

# Using a Linguistic Approach to Represent Terminology in an Ontology

Amparo Alcina<sup>1</sup>

<sup>1</sup> University Jaume I, Avda. Sos Baynat s/n, Castelló, 12071, Spain

## Abstract

Abstract text. Ontology models of lexical and terminological resources need to include the linguistic dimension in a more coherent and adequate way. In general, ontologies describe objects, but the linguistic approach consist of describe terms as words, not as codes or labels. In this work, we will present our ontology model, where *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. This model of lexical representation can enhance the use of ontologies to make dictionaries and contribute to the Web Semantic and Linked Data Linguistics.

## Keywords

Terminology, dictionaries, ontology, natural language processing

## 1. Introduction

The increased interest in having terminological resources based on the OWL-RDF language of the Semantic Web by experts in natural language processing or knowledge engineering contrasts with the fact that ontologies are not considered adequate repositories as lexical or terminological resources according to some authors [1-3].

On the one hand, these ontologies are valuable because, like lexical networks, they organize terminology according to semantic or conceptual relationships. This is evident in the case of termologists who work according to ISO principles and standards (ISO 1087 and 704). In the same case, we have the European projects that have developed models to incorporate linguistic information into ontologies (such as OntoLex, Lemon) and the thematic networks and communities, such as ELEXIS, NEXULINGUARUM and TOTh, which promote the work of lexicon and terminology based on lexical networks. and ontologies.

On the other hand, these initiatives encounter the problem of representing the lexicon and terminology in such a way that the onomasiological perspective (conceptual representation) and the semasiological perspective (representation of linguistic data) are adequately combined. These shortcomings dissuade many lexicographers and terminologists from using ontologies in the development of their resources.

In our opinion, if the formalization of the linguistic dimension and the conceptual dimension in ontologies are properly combined, their use in the field of linguistics could increase considerably, which would benefit the Semantic Web and Linked Data Paradigm. Therefore, the challenge is to represent the linguistic dimension of natural language terminology in an ontology, avoiding the problems it currently suffers from.

In the first part, we show the reasons why lexical networks or ontologies do not satisfy some of the expression needs of the linguistic dimension of terms until now. In the second part, we present our proposal to extend the ontological model with the linguistic dimension. In contrast to other models, our

---

*1st International Conference on "Multilingual digital terminology today. Design, representation formats and management systems", June 16 – 17, Padova, Italy*

EMAIL: [alcina@uji.es](mailto:alcina@uji.es)

ORCID: 0000-0002-4931-564X



© 2022 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

model allows to formalize the linguistic data as part of the ontology, not of the metadata, which avoids the weaknesses identified by linguists.

For the development of our proposal, we have followed the ontology creation methodology of Knowledge Engineering [4, 5]. For the implementation of the model and the data we have used the Protégé ontology editor [6].

## 2. Objects vs. Words Formalisation

A number of authors concluded that ontologies does not meet all the needed requirements to organize words. The main problems of using lexical networks or linguistic ontologies (such as WordNet, Cyc, SIMPLE) to represent the lexicon have been pointed out by Hirst [1] and other authors [2-8]. In short, the main difficulties presented by these models appear to be:

1. Deficiencies in representing the meaning of the lexicon
2. No correspondence between the ontological categories in an ontology and the semantic categories of the words
3. No correspondence between categories necessary for classifying the objects of the world and categories necessary for classifying the lexicon
4. No correspondence between the organization of words according to semantic categories, and the organization of words according to morphological, grammatical or other linguistic categories

When we analyze these problems, we realize that the real aim of ontologies, including lexical networks and so-called linguistic ontologies, is to structure the entities of the world. They follow a philosophical conception, whose aim is to describe objects. In these ontologies, terms are considered “codes” or “labels” which refer to these concepts (classes or set of objects), following an approach base on the referential semantics. That is evident in the following aspects:

1. the identification of concept and word: each concept in ontology is intended to describe real world entities (e.g. aeroplanes in aeronautics, plants in botany).
2. the exclusively referential (non-semantic) description of concepts. In these ontologies, there is no description or formalisation of the meaning of the concept, apart from the specification of a more general concept. Instead, they allude to a semantics by reference: concept is given by the set of the objects of the world to which it refers.
3. the different levels of language description, such as phonological, morphological, grammatical, syntactic, etc., are not represented. Some models, such as SKOS or Lemon, propose the insertion of this data on the periphery of the ontology, as tags in the metadata.
4. the aspiration to a universal ontology. Some ontologies seek/proclaim the universality of ontological representation, which they also extend to a linguistic universality, which is a contradiction. The existence of words is inextricably linked to the linguistic system in which they develop and are used.

In Linguistics, however, the terms are not considered mere codes, but elements that interact with each other and with other elements within a complex system, the linguistic system. This complex system does not fit under the semantics by reference approach. A more complex model capable of modeling all the dimensions of the terms in natural language, not just the conceptual dimension, is needed.

## 3. An ontological model to represent dictionaries: ONTODIC model

Ontologies, in knowledge engineering (descriptive 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 morphological, grammatical and semantic aspects, with examples of the representation of cases of polysemy and synonymy and the representation of equivalences in other languages. 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.

The linguistic approach we propose solves the main problems of using ontologies to represent natural language. This ontological model is based on the following principles:

1. it represents the lexicon and its conceptualization using the main elements of ontology (such as classes, relations and individuals, axioms, etc.), not tags or metadata
2. it formalizes the differences between form and meaning, allowing the representation of asymmetric relations such as polysemy or synonymy
3. it allows the different levels of language (phonology, morphology, grammar, semantics) to be represented, with the necessary and sufficient level of detail
4. it assumes that all languages have their own formal and conceptual configuration, and allows each of them to be represented independently, rather than assuming a universal or supposedly objective representation.

We believe that this model of lexical representation will enhance the use of ontologies by linguists (lexicographers, translators and linguistics researchers) to make dictionaries, which will contribute to the increased availability of high-quality lexical resources in the OWL-RDF format on the Semantic Web.

## 4. Acknowledgements

I would like to thank the reviewers of a previous version of this paper for their valuable comments.

This research is part of the project “PRO-ONTODIC: Protocolos para la creación de diccionarios terminológicos basados en ontologías (Modelo ONTODIC)” Ref. UJI-B2018-65, funded by the University Jaume I of Castellón, Spain.

## 5. References

- [1] G. Hirst, *Ontology and the Lexicon*, in: S. Staab, R. Studer (Eds.) *Handbook on Ontologies*, Springer, Berlin, 2004, pp. 209-229.
- [2] P. Edmonds, G. Hirst, *Near-Synonymy and Lexical Choice*, *Computational Linguistics* 2 (2002) 105-144.
- [3] P. Cimiano, E. Montiel-Ponsoda, P. Buitelaar, M. Espinoza, A. Gómez-Pérez, *A note on ontology localization*, *Applied Ontology* 2 (2010) 127-137.
- [4] E. Montiel-Ponsoda, *Multilingualism in Ontologies. Multilingual Lexico-Syntactic Patterns for Ontology Modeling and Linguistic Information Repository for Ontology Localization*, Ph.D. thesis, Universidad Politécnica de Madrid, Madrid, 2011.
- [5] M.-C. L'Homme, G. Bernier-Colborne, *Terms as labels for concepts, terms as lexical units: A comparative analysis in ontologies and specialized dictionaries*, *Applied Ontology* 4 (2012) 387-400.
- [6] P. Buitelaar, P. Cimiano, P. Haase, M. Sintek, *Towards Linguistically Grounded Ontologies*, in: L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, E. Simperl (Eds.) *The Semantic Web: Research and Applications*, Springer Berlin Heidelberg, 2009, pp. 111-125.
- [7] P. Cimiano, J. McCrae, P. Buitelaar, *Lexicon Model for Ontologies: Final Community Group Report* 10 May 2016, 2016. <https://www.w3.org/2016/05/ontolex/>.
- [8] A. Parvizi, M. Kohl, M. González, R. Saurí, *Towards a Linguistic Ontology with an Emphasis on Reasoning and Knowledge Reuse*, in: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC16)*, European Language Resources Association (ELRA), Portorož, Slovenia, 2016, pp. 441-448.