TOTh 2022

Terminologie & Ontologie : Théories et Applications

Terminologie & Ontologie: Théories et Applications

Actes de la conférence

TOTh 2022

Université Savoie Mont Blanc

2 & 3 juin 2022

Les ouvrages TOTh précédents sont disponibles:
- sur le site de l'Université Savoie Mont Blanc (btk.univ-smb.fr/livres)
- sur le site du Comptoir des Presses d'Universités (www.lcdpu.fr)
- ou auprès de: contact@toth.condillac.org

Terminologie & Ontologie: Théories et Applications

Actes de la conférence
TOTh 2022

Université Savoie Mont Blanc

2 & 3 juin 2022

http://toth.condillac.org

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

Comité scientifique

Président du Comité scientifique: Christophe Roche

Comité de pilotage

Rute Costa	Universidade Nova de Lisboa
Humbley John	Université Paris 7
Kockaert Hendrik	University of Leuven
Christophe Roche	Université Savoie Mont Blanc

Comité de programme 2022

Le comité de programme est constitué chaque année à partir du comité scientifique de TOTh en fonction des soumissions reçues. La composition du comité scientifique est accessible à l'adresse suivante: http://toth.condillac.org/committees

Amparo Alcina	Universitat Jaume I – Spain
Xiaomi An	Renmin University – China
Albina Auksoriute	Institute of the Lithuanian Language – Lithuania
Bruno Bachimont	Université Technologie de Compiègne – France
Jean-Paul Barthès	Université Technologie de Compiègne – France
Christopher Brewster	Maastricht University – Netherlands
Danielle Candel	CNRS, Université Paris Diderot – France
Sylviane Cardey	Université de Franche-Comté – France
Stéphane Chaudiron	Université de Lille 3 – France
Rute Costa	Universidade NOVA de Lisboa – Portugal
Eric De La Clergery	INRIA – France
Dardo De Vecchi	Kedge Business School – France
Thierry Declerck	DFKI – Germany
Valérie Delavigne	Université Paris 3 – France
Sylvie Desprès	Université Paris 13 – France
Juan Carlos Diaz Vasquez	EAFIT University – Colombia
Pamela Faber	Universidad de Granada – Spain
Christiane Fellbaum	Princeton University – USA
Cécile Frérot	Université Stendhal Grenoble 3 – France
Iolanda Galanes	Universidade de Vigo – Spain
Teodora Ghiviriga	Alexandru Ioan Cuza University – Romania
Rufus Gouws	University of Stellenbosch – South Africa
Frédérique Henry	ATILF – France
John Humbley	Université Paris 7 – France
Yangli Jia	University of Liaocheng – China
Kyo Kageura	University of Tokyo – Japan
Barbara Karsch	BIK Terminology – USA
Hendrik Kockaert	University of Leuven – Belgium
Héba Lecocq	Université Sorbonne nouvelle – France
Hélène Ledouble	Université de Toulon – France

Patrick Leroyer   Aarhus University – Denmark
Georg Löckinger   University of Applied Sciences Upper Austria – Austria
Rodolfo Maslias   TermCoord, European Parlement – Luxembourg
Candida Jaci de Sousa Melo   Universidade Federal do Rio Grande do Norte – Brazil
Fidelma Ní Ghallchobhair   Foras na Gaeilge, Irish-Language Body – Ireland
António Pareja Lora   Universidad Complutense de Madrid – Spain
Sandrine Peraldi   University College Dublin – Ireland
Silvia Piccini   Italian National Research Council – Italy
Suzanne Pinson   Université Paris Dauphine – France
Maria Pozzi   el colegio de méxico – Mexico
Bihua Qiu   China National Committee for Terms in Sciences and Technologies – China
Jean Quirion   Université d'Ottawa – Canada
Renato Reinau   Université de Genève – Switzerland
Christophe Roche   Université Savoie Mont Blanc – France
Mathieu Roche   CIRAD – France
Micaela Rossi   Università degli studi di Genova – Italy
Klaus-Dirk Schmitz   Cologne University – Germany
Marcus Spies   University of Munich – Germany
Frieda Steurs   University of Leuven – Belgium
Anne Theissen   Université de Strasbourg – France
Philippe Thoiron   Université Lyon 2 – France
Kara Warburton   City University of Hong Kong – China
Maria Teresa Zanola   Università Cattolica del Sacro Cuore – Italy

Avant-propos

La seizième édition de la conférence TOTh s'est tenue, selon une habitude bien établie, les jeudi et vendredi de la première semaine de juin. Pour la deuxième année consécutive, TOTh s'est déroulée à la fois en présentiel et à distance. Ainsi, nous avons pu renouer avec les plaisirs qu'offrent le présentiel et les moments de sociabilité comme le dîner de TOTh dans la ville historique de Chambéry, tout en permettant à un plus grand nombre de suivre et de participer à nos travaux.

Notre collègue Platon Pétridis, Professeur d'Archéologie Byzantine à l'Université d'Athènes, a ouvert la Conférence TOTh 2022 sur le thème de la classification et appellation des objets dans les contextes archéologiques et pour qui « une nouvelle approche en matière de terminologie et de taxinomie dans les sciences archéologiques s'avère donc indispensable». Un sujet proche de nos préoccupations portant sur l'organisation des objets et leur dénomination.

Quelques chiffres donneront une idée de cet événement initié en 2007.

– Quatorze communications ont été retenues pour publication sur les 17 présentées après une sélection rigoureuse par un comité de programme international issu de 24 pays. La diversité des sujets abordés, tant théoriques que pratiques, illustre la richesse et le dynamisme de notre discipline.

– Sur les 87 personnes enregistrées, plus de 50 ont suivi de manière assidue les présentations. 18 pays étaient représentés : Albanie, Allemagne, Argentine, Chine, Chypre, Congo, Espagne, Estonie, États-Unis, France, Grèce, Italie, Lettonie, Lituanie, Niger, Pays Bas, Portugal, Suisse.

Les actes sont publiés aux Presses Universitaires Savoie Mont Blanc. Ceux des années précédentes sont accessibles à partir du site de la conférence (http://toth.condillac.org/) et des Presses Universitaires Savoie

Mont Blanc (https://btk.univ-smb.fr/livres/?fwp_collections_revues=-terminologica).

Je terminerai en remerciant le Ministère de la Culture, et plus précisément la Délégation Générale à la Langue Française et aux Langues de France, l'Université Savoie Mont Blanc et l'École Polytech Annecy-Chambéry pour leur support et leur aide financière à l'organisation de la conférence et à la publication des actes.

<div align="right">

Christophe Roche
Président du Comité scientifique

</div>

Sommaire

Implementation of terms and their grammatical category in an ontology. The ONTODIC model

Amparo Alcina

Universitat Jaume I
alcina@uji.es, http://tecnolettra.uji.es

Abstract. Terminology dictionaries contain different linguistic information about the terms in their entries. Depending on the purpose and function of the dictionary, in each entry we will find the term, its grammatical description (such as grammatical category, gender and number), definition, synonyms and equivalences in one or more languages. Terminology databases present files with the same type of information. To create terminological dictionaries based on ontologies we have designed an ontological model, ONTODIC.
In this paper, we present the main aspects of the ontological model ONTODIC and explain how the grammatical information of the terms is represented in the ontology.

1. Introduction

Terminology dictionaries contain different linguistic information about the terms in their entries. Depending on the purpose and function of the dictionary, in each entry we will find the term, its grammatical description (such as grammatical category, gender and number), definition, synonyms and equivalences in one or more languages. Terminology databases present files with the same type of information.

Current ontological models focus on the representation of concepts, their relationships and the hierarchical representation of the gener-

ic-specific relationship. Terms are represented as labels that link to concepts in the hierarchy. These tags and the grammatical and other information associated with the terms appear as metadata or ontology annotations (Cimiano *et al.* 2020, Bosque-Gil *et al.* 2019, McCrae *et al.* 2017, Cimiano, McCrae y Buitelaar 2016).

From a linguistic approach, these ontological models present two types of problems. On the one hand, they cannot be considered terminological resources because the management of terms as labels does not allow them to participate in the inferential and reasoning processes. On the other hand, these models are based on the univocal relationship between concept and term.

Other models, such as the one presented by Schalley (2019), follow an organization of the data more in line with the objectives of linguistics to present, where appropriate, the linguistic typology. In our model, the ONTODIC model, we seek a model that allows the representation of terminological data according to the needs of linguists and dictionary users. In this model, both concepts and terms are represented as main elements of the ontology (classes, individuals, properties). For its development, we have based ourselves on the knowledge of the domain of Linguistics and we have followed the methodology of creating ontologies of Knowledge Engineering (Gómez-Pérez, Fernández López y Corcho 2004, Guarino y Welty 2004). For the implementation of the model tests we have used the Protégé ontology editor (Musen 2015).

In this paper, we present the main aspects of this ontological model and explain how the grammatical information of the terms is represented in the ontology.

2. An ontological model to represent dictionaries: ONTODIC model

Ontologies, in knowledge engineering (description logic), are applied to any field and have the objective of organizing objects (individuals) under concepts (in classes). So, we have organized a sample of terms including the ontological and linguistic dimensions following the

ontology creation methodology and using the appropriate tools, starting from a linguistic perspective.

In our ontology model, *words* or *terms* are the individuals that are the object of classification and linguistic concepts (whether grammatical, as a noun or verb, or morphological, as a full or derived form) constitute the classes into which the terms are classified. We have thus, from a linguistic approach:

1. terms as elements of a linguistic system, which are represented as 'individuals'
2. linguistic concepts or 'classes' under which the terms are classified
3. linguistic relationships of various kinds that exist between the terms in a linguistic system formalized as 'object properties'

In this work, we will present this configuration by applying it to the terminology of the Spanish ceramics industry and analyse its grammatical aspects, with examples of the representation of Spanish terminology. We will examine the peculiarities of this configuration of elements in comparison to other configurations and in comparison to the methodology habitually used in knowledge engineering.

3. Grammatical description of terms

In dictionaries and terminology databases, terms are described with information about their part of speech. It indicates, for example, whether they are verbs, nouns, adjectives, adverbs, the grammatical number (singular or plural) and their grammatical gender (feminine or masculine). Thus, for example, the term esmalte is characterized by being a noun, masculine, singular; or the term esmaltar because it is a transitive verb.

The different linguistic concepts that we use to describe the terms grammatically have been represented as classes in the ontology. Thus, they constitute ontology classes: Noun, Verb, Singular, Plural, Feminine or Masculine, among others. It can be seen in Figure 1.

Implementation of terms and their grammatical category in an ontology.
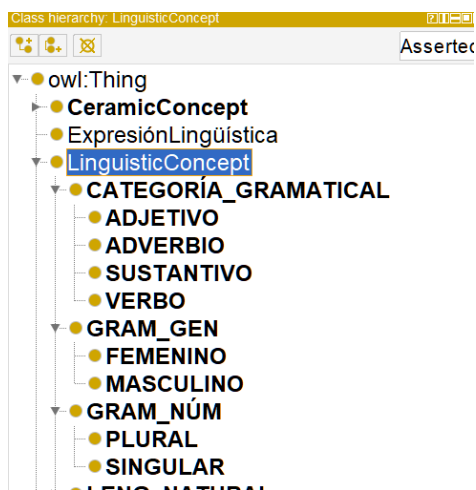The ONTODIC model

Fig. 1 – Grammatical concepts in the ontology

In ONTODIC, each term is represented as an individual and is linked to the class or classes that adequately describe them. Thus, for example, the terms *horno* 'oven', *tolva* 'hopper', and *rugosímetro* 'profilometer' are instances of the Noun class. Furthermore, *oven* and *profilometer* are instances of the Masculine and Singular classes, and the term *tolva* 'hopper' is an instance of the Feminine and Singular classes.

The implementation of grammatical concepts as classes in the ontology follows from the application of general linguistic theory. Thus, nouns constitute the set or class of words, or terms, with the capacity to function as the nucleus of a noun phrase in language and texts. Thus, the noun class can be formalized in the ontology as:

Class	Object property	Class
Sustantivo 'Noun'	funcionaComoNucleoDe 'worksAsHeadOf'	Sintagma nominal 'Nominal phrase'

Where NOUN and NOMINAL PHRASE are classes and 'worksAsHeadOf' is a property that links both classes.
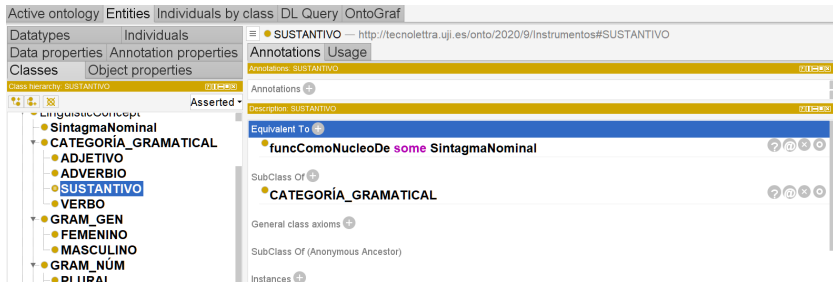
Fig. 2 – Description of the class Sustantivo ('Noun')

In Linguistics, the analysis of the behavior of the terms in the texts leads us to conclude to which class they belong. For example, in the following contexts (contexts 1 and 2), extracted from our corpus of ceramic texts, TXTCeram, we observe that the term *tolva* 'hopper' functions as the nucleus of the noun phrase *la tolva* 'the hopper'.

(1) La alimentación del horno, se lleva a cabo desde la tolva mediante un tornillo sinfín. (CE021-0e)

'Furnace feeding is carried out from the hopper by means of an endless screw.'

(2) El problema es particularmente frecuente en las producciones de gres porcelánico donde la segregación granulométrica en la tolva de la prensa provoca el llenado de los alvéolos exteriores con material de granulometría más grande. (CE027-5e)

'The problem is particularly frequent in porcelain stoneware productions where the granulometric segregation in the press hopper causes the filling of the external alveoli with material with a larger granulometry.'

The fact that the terms fulfill the properties associated with a linguistic category is what makes them part of that linguistic category. In ontology, we therefore represent linguistic categories as classes, which

are described by object properties. Language elements that satisfy these properties are classified as instances of these classes.

This configuration has the consequence that the individuals of this ontology will be instances not of one but of several classes. In the figure Figure 3, we observe that the term tolva 'hopper', represented as an individual, is an instance of the SUSTANTIVO 'NOUN', FEMENINO 'FEMININE', SINGULAR 'SINGULAR' classes (among others). This gives rise to a *multidimensional* ontology, in which the representation of disjoint classes almost never (if not never) has a place.
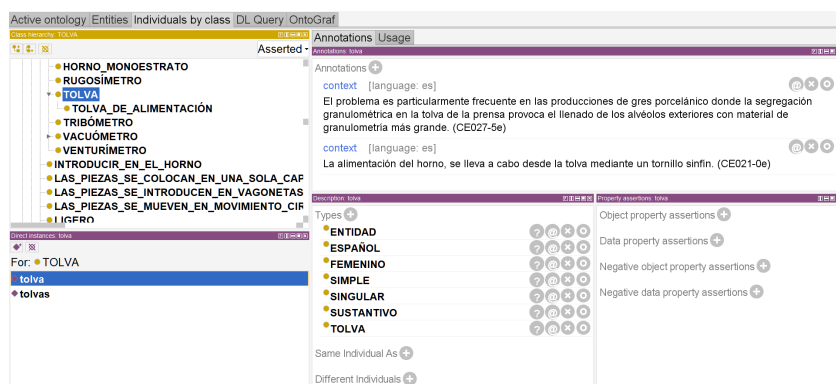

Fig. 3 – Representation of tolva 'hopper'
as instance of different classes

4. Conclusions and future work

In this work we have shown a useful ontological model to formalize the lexicon of natural language and that, in turn, allows the integration between the ontological and linguistic dimensions.

The model that we have presented overcomes the problems that linguists attribute to ontologies, which can facilitate their use by terminologists and translators, which will contribute to further development of linguistic ontologies and their higher quality in the semantic web.

In the future, we hope to develop different modules of the model, in addition to the grammatical aspects, for instance, morphological or how to represent collocations of terms.

We would like this work, developed from a linguistic approach and using only the technology of the Protégé editor, to benefit from the contributions and collaboration of experts in ontology, knowledge engineering and Semantic Web.

References

Bosque-Gil, J. *et al*. 2019. The OntoLex Lemon Lexicography Module. Final Community Group Report 17 September 2019. Edited by Julia Bosque-Gil and Jorge Gracia.

Cimiano, P. *et al*. 2020. *Linguistic Linked Data. Representation, Generation and Applications*.

Cimiano, P., J. McCrae y P. Buitelaar. 2016. Lexicon Model for Ontologies: Final Community Group Report 10 May 2016. W3C Ontology Lexicon (Ontolex) Community Group.

Gómez-Pérez, A., M. Fernández López y O. Corcho. 2004. *Ontological Engineering*. New York: Springer.

Guarino, N. y C. A. Welty. 2004. "An Overview of OntoClean." In *Handbook on Ontologies*, edited by Steffen Staab and Rudi Studer. Berlin: Springer.

McCrae, J. *et al*. 2017. "The OntoLex-Lemon Model: Development and Applications." In *eLex 2017. Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, edited by Iztok Kosem, Carole Tiberius, Milos Jakubícek, Jelena Kallas, Simon Krek and Vít Baisa, 587-597. Leiden: Lexical Computing CZ.

Musen, M. A. 2015. "The Protégé project: A look back and a look forward." *AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence* 1 (4). doi: 10.1145/2557001.25757003.

Schalley, A. C. 2019. "Ontologies and ontological methods in linguistics." *Language and Linguistics Compass* 13 (11). doi: 10.1111/lnc3.12356.

Acknowledgement