Christophe Jouisi
IDIST, University CHARLES-DE-GAULLE - Lille 3
& LALIC, University La Sorbonne, France

System of Types + Inter-concept Relations Properties: towards Validation of Constructed Terminologies?

Abstract: Building reference terminology (i.e. conceptual structure: identification of terms or concepts names related to a specific field of knowledge and the logico-semantic relations they hold) from extensive textual data is not a simple task when the designer has to examine a new field of knowledge. We observe special problems due to the fact that inter-concept relations are not very often specified in Knowledge-Based Systems, Thesaurus Construction (Documentary Databases, Information Retrieval) and Terminological Database construction, etc. For that reason, it is sometimes difficult to validate the constructed representation. In this proposal, we associate properties to relations by introducing the relations into a system of specifications. Therefore, it is possible to check, not the validation of the constructed conceptual structure, but its coherency in comparison to the established relations between the concepts, which is the necessary condition to validate a semantic representation.

1. Introduction

The aim of recent studies in semantics is to integrate into concept representations models issued from logic and/or from artificial intelligence (conception of knowledge-based systems). The fundamental problem entails the necessity to structure the knowledge in a micro-field and next to check that all constructed representations are valid. However different recent studies on terminology try to better specify and to extend terminological relations (see for instance Molhot 96, Green 96, and Bean 96). Several teams developed systems in order to "navigate" in networks of concepts (Terminological Databases): Termisti system (Van Campenhoudt 1996), Code system, COGNITERM project or Ikarus system (Meyer 1994), among others.

In this proposal, we associate properties to relations by introducing the relations into a system of specifications. Therefore, it is possible to check, not the validity of the constructed conceptual structure, but its coherency in comparison to the established relations between the concepts, which is the necessary condition to validate a representation.

The system of meanings proposed is based on a structured set of Semantic Primitives: types, relations, and properties. Also, before presenting it in section 3, it will be useful to define precisely what we understand by Semantic Primitives in section 2. The data-processing system that automatically checks coherence of Terminologies is finally described in section 4.

2. What we understand by Semantic Primitives?

Nowadays, two great theories are still facing each other: Universalism and Relativism. Universalism holds that reality is universal, it is always the same for everybody in any place but each language expresses this reality in its own way.

The frontier between 'res' and 'verba' is crucial: the concepts dealt with are alike but the words used to express it are different. It is always possible to translate a text since the concepts behind the words are the same in any language. Philosopher K. Popper wrote: "The fact is, that even totally different languages are not untranslatable" (Popper 1970, P. 56).

On the other hand, Relativism holds that thoughts and language are undoubtedly linked together. Thoughts are limited and structured by the language spoken. For Wilhelm Von Humbolt, founder of the relativist vision of language, our thoughts stem from the language we speak. Translation is no longer seen as a pure copy of the original text but as a reproduction of the way a translator understands a text at a given time. This has been labelled as the
**Hermeneutic** vision. Language is seen as an instrument used by man in order to classify and understand. The world we live in is no longer considered as a universal and standstill reality: our thoughts and our language determine it.

American Theorists such as Sapir and Whorf elaborated this theory (*The Sapir-Whorf theory*) and reconsidered the relativist theory (Sapir, 58 [1929], Whorf, 1956). In a classical article, Sapir explained in 1929: "Human beings do not live in the objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language which has become the medium of expression for their society. It is quite an illusion to imagine that one adjusts to reality essentially without the use of language and that language is merely an incidental means of solving specific problems of communication or reflection. The fact of the matter is that the 'real world' is to a large extent unconsciously built upon the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached... We see and hear and otherwise experience very largely as we do because the language habits of our community predispose certain choices of interpretation." (Sapir 1958 [1929], p. 69). Whorf who declared in another famous paper then promoted this point of view in the ‘30s: "We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organised by our minds - and this means largely by the linguistic systems in our minds. We cut nature up, organise it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organise it in this way - an agreement that holds throughout our speech community and is codified in the patterns of our language. The agreement is, of course, an implicit and unstated one, but its terms are absolutely obligatory; we cannot talk at all except by subscribing to the organisation and classification of data which the agreement decrees." (Whorf 56 [1940], pp. 213-14)

We admit that too many notions we believed were universal are in fact very local and are not linked to a general linguistic mechanism which would take into consideration the variety of languages. Nevertheless, our hypothesis based on Descles work states that: "The superficial differences that exist between the languages are neutralised by the abstract invariances that encompass a cognitive mechanism managed by the language activity" (Descles 90, P. 271). More precisely, by partially integrating the **Localist Hypothesis** (that suggests that languages structure is closely determined through space categorisation), we promote the idea that **some Elementary Semantic Primitives are common to at least several groups of language** (for instance, some Indo-European languages). The categorisation done by the perception organs (Sowa 84) and more precisely by our visual perception organs determine some primitives. In order to understand precisely the primitives we present in the following pages, it is useful to select several levels of representation. These levels stem from a linguistic model.

### 3. Typology of Inter-concept relations

We propose to structure Terminology in using a set of semantic relations. This set originates from a global model of language processing: **Applicative and Cognitive Grammar** (ACG, Descles 90), and an extension of this model to Terminology (Jouis, Mustafa 95, 96).

#### 3.1. The ACG Architecture

The ACG is an extension of the Universal Applicative Grammar (Shaumyan 87). It postulates three levels of representation of languages:
1. The Phenotype Level (or Phenotype) where the particular characteristics of natural languages are described (order of words, morphological cases, etc.). Each language is apprehended in the diversity of linguistic expressions which are directly observable. The linguistic expressions of this level are concatenated linguistic units.

2. The Genotype Level (or genotype) where grammatical invariants and structures that are underlying to sentences of phenotype level are expressed. The genotype level is structured as a formal language called genotype language. It is described by a grammar called Applicative grammar (Biskri & Desclés 97). Descriptions are represented in the form of Applicative expression formulated with operators and operands of different types.

3. The Cognitive Level where the meanings of lexical predicates are represented by semantic cognitive schemes. This level constitutes the knowledge representation associated to one text. Representations of levels two and three are expressions of typed combinatory logic (Curry & Feys 1958). With the cognitive level, the ACG proposes a set of semantic concepts, which defines an organised system of meanings.

3.2 Cognitive level: Primitives
   In this system, we distinguish four categories of primitive:

   1. Elementary Semantic Types of entities,
   2. Formator Operators in order to create more complicated types from elementary types (lists, arrays, functional types, etc.),
   3. Fundamental Static Relations between entities. Static relations permit the description of some states (static situations) related to an area of knowledge. Static situations remain stable during a certain temporal interval where neither the beginning nor the end are taken into account. We distinguish more than twenty relations.
   4. Fundamental Evolutive relations between terminological units. Evolutive relations permit descriptions of processes or events related to an area of knowledge (terminological unit denoting dynamical situations): movements, changes of state, conservation of a movement, iterations, intensity, variation, constraints, causes, etc.

   In this proposal, we will detail only part of the primitives of the ACG: Semantic Types and Static Relations because these are the primitives which are actually used in our ACG extension to Terminology. For a description of Evolutive relations, see (Descles 90), (Jouis 93).

3.2.1 Semantic types
   On the cognitive level of ACG, we distinguish a certain number of elementary types of entities. For instance:
   
   Individualisable entities are the ones which can be designated and shown by pinpointing. They may be counted individually or regrouped by classes. A quantification operator allows building a class more or less determined by individualised exemplars. For instance, entities such as John, table, chair, furniture, man, child are distinctive.

   Massive entities such as water, sea, sand, wine, butter are not distinctive entities. However, we may notice that a certain number of operators (classificators) can individualise massive notions: a glass of water, a blob of sand, an arm of the sea, a bottle of wine, a piece of butter.

   Distributive classes regroup individual entities with one identical property. For instance, to-be-a-square represents a class of individuals or "concepts".

   Collective classes are distinguished from the Individualisable entities because they represent the objects that form a "whole" from more elementary objects. Thus, geographical
entities, army, molecule (which is formed of different types of atoms which themselves are formed of...), human body, etc. represent collective classes.

Place, as a semantic type, is conceptualised as a set of positions, each position being assimilated to a point. To each entity (individualisable, collective or massive) we can relate a set of places. For instance, Paris, garden, house, can be on one hand seen as individualisable entities (Paris is a city) and on the other hand each individualisable entity determines a given place (I am in Paris).

3.2.2 Static Relations

Static relations are structured and independent from a particular domain. They are binary relations. We distinguish more than twenty relations, particularly:

- Identification (or equivalence) between two entities (example: "rendezvous" has the same meaning as "appointment").
- Incompatibility among entities (There is incompatibility between protons and electrons).
- Measures (This solution is pH 5).
- Cardinality (On January 1st, 1990, there was in France 169,051 doctors).
- Comparisons (Probability getting a red ball is superior to 0.5).
- Inclusion among distributive classes (Bacterium are microorganisms).
- Belonging of one individualisable entity to another distributive class (PI is a real).
- Localisations of one entity in one place (There are black shoes in the bag).
- Relations part/whole among collective classes (Fluorine goes into the composition of bones and teeth).
- Possession (John has got a car).
- Attribution (Every registration is characterised by a number)...

Relations are a part of a specification network built on a general terminological relation schema (i.e. a coherent system of meanings of relations).

![Figure 2: Specification Network of Static Relations](image)

The general schema of relation ("an entity X is in relation with an entity Y") is further specified according to the algebraic properties in more precise relations that are axiomatically attributed to them (see Figure 1).

In this system, a relation may be specified in other more precise relations in terms of its
properties: (1) its functional type (the semantic type of arguments of the relation); (2) its algebraic properties (reflexivity, symmetry, transitivity, etc.); (3) its combinatorial relations with other entities in a same context (the part of the text where a concept is defined for instance). For example, inclusion among distributive classes is not reflexive, but rather antisymmetrical and transitive. Moreover, in the same context, it is incompatible with some relations such as the belonging of one individual entity to one distributive class. A relation may be specified in other more precise relations in terms of its properties. For instance, localisation is specified in:

- Localisation in the interiority of a place (oscillating stands in the first zone). For instance, this relation is transitive, antisymmetrical and not reflexive.
- Localisation at the exteriority of a place (Limitor is exterior to the third zone). For instance, this relation is not transitive and never reflexive.
- Localisation on the edge-line (boundary) of a location (Algiers is at the seaside). For instance, this relation is incompatible with interiority and exteriority and more precise than the localisation at the closure of a locality.
- Localisation at the closure of a locality (Boulogne is located in the suburb of Paris). For instance, this relation is incompatible with interiority and redundant with the localisation on the edge-line and interiority.

4. Towards validation of constructed conceptual structures

One of the delicate points is the validation of a reference terminology during its construction. This validation can only be realised with the cooperation of specialists (i.e. field specialists, terminologists and/or knowledge engineers).

With the whole-defined relations by logico-semantic properties, it is possible to control the coherency of representations by checking that the properties of the relations are well applied (respected). The coherency is a necessary condition in order to validate conceptual structures. A Processing module, testing these properties, is in course of development in SEEK system (Jouis 93, Jouis & Mustafa 95). It integrates specific procedures of control for each property. This module uses GRAPHLET (Himsolt 94), a Graph Drawing Processing system. Procedures are released from the moment one introduces a new relation between two concepts. In the example below (see Figure 2), we present an extract of the concepts and relations representation using GRAPHLET system within the framework of project SEPCI.
Following, we give some examples of incoherent terminology representations detected automatically by our system. For instance, the relation of inclusion between two distributive classes \( D_1 \) & \( D_2 \) is not reflexive, but transitive and asymmetric (see Figure 3).

![Figure 3: Inclusion is not reflexive, but transitive and asymmetric.](image)

When a relation of inclusion is established between \( D_1 \) & \( D_2 \), it becomes incompatible with others relations as Part/Whole relations (see Figure 4).

![Figure 4: Inclusion and Part/Whole are incompatible in the same context.](image)

After that, it is impossible to have a relation of inclusion from \( D_2 \) towards \( D_1 \), or a part/whole relation between \( D_1 \) & \( D_2 \) in a same context. The coherence check of the representations shown leads to test several properties of various relations. A typical situation is given in Fig. 5.

![Figure 5: Typical situation where it is necessary to check coherence by testing several properties of various relations.](image)

Take for instance two distributive classes \( D_1 \) and \( D_2 \) which are declared incompatible (i.e. a specification of the opposition relation). If you want the under-hierarchy of concepts resulting from \( D_1 \) to remain coherent, it is no longer possible to have inclusions or memberships with the under-hierarchy resulting from \( D_2 \) (and reciprocally of \( D_2 \) towards \( D_1 \)).

6. Concluding Remarks

By introducing the relations into a system of specifications (types, relations, properties) it is possible to check, not the validation of the constructed conceptual structure, but its coherency in comparison to the established relations between the concepts, which is the necessary condition to validate a representation. The structured set of the relations consists of
semantic invariants. It is independent of a field of knowledge. However, we do not claim that it is complete. It is only one working hypothesis which should be validated. Our approach does not exclude the need for adding other relations more precise or specific to a particular field of knowledge.

Acknowledgements

These considerations are partially results issued from an ongoing studies within a program sponsored by the "Association des Universités Entièrement ou Partiellement de Langue Française" (AUPELF) - an international association whose mission is to promote the dissemination of French as a scientific medium. This research aims to evaluate software capabilities in automatic terminology building from corpuses in French (ARC A3). The author is also very grateful to Pr. Desclés (Paris-Sorbonne) for the discussions very profitable.

Notes

14 This hypothesis is confirmed through some experiences carried out to check the existence of cognitive invariances on colour perception by man (Kay 83).
15 For a more complete description of the relations' properties, see (JOUIS 93, pp 146-223).
16 Each object’s occurrence, in a particular pragmatic environment, determines a location (i.e. place or “neighbourhood” in the topology meaning). Primitives of position can be defined by calling upon some basic concepts of general topology. A location is then visualised either in its interiority, or in its exteriority (excluding its interiority and its closure), or as a closure (limits and interiority). We introduce the operators of topological determination of a place x: in(x), ex(x), fr(x) and fe(x), respectively determining the interior, the exterior, the boundary and the closure of x. For any place x, we have for instance:

\[
in(x) \subset x \subset fe(x)
\]

\[
fr(x) \subset fe(x) \text{ (because } fe(x) = x \cup fr(x))
\]

\[
x \cap ex(x) = \emptyset
\]

\[
fr(x) = co(in(x)) \cap co(ex(x))
\]

The properties of these four operators enable us to establish properties of localisation’s relations.

17 I.e. \( \forall x, \text{ NOT } (x \text{ loc-ex } x) \). It is about a property stronger than non-reflexivity which would give: \( \text{ NOT } (\forall x, (x \text{ loc-ex } x)) = \exists x \text{ NOT } (x \text{ loc-ex } x) \).

18 Project SEPCI ("Expert system of Placement of Commands and Informations") was carried out jointly by IPSE (Institute Polytechnic of SEVENANS, BELFORT) & by the company GEC-ALSTHOM. SEPCI is a system of assistance to the disposition of devices (information and commands). These devices will be installed in Man/machine interfaces, following the rules of ergonomics. The interfaces are desk type (for instance the desk used in the TGV - high-speed Train). The two ergonomicists tasks (Mr. Benchekroun, engineer in ergonomics at GEC-ALSTHOM and M.L. Roberty, Researcher at IPSE, Department of ergonomics), consisted in taking into account the constraints imposed by the manufacturing, but also to optimise the security and the working conditions of the driver.

References


Advances in Knowledge Organization, Vol.6(1998)


