LREC 2012 - Eighth international conference on Language Resources and Evaluation
Istanbul - 21-27 May 2012. pp. 2626-2630

# Ontoterminology: How to unify terminology and ontology into a single paradigm

**Christophe Roche**
Condillac Research Group – Listic Lab.
University of Savoie – Campus Scientifique
F – 73 376 Le Bourget du Lac cedex
E-mail: christophe.roche@univ-savoie.fr

**Abstract**

Terminology is assigned to play a more and more important role in the Information Society. The need for a computational representation of terminology for IT applications raises new challenges for terminology. Ontology appears to be one of the most suitable solutions for such an issue. But an ontology is not a terminology as well as a terminology is not an ontology. Terminology, especially for technical domains, relies on two different semiotic systems: the linguistic one, which is directly linked to the "Language for Special Purposes" and the conceptual system that describes the domain knowledge. These two systems must be both separated and linked. The new paradigm of ontoterminology, i.e. a terminology whose conceptual system is a formal ontology, emphasizes the difference between the linguistic and conceptual dimensions of terminology while unifying them. A double semantic triangle is introduced in order to link terms (signifiers) to concept names on a first hand and meanings (signified) to concepts on the other hand. Such an approach allows two kinds of definition to be introduced. The definition of terms written in natural language is considered as a linguistic explanation while the definition of concepts written in a formal language is viewed as a formal specification that allows operationalization of terminology.

**Keywords:** terminology, ontology, ontoterminology

## 1. Introduction

Terminology is assigned to play a more and more important role in the Information Society. The need for a computational representation of terminology for IT applications (computer aided translation, multilingual information retrieval, specialized encyclopaedias, semantic web, etc.) raises new challenges for terminology. Ontology, understood as a shared and formal specification of a domain conceptualisation, appears to be one of the most suitable solutions for such an issue. But an ontology is not a terminology whose terms would be lexicalised concepts as well as a terminology is not an ontology whose concepts would be term meanings. Terminology, especially for technical domains, relies on two different semiotic systems: the linguistic one, which is directly linked to the "Language for Special Purposes" used in writing technical documents and the conceptual system, which describes the domain knowledge. These two systems must be both separated – let us recall that the conceptual and lexical structures do not match [Roche 2007] – and linked.

The new paradigm of ontoterminology, i.e. a terminology whose conceptual system is a formal ontology, emphasizes the difference between the linguistic and conceptual dimensions of terminology while unifying them. A double semantic triangle is introduced in order to link the linguistic notions to the ontological ones; terms (signifiers) are linked to concept names on a first hand when meanings (signified) are related to concepts on the other hand. Such an approach allows two kinds of definition to be introduced. The definition of terms written in natural language is considered as a linguistic explanation – and then does not require to be standardised – while the definition of concepts written in a formal language is a formal and computational specification that allows operationalization of terminology for IT applications. Let us bear in mind that using a formal language like logic allows agreement commitment about concept definitions and their standardisation.

The article is divided into 4 parts. The first two set down the definitions of terminology and ontology used in this work. The third part introduces the new paradigm of ontoterminology and the last one presents an application-oriented validation of our work, the ASTECH FP6 European project, a multilingual information retrieval system in the domain of renewable energy.

## 2. Terminology

Although some people denies the independent status of terminology as a discipline [Sager 1990] claiming that it is a part of applied linguistics [Pavel and Nolet 2001], terminology appears as an independent discipline in the twentieth century [Wüster 1968]. The General Theory of Terminology, as defined by H. Felber [Felber 1984] on the basis of Wüster' lecture notes, is an attempt to distinguish terminology from linguistics [Cabré 2003].

Defined as a "set of designations belonging to one special language " [ISO 1087-1], the main goal of terminology is to eliminate ambiguity from technical languages by means of standardisation. In order to achieve such a goal, the General Theory of Terminology (Wüster) postulates the priority of the concept over the designation (term) and its universality independently of the diversity of languages. Felber [Felber 1984] gave three definitions of terminology that illustrate the importance of concept over the term: "inter- and transdisciplinary field of knowledge dealing with concepts and their representations (terms, symbols, etc.)"; "aggregate of terms, which represent the system of concepts of an individual subject field"; "publication in which the system of concepts of a subject field is represented by terms". A term is then any "conventional symbol representing a concept defined in a subject field" [Felber 1984].

Although the concept is the core point of the General Theory of Terminology (GTT), and unlike artificial intelligence, the main goal of terminology is not to represent concepts in order to manipulate (compute) them. The concept in terminology – an "unit of knowledge created by a unique combination of characteristics" [ISO 1087-1] – does not exist in itself. It exists only through its definition written in natural language or in a semi-formal language. Let us recall that "The terminology work dealt within this International Standard is concerned with terminology used for unambiguous communication in natural, human language. The goal of terminology work as described in this International Standard is, thus, a clarification and standardization of concepts and terminology for communication between humans." [ISO 704].

The need for a computational representation of concepts in terminology comes from IT application like (multilingual) Content Management Systems, (multilingual) Information Retrieval, Specialized Encyclopaedias or Semantic Web. Such applications raise new issues for Terminology. Although the "Principles of Terminology" of the GTT propose a certain number of paradigms demonstrating a scientific ambition to order reality – a mathematical structuralism – based on connected concept systems, some inaccuracies in these paradigms' definition make terminology operationalization difficult and explain knowledge engineering's leader status in this field. From the computational perspective alone, the Principles need to be re-examined [Roche 2008] in order to specify from a logical point of view some principles as well as to introduce new features from Artificial Intelligence and Knowledge Representation like valuable attributes. In this context, ontology appears as one of the most suitable and promising approach.

## 3. Ontology

Ontology of the knowledge engineering shares a similar goal with terminology: to enable communication and knowledge sharing between agents either human or software. It also relies on a similar principle: a shared and common conceptualisation based on standardisation. We can claim that ontology [Staab et al. 2004], [Gomez-Perez et al. 2004], [Roche 2003] represents one of the most promising and useful approach for terminology and its operationalization.

Nevertheless, we have to keep in mind that an ontology, defined as a "specification of a conceptualisation", is first of all "a description (like a formal specification of a program) of the concepts and relationships that can exist" [Gruber et al. 1993] in order to manipulate them either from a logical or computational point of view. Thus an ontology is not a terminology even if some definitions can suggest the opposite: "an [explicit] ontology may take a variety of forms, but necessarily it will include a vocabulary of terms and some specification of their meaning (i.e. definitions)" [Ushold et al. 1996]. Ontology doesn't take into account the linguistic dimension of terminology. As a matter of fact, terms cannot be reduced to arbitrary words (lexicalized concepts) or labels stuck onto concepts. Terms of usage, standardised terms, lexical forms (including terminological variations and reductions, rhetorical figures like ellipsis, etc.), connotative information as well as linguistic relationships, which are not taken into account in ontology, are central features in terminology. Furthermore an ontology does not necessarily define a valid conceptual system of a terminology.

As a matter of fact, an ontology is defined according to a given theory using a formal (or semi-formal) language following the epistemological principles of this artificial language. It implies that the way ontology and concept are defined depend directly on the formal language being used. For instance, description logic [Baader et al. 2003] is an example of logic adapted to knowledge representation whereas frame representation languages [Wright et al. 1984], in spite of the criticism of Woods [1975], provide semi-formal and more human-readable languages. The Web Ontology Language [OWL], combines the advantages (and disadvantages) of these two approaches.

## 4. Ontoterminology

Writing scientific texts and modelling knowledge are two different activities involving different languages – natural versus formal. They may define several different points of view about the "world". We claim that domain ontology, viewed as a scientific knowledge, can be defined independently from the different linguistic ways of speaking of it [Roche 2007] – even if useful information can be extracted from corpus [Buitelaar et al. 2005], [Daille et al. 2004].

Nevertheless a conceptualisation is more than a computational or formal representation of concepts. It must be guided by epistemological and terminological principles – and logic and computational languages are neither epistemological nor linguistic. These formalisms must be extended with for example rigid predicates

[Guarino et al. 1994a].

Terminology is not only the science of terms (specialized lexical units), but also the science of objects (which populate the world) which requires an epistemological approach.

Referring first the domain ontology in the terminology work had led us to introduce a new paradigm for terminology called ontoterminology: a terminology whose conceptual system is a formal ontology relying on epistemological principles [Roche 2009]. Like the General Theory of Terminology, ontoterminology is based on an onomasiological approach: “concepts in expert knowledge became the starting point in terminological analysis” when other ontology-oriented approaches like termontography [Temmerman 2000] remains semasiological: “terms (linguistic expressions) in texts became the starting point in terminological analysis”. Our intention is not to compare the two approaches, their goals remain different: the former focuses on conceptualisation whereas the latter focuses on specialised vocabulary. We should just bear in mind that the lexical structure extracted from a corpus does not match the conceptual structure directly defined by experts using a formal language: “saying is not modelling” [Roche 2007] (figures 1 and 2).

Concepts in ontoterminology exist in their own right. This implies that terms are separated from concepts as well as term definitions (written in natural language) are separated from concept definitions (written in formal language). This gives the possibility to manage the conceptual and linguistic dimensions of terminology and provide two kinds of definition: the first formally defines the concept whereas the second explains the term and its usage from a linguistic point of view.

The classic semantic triangle [Ogden and Richards 1923] had been extended towards a double semantic triangle (figure 1) [Roche 2007] in order to express the difference between the meaning of a term and the concept of the domain ontology as encountered in terminology. This double semantic triangle lays stress on the two linguistic and conceptual semiotic systems that compose each terminology. It identifies the different elements involved into the meaning building process as well as their relationships. Then a concept is not the signified and the identifier is not the signifier.

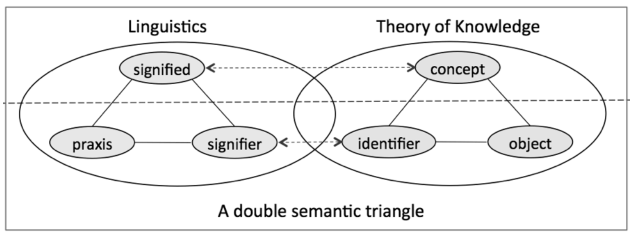


Figure 1: A double semantic triangle

Like classical terminology, ontoterminology enables standardisation of language. But unlike classical terminology ontoterminology preserves the diversity of language between different communities of practice since they share the same domain and standardised conceptualisation. In point of fact, two different terms can denote the same concept whose identifier should be built so that we understand the right place of the concept in the ontology. Standardised terms can be built from concept identifiers, even if they are not used in text – they are necessary for term meaning and understanding. For example “voltage relay” in English and “relais de tension” in French denote the same concept of <Voltage threshold relay> when the standardised term in French is “relais à seuil de tension” (in order to well distinguish the linguistic system and the conceptual system, terms are written between quotation marks, e.g. “voltage relay” when concepts are written between chevrons starting in upper case, e.g. <Voltage threshold relay>).

## 5. An application

Ontoterminology was validated and is currently used in different industrial applications like multilingual Information Retrieval Systems and Specialized Encyclopaedias for knowledge capitalisation. One of them is the ASTECH Project.

ASTECH (Advanced Sustainable Technologies for Heating and Cooling Applications) is a FP6 European project whose main goal is to promote the usage of renewable energy technologies for heating and cooling purposes. It concerns solar thermal, heat pumps, biomass and thermal energy storage technologies for domestic and large-scale applications. The aim of this project is to share information between providers and users of renewable energy technologies in Europe.

A multilingual search engine relying on the ontoterminology principle is available for the registered partners. Documents and information can be posted and searched in 9 languages (Bulgarian, English, French, German, Greek, Hungarian, Italian, Polish and Spanish). The documents are not translated for reasons of time and cost.

The first stage of the project was to build the common language-independent ontology using the OCW (Ontology Craft Workbench) methodology and environment where a concept is defined by specific differentiation [Roche 2001]. In a second stage each partner defined its own terminology, it means the different terms, including terminological variations, associated to the shared concepts – taking into account a new language requires only to define the corresponding terms. Then each document was classified onto concepts by a linguistic analysis of its content – a concept, like a folder, contains all the relevant documents whatever its writing language. Looking for information is done in his/her mother language. The query is analysed as a document in order to identify the denoted concepts. The reply contains all the relevant documents whatever its writing language. The figure 2 presents an example of a user search. The user browses the document database through the ontology (the left pane of the user interface provides an interactive navigation into the ontology) and selects a concept – the ontology is displayed in the user's language. The result is an ordered list of texts whatever of the writing language – here the first documents are written in English whereas the query is done in French.

Using both linguistic relationships between terms, e.g. synonymy, and formal relationships between concepts, e.g. logical properties of the subsumption, allows the relevance of the search results to be improved. Noise (returned irrelevant documents) is reduced as well as silence (missing relevant documents). A semantic signature, i.e. a vector of concepts with coefficients of importance, is associated to each document when it is classified. In a same manner, a semantic signature is associated to the query. A distance calculation between vectors puts the results into order.

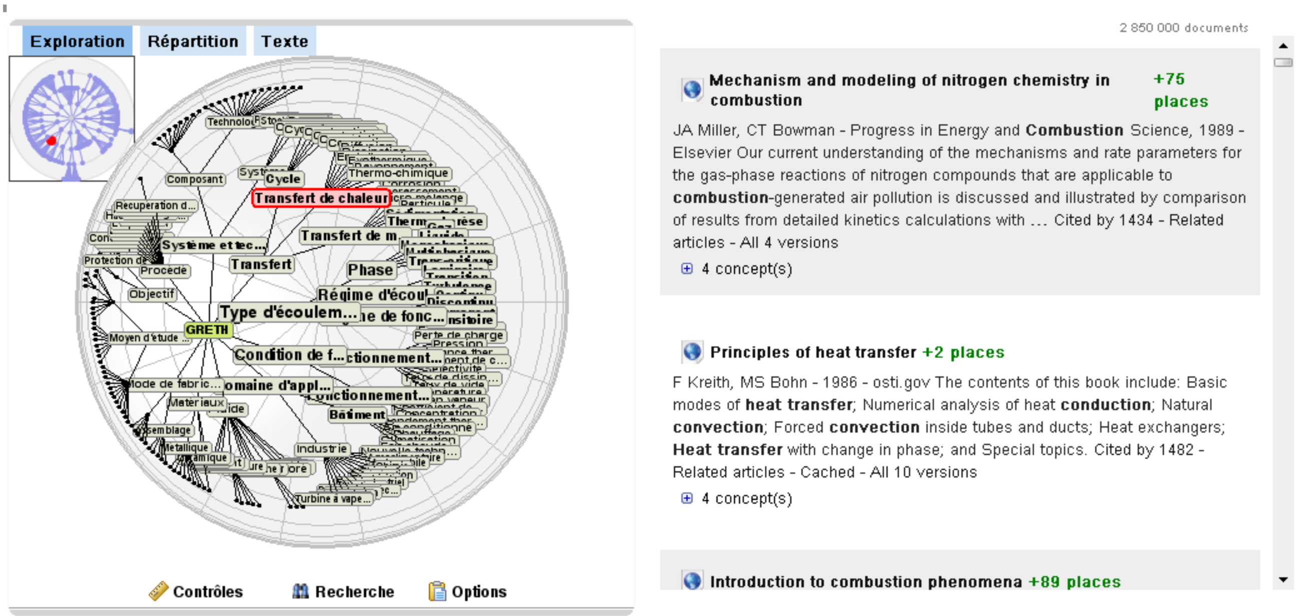


Figure 2: An ontology-oriented Content Management System

## 6. Conclusion

The need for a computational representation of the meaning of terms for IT application (semantic web, information retrieval, content management system, knowledge base system, etc.) raises new challenges for terminology. Although ontology appears to be one of the most interesting solutions, an ontology is not a terminology. Indeed, terms cannot be reduced to lexicalized concepts or to labels stuck onto concepts. Furthermore, a domain conceptualisation in terminology is more than a formal representation of concepts.

In order to take into account the two linguistic and conceptual dimensions of terminology, a new paradigm for terminology, called ontoterminology, has been introduced. An ontoterminology is a terminology whose conceptual system is a formal ontology relying on epistemological principles. Concepts and terms in ontoterminology exist for themselves. It means that terms are separated from concepts as well as term definitions written in natural language are separated from concept definitions written in a formal language. This makes it possible to manage the two linguistic and conceptual dimensions of terminology and provide two kinds of definition: the first formally defines the concept whereas the second explains the term and its usage from a linguistic point of view.

Ontoterminology was validated and is currently used in different industrial applications like multilingual Information Retrieval Systems and Specialized Encyclopaedias for knowledge capitalisation. Furthermore, ontoterminology allows new IT applications like knowledge mapping and browsing where ontology is viewed as a conceptual map in which experts navigate along the "is-a" and "part-of" relationships in order to access to information attached to the concepts.

## 7. References


Alexeeva, L.M. (2006). "Interaction between Terminology and Philosophy. Theoretical Foundations of Terminology Comparison between Eastern Europe and Western Countries". Würzburg: Ergon Verlag

Baader, F., Calvanese, D., McGuiness, D., Nardi, D., Patel-Schneider, P. (2003). "The Description Logic

Handbook". Franz Baader, ed. et al. Cambridge University Press, 2003

Budin, G. (2001). "A critical evaluation of the state-of-the-art of Terminology Theory". ITTF Journal,12. Vienna: TermNet

Buitelaar, P., Cimiano P., Magnini B. (2005). "Ontology Learning from Text: Methods, Evaluation and Applications". (Frontiers in Artificial Intelligence and Applications, Vol. 123). P. Buitelaar (Editor) Ios Press Publication (July 1, 2005)

Cabré T. "Theories in terminology". In Terminology 9:2 (2003), pp 163-199.

Daille B., Kageura K., Nakagawa H. and Chien L.F. (2004). "Recent Trends in Computational Terminology". (Special issue of Terminology 10:1). Benjamins publishing company.

Felber, H. (1984). "Terminology Manual". Unesco (United Nations Educational Scientific and Cultural Organization) – Infoterm (International Information Centre for Terminology).

Gomez-Perez, A., Corcho. O., Fernandez-Lopez, M. (2004). "Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web". Asuncion Gomez-Perez, Oscar Corcho, Mariano Fernandez-Lopez, Springer 2004

Grice, H.P (1957) "Meaning". Philosophical Review, n°66. pp 377-88

Gruber, T. (1993). "A Translation Approach to Portable Ontology Specifications". Knowledge Systems Laboratory September 1992 - Technical Report KSL 92-71 Revised April 1993. Appeared in Knowledge Acquisition, 5(2):199-220

Guarino, N., Carrara, M., and Giaretta, P. 1994a. "An Ontology of Meta-Level Categories". In J. Doyle, E. Sandewall and P. Torasso (eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR94). Morgan Kaufmann, San Mateo, CA: 270-280.

ISO 1087-1:2000. Terminology work-Vocabulary-Part1: Theory and application. International Organization for Standardization.

ISO 704:2000. Terminology work - Principles and methods. International Organization for Standardization.

Lyons, J. (1995). "Linguistic Semantics: An Introduction". Cambridge University Press.

Madsen, Bodil Nistrup & Hanne Erdman Thomsen (2008). "Terminological Principles Used for Ontologies." Managing ontologies and lexical resources. TKE 2008. Copenhagen: ISV.

Pavel, S. & Nolet, D. (2001). Handbook of Terminology. Minister of Public Works and Government Services Canada 2001, Catalogue No. S53-28/2001

Ogden C.K. & Richards I.A. (1923). "The meaning of meaning". New York : A Harvest Book, Harcourt, Brace & World, (reprint).

OWL Web Ontology Language: http://www.w3.org/TR/owl-features/

Roche C. (2001). "The "specific-difference" principle: a methodology for building consensual and coherent ontologies". IC-AI 2001, Las Vegas USA, June 25-28

Roche, C. (2003). "Ontology: a Survey". 8th Symposium on Automated Systems Based on Human Skill and Knowledge. IFAC, September 22-24, Göteborg

Roche, C. (2007). "Le terme et le concept : fondements d'une ontoterminologie". TOTh 2007 (Terminologie & Ontologie : Theories et applications), pp 1-22, Annecy, France, 1er juin 2007.

Roche, C. (2007). "Saying is not modelling". NLPCS 2007 (Natural Language Processing and Cognitive Science); pp 47 – 56. ICEIS 2007, Funchal, Portugal, June 2007.

Roche, C. (2008). "Should Terminology Principles be re-examined?". English version of the article: « Faut-il revisiter les Principes terminologiques ? ». TOTh 2008 (Terminology & Ontology : Theories and applications), pp 53-72, Annecy, France, 5 & 6 June 2008.

Roche, C., Calberg-Challot, M., Damas, L., Rouard. P. (2009). "Ontoterminology: A new paradigm for terminology". KEOD 2009. International Conference on Knowledge Engineering and Ontology Development, 5-8 October, Madeira (Portugal)

Sager, J. (1990). "A Practical Course in Terminology Processing". John Benjamins Publishing Company

Sapir, E. (1921). "Language : An Introduction to the study of speech". Docer Publications, 2004. (Originally published by Harcourt, Brace and Company, 1921)

Saussure F. (reprint 1998). "Course in General Linguistics". Open Court Publishing Co ,U.S.

Whorf, B.L. (1956). "Language, Thought and Reality". The MIT Press

Staab, S., Studer, R. (2004). "Handbook on Ontologies". Steffen Staab (Editor), Rudi Studer (Editor), Springer

Temmerman R. (2000). "Towards New Ways of Terminological Description. The Sociocognitive approach". Amsterdam/Philadelphia: John Benjamins.

Tricot C., Roche C. (2005). "Visual Information Exploration: A Return on Experience in Knowledge Base Management," in ICAI'05 - The 2005 International Conference on Artificial Intelligence. Las Vegas, Nevada, USA

Ushold, M., Gruninger, M. (1996). "Ontologies: Principles, Methods and Applications". Knowledge Engineering Review, Vol. 11, n° 2, June 1996. Also available from AIAI as AIAI-TR-191

Woods, W. (1975). "What's in a Link: Foundations for Semantic Networks". Representation and Understanding: Studies in Cognitive Science, 35-82, edited by D.G. Bobrow and A.M. Collins, New York: Academic Press

Wright, S.E., Budin, G. (1997). "Handbook of Terminology Management", volume 1 and 2. John Benjamins Publishing Company

Wright, J. M., Fox, M.S., Adam, D. (1984). "SRL/1.5 Users Manual." Technical report. Robotics Institute, Carnegie-Mellon University

Wüster, E. (1968). "The Machine Tool – An interlingual Dictionary of Basic Concepts". London: Technical Press