaining Knowledge Bases," *Proceedings of the IEEE*, vol. 74, no. 10, October 1986, pp. 1414-1421.

[Huhns and Stephens, 1988] M. N. Huhns and L. M. Stephens, "Knowledge Classification in CYC (Research Note)," MCC Technical Report No. ACA-AI-091-88-Q, April 1988.

[Huhns and Stephens, 1989] M. N. Huhns and L. M. Stephens, "Plausible Inferencing Using Extended Composition," *Proceedings of IJCAI-89*, Detroit, MI, August 1989, pp. 1420-1425.

[Lenat and Guha, 1990] D. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*, Addison Wesley Publishing Company, Inc., Reading, Massachusetts, 1990.

[Stanat and McAllister, 1977] D. F. Stanat and D. F. McAllister, *Discrete Mathematics in Computer Science*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1977.

ISSN: 2324-9773

Stephens, L. M. (1990). The classification of semantic relations based on primitive properties. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 161-170. doi: 10.7152/acro.v1i1.12476

[Winston et al., 1987] M. E. Winston, R. Chaffin, and D. Herrmann, "A Taxonomy of Part-Whole Relations," *Cognitive Science*, vol. 11, 1987, pp. 417-444.

ISSN: 2324-9773