Onomasiological dictionaries and ontologies

Patrícia Cunha França
University of Minho, Portugal

There has been an increasing interest in ontologies over the last years by the computer science. This interest can be attributed to the advent of what has been known as Semantic Web. Consequently, a persistent interest in Onomasiological Lexicography has arisen and several authors have written about the relation between thesaurus and ontologies, determined to build bridges between the two. The aim of this paper is to light up the discussion between the similarities and differences between onomasiological dictionaries and ontologies. It presents some definitions for concepts such as onomasiology, onomasiological dictionaries, ontologies, formal ontologies, linguistic ontologies. It also intends to demonstrate that the differences pointed out by some authors to distinguish onomasiological dictionaries from ontologies are not quite evident.

1. Introduction

> *Among the wide Spectrum of information representation and retrieval tools are thesauruses and ontologies, which are the most often linked in bibliography, even though they come from very different disciplinary areas.*
> Arano, 2005

There has been an increasing interest in ontologies over the last years by the computer science. This interest can be attributed to the advent of what has been known as Semantic Web[1]. Consequently, a persistent interest in Onomasiological Lexicography has arisen. It is recurrent to observe a mention to dictionaries, mainly thesaurus, when reading about ontologies in recent research. Several authors, particularly Arano (2005), Moreira, Alvarenga & Oliveira (2004), Hirst (2004), Oltramari & Vetere (2008) and Wielinga *et al*. (2001), have written about the relation between thesaurus and ontologies, determined to build bridges between the two representational instruments.

The aim of this paper is to light up the discussion between the similarities and differences between onomasiological dictionaries and ontologies. We will try to demonstrate that the differences pointed out by some authors are not quite evident.

The article is organized as follows. Section 2. is an attempt to clarify the concepts of onomasiology, onomasiological dictionary, ontology, lexical database and linguistic ontology. Section 3. presents several objections to the arguments sustained by some authors to distinguish between dictionaries and ontologies. Section 4. is an attempt to refute some of the arguments provided in section 3. Finally, in section 5. the concluding remarks will be presented.

2. Defining concepts: from onomasiology to ontology

Grzega and Schöner define onomasiology as the field, within Lexicology, that aims 'to find the linguistic forms, or the words, that can stand for a given concept/idea/object' (Grzega; Schöner 2007: 7). An onomasiological dictionary has been traditionally defined as a dictionary where words are grouped by semantic fields. While in a semasiological dictionary the entries appear, in general, in alphabetic order, in an onomasiological dictionary, words are usually grouped by lexical fields. In an onomasiological dictionary an entry represents a

[1] http://www.w3.org/

concept, an idea. This particular type of lexicographic product has been classified under different terms. For instance, Haensch (1982) groups, under the same concept, the ideological dictionary, the analogical dictionary, the pictorial dictionary, the dictionary ordered by subjects and the synonym dictionary. Similarly, Martínez de Sousa (1995) suggests that an onomasiological dictionary can be set in the same concept as an ideological dictionary, a conceptual dictionary, an analogical dictionary or a codifying dictionary. Béjoint, on the one hand, uses the terms 'thesaurus' and 'lexicon', but, on the other hand, also uses the terms '*conceptual* (or *ideological*, *notional*, *onomasiological*, *semantic*) *dictionary*' (Béjoint 2004: 15). One paradigmatic and uncontroversial example of an onomasiological dictionary is the famous *Thesaurus* of Peter Mark Roget. Another well-known example is the *Begriffssystem* de Hallig and Wartburg (1962).

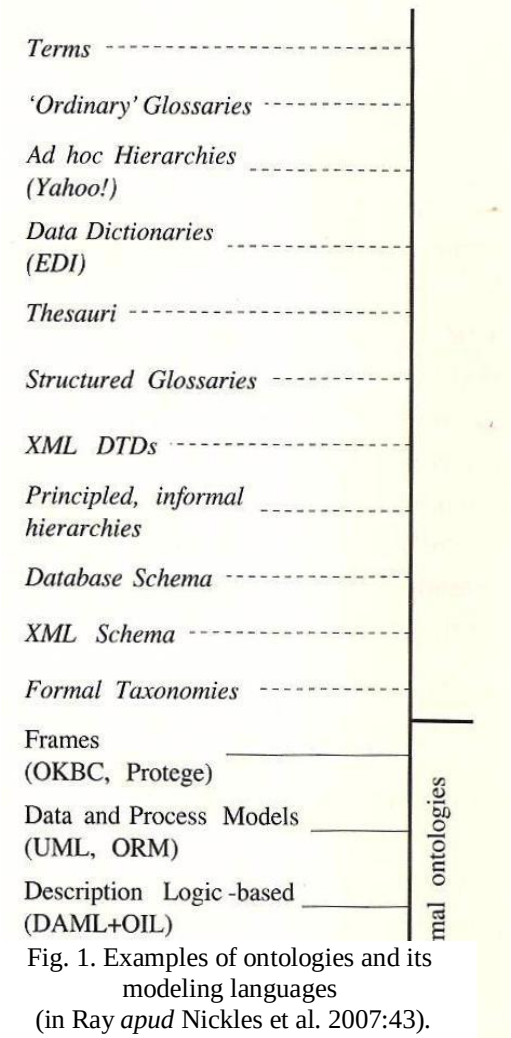

Fig. 1. Examples of ontologies and its modeling languages (in Ray *apud* Nickles et al. 2007:43).

The most widely cited definition of ontology in recent research comes from the Computational Science. It is given by Gruber (Gruber 1993), who defines an ontology as 'an explicit specification of a conceptualization'. It is also accepted that the definition of what exactly is ontology has not been attained consensually[2]. Ontology, in Computational Science, can be by and large defined as a computational artifact designed for a specific purpose, in which terms, linked together in different ways, represent instances in the world[3]. A linguistic ontology can be defined as a particular kind of ontology where terms, also linked together, are seen as linguistic objects and not exactly instances in reality. Although there is a general controversy in considering lexical databases ontologies in a strict sense, authors like Magnini and Speranza claim that they are a 'particular kind of ontology' (Magnini; Speranza 2003: 43). The most well-known linguistic ontology is WordNet[4].

To make the issue less controversial it might be important to distinguish, initially, between a formal and a non-formal ontology. Within the Computational Science, ontology is defined by the language used to build it. Some of these languages are considered more formal than others. For instance, Fig. 1 shows an illustration of the different degree of formalism used by different computational artifacts. As it can be seen, *thesauri*

[2] For a detailed discussion about the problems of defining the concept see Guarino (1998), Guarino & Giaretta (1995) and Smith (2004). Here we will consider that formal ontology is a computational artifact designed for a certain purpose and for a specific context; it is made by terms that represent things in the world.

[3] A further discussion about the essence of this world – whether it is the world of our discourse or the world in a realistic sense - can be seen in França (2009), Grenon (2008) and Smith (2008).

[4] http://wordnet.princeton.edu/wordnet/

are on the top of the image, they are not considered formal ontologies, whereas data models (represented in UML), for instance, are considered formal ontologies. Two examples of well-known formal ontologies are SUMO[5] and BFO[6]; both use OWL[7] as their language.

Nevertheless, there are important theoretical issues to be discussed when defining ontology. Some important questions, that do not concern the language in which ontologies are written, need to be taken into consideration. As mention above, considering a computational artifact, like WordNet, an ontology might not only depend on the language in which these artifacts are built. A more careful look at the content is just as important, if not more important, as the way by which this content is presented. In the next lines some of these theoretical issues will be discussed.

3. On the differences pointed out between dictionaries and ontologies

At it has been stated above, there have been several attempts to establish a relation between dictionaries and ontologies. As seen from the examples given in Fig. 1, dictionaries can be considered simple ontologies. One of the most obvious similarities between the two is that both use terms and relations. Hierarchical relations can be found in both onomasiological dictionaries and ontologies. In fact, some hierarchical relations can also be found in a semasiological dictionary, even though they are implicit. Hirst, for example, states that a semasiological dictionary includes an implicit ontology or, at least, a semantic hierarchy, when establishing basic Aristotelian definitions:

> For example, if *automobile* is defined as a *self-propelled passenger vehicle that usually has four wheels and an internal-combustion engine*, then it is implied that *automobile* is a hyponym of vehicle and even that automobile IS-A VEHICLE; semantic or ontological part-whole relations are also implied.
> (Hirst 2004: 223)

In spite of these resemblances, some authors, like Nickles et al. (2007), Smith (2004) and Hirst (2004), argue that there are fundamental differences to point out to ontologies and dictionaries. These differences can be summarized in four key points.

First, it is argued that ontologies use formal language, while dictionaries use natural language. In fact, one of the main characteristics of formal ontologies is to use a language that can be processed by machines, while dictionaries are intended to be read by people. As Nickles *et al.* indicate, '[n]o machine is currently capable of understanding a dictionary in any realistic sense of the word 'understanding'' (Nickles et al. 2007: 45).

Second, the dictionary is descriptive, it provides definitions of words and notes on how those words are used in a specific time and place. A formal ontology is prescriptive and normative; it determines, specifically, in a formal language, what a given term means. This is, in fact, one of the paradigmatic characteristics of ontologies. There are authors that sustain that a term in an ontology must mean exactly one and only one thing. There is no place in ontologies for ambiguities (Nickles et al. 2007: 45; Smith 2006).

[5] http://www.ontologyportal.org

[6] http://www.ifomis.org/bfo

[7] OWL is the most recommended knowledge representation language for building formal ontologies. It is recommended by the WWW Consortium and is based on Description Logic. For a comprehensive study of OWL see Lacy 2005.

Third, authors maintain that a term in an ontology is not a word but a concept. If it is true that terms in an ontology may receive names, which correspond to words or combination of words, so they can be understood by humans, notwithstanding, those names might well be replaced by arbitrary codes without losing their formal proprieties[8]. Ontological relationships are also different from those in a dictionary.

Finally, there is still another argument that sustains that ontology is different from an onomasiological dictionary. Hirst, for instance, defends that ontology deals with objects in the real world, it is 'a set of categories of objects or ideas in the world, along with certain relationships among them; it is not a linguistic object' (Hirst 2004: 8). As it was stated above, it is assumed that ontology represents instances in the real world. For this reason, it is argued that ontology can not be considered a linguistic object. Ontological relations are, thus, fundamentally different from lexical relations (Hirst 2004: 8).

4. Dictionaries and ontologies are not quite different

Some objection can be found when reading the arguments exposed above. First of all, it is important to start from the assumption that an onomasiological dictionary fits the category of dictionary. As it was demonstrated in Part 2 of this article, an onomasiological dictionary is a peculiar type of dictionary, but it is still a dictionary. If we do not assume that, the following discussion would be worthless.

In what refers to the first argument, if it is true that ontologies are built with formal languages, in order to be processed by computers, it is also true that ontologies are to be read by people. In fact, the great challenge when thinking of modeling languages used in ontologies is precisely this: how can we obtain the expressive power to describe content that can be processed by machines and, at the same time, allow it to be understood by humans? This is exactly what Lacy defends when he refers to Owl (Lacy 2005: 43):

> developers of Owl wanted to make the language intuitive for humans and to have sufficient power to describe machine-readable content needed to support Semantic Web applications. (Lacy 2005: 43).

In what regards the second argument, if theoretically a dictionary is based on a descriptive nature we cannot escape from pointing it a prescriptive character. In theoretical terms, a dictionary describes the language used by speakers at a given time and in a specific place, but it might be important to pay attention to the words of Jonathan Green, referring to Samuel Johnson and Noah Webster, the famous lexicographers of the 18th century:

> What both men were doing, although neither articulated it as such, was playing God. Or if not God, then at least Moses, descending from Sinai with the tablets of the law. For them their role was not simply to select a word list, define it, and make it available to the reading public; in addition they took on the priestly task of revealing a truth, in this case a linguistic one. (Green 1996: 5).

As for the third argument, which states that ontologies can be distinguished from the dictionaries for dealing with concepts, it must be said that this is not quite exact. Onomasiological dictionaries are known for dealing with concepts, unlike semasiological dictionaries, although there are important issues to be clarified regarding the definition of the

[8] As Nickles *et al.* point out; the linguistic items used in formal ontologies are not formal linguistic signs in its full sense, with form and content, but only strings of bytes (Nickles et al. 2007: 32).

word 'concept'[9]. But there is another objection to be made to this argument. Although an ontology is made to be processed by computers, it must remain intelligible for humans, so it can be used. Nickles et al. state that one of the challenges we are facing today in studying language and ontologies is establishing a satisfactory link between formal ontologies and linguistic expressions (Nickles et al. 2007: 44).

With this last argument we can also make an objection to Hirst. Although there are no objections to the assumption that a semasiological dictionary consists of words, understood as linguistic objects, as opposed to an ontology, where words are taken as representative items of the instances in a real world, the same can not be said of an onomasiological dictionary. Onomasiological dictionaries focus on concepts, not on the linguistic content. In the same way, it is not correct to assume that ontology deals with concepts rather than with linguistic signs. Linguistic ontologies, like WordNet, deal specifically with linguistic signs.

Johansen presents an interesting metaphor that intends to share some light to the old discussion between words, concepts and reality. Johansson argues that the Optical Metaphor is a good evidence to understand the distinction between Use and Mention (Johansson 2008: 300). The author suggests that when we are looking *at* language we are aware of the linguistic entities as such, while when we are looking *through* language, we are not quite aware of the linguistic entities:

> When, for example, one is conveying or receiving information in a language in which one is able to make and understand language acts spontaneously, one is not looking at the terms, concepts […] in question […]. Rather, one looks through these linguistic entities in order to see the information (facts, reality, or objects) in question. We are looking at linguistic entities, in contrast, when for example we are reading dictionaries and terminologies. (Johansson 2008: 300).

Johansson (2008: 303) states that WordNet can hardly be considered an ontology because it looks *at* language and not *through* it:

> All ontologies in information science contain terms. […] the experts in the various specialized domains of knowledge generally look through the terms. However, an ontology such as WordNet presents a special case, for (if it is to be called an ontology at all) it is an ontology of terms and meaning; it is like a dictionary, not like a taxonomical textbook. […].
> It is clear that the term 'cat' is mentioned and not used in WordNet. Both the scare quotes around the term 'cat' and the fact that it is preceded by the term 'noun' makes it clear that WordNet contains no talk of real cats. (Johansson 2008: 303)

As it was referred above, an ontology is, by and large, defined as a computational artefact where terms represent instances in the world. Johanssons' arguement might be a good solution to distinguishing between lexical databases (that look *at* language) and ontologies (that look *through* language), but, just as a common language dictionary can hardly presents only linguistic information (in the sense provided by Johansson), the same occurs in a lexical database or in a linguistic ontology.

If we carefully examine WordNet we will find that some of the relations provided are ontological relations. Although, in a first query, the entries are group grammatically, the relations provided can hardly be considered purely lexical. Consider the word 'cat' as an example. WordNet provides two grammatical categories for the term: noun and verb. Notwithstanding, the first definition provided is 'feline mammal usually having thick soft fur

[9] For a detailed discussion about the term 'concept' and its use in Linguistics see França (forthcoming).

and no ability to roar: domestic cats; wildcats'[10]. Moreover, WordNet makes a clear distinction between synset (semantic) relations (under 'S') and word (lexical) relations (under 'W'). All relations under the entry 'cat' are synset relations and not lexical relations. It will be interest to question about what kind of information is being provided with the expression 'usually having thick soft fur and no ability to roar'.

5. Concluding remarks

As has been shown, there are fundamental similarities between onomasiological dictionaries and ontologies. The question that immediately arises from the discussion above is how we can classify information that one receives when looking through language. It might be consider ontological information and/or semantic information. The answer to this question goes beyond the aim of this article. It is, nonetheless, an important question to be answer by an ontologist.

To consider an ontology a linguistic object might not depend on the nature of the ontology itself, but on the purpose for which it is built, how the terms are defined and how the user will use the ontology. It is also important to distinguish between different stages of the process of building an ontology. As Johansson sustains,

> All ontologies in information science contain terms. The builders of such ontologies look mainly *at* the terms in question, whereas the users of ontologies look mainly *through* them. Like the users, the experts in the various specialized domains of knowledge generally look through the terms. (Johansson 2008: 303).

Furthermore, it could be of great interest a more protracted analysis of some critics made to onomasiological dictionaries at the time they were published. It may prove crucial to the development of modern ontologies. The conclusions drawn can also be of great interest to theoretical Lexicography.

[10] http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o7=&o5=&o1=1&o6=&o4=&o3=&s=cat&i=1&h=100000000000#c

Bibliography

Arano, S. (2005). *Thesaurus and ontologies*. [on-line publication]. http://www.hipertext.net/english/pag1009.html [access date: 17 September 2009].

Béjoint, H. (2004). *Modern lexicography: an introduction*. Oxford: University Press.

França, P. (2009). 'Conceitos, classes e/ou universais: com o que é que se constrói uma ontologia?'. In *Linguamática*, nº 1, Maio, 2009. 105-122. ISSN: 1647-0818. [also available on-line]. http://www.linguamatica.com/index.php/linguamatica [access date: 31October 2009].

França, P. (forthcoming). *Ontologia e ontologias: contributos teóricos para uma perspectiva transdiciplinar*. [Master thesis]. Braga: University of Minho (ILCH).

Green, J. (1996). *Chasing the Sun: dictionary makers and the dictionaries they made*. New York: Henry Holt and Company, Inc.

Grenon, P. (2008). 'A Primer on Knowledge Representation and Ontological Engineering'. In Munn & Smith (eds.) *Applied Ontology*. Frankfurt: Ontos/Verlag. 57-82.

Gruber, T. (1993). *Towards principles for the design of ontologies used for knowledge sharing*. [on-line publication]. http://www.itee.uq.edu.au/~infs3101/ Readings/OntoEng.pdf [access date: 21 October 2009].

Grzega, J.; Schöner, M. (2007). *English and General Historical Lexicology*. [on-line publication]. http://www1.ku-eichstaett.de/SLF/EngluVglSW/OnOnMon1.pdf [access date: 31 October 2009].

Guarino, N. (1998). 'Formal Ontology and Information Systems'. [on-line publication]. http://www.loa-cnr.it/Papers/FOIS98.pdf [access date: 31 October 2009].

Guarino, N. (1998). 'Formal Ontology and Information Systems'. [on-line publication]. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.29.1776 [access date: 17 February 2010].

Guarino, N.; Giaretta, P. (1995). 'Ontologies and Knowledge Bases. Towards Terminological Clarification'. [on-line publication]. http://www.loacnr.it/Papers/KBKS95.pdf [access date: 17 September 2008].

Haensch, G. (1982). 'Tipología de las obras lexicográficas'. In Haensch, G.; Wolf, L.; Ettinger, S.; Werner, R. *La lexicografía. De la lingüística teórica a la lexicografía práctica.* Madrid: Gredos. 95-187.

Hallig, R.; Wartburg, W. (1963). *Begriffssystem als Grundlage für die Lexikographie / Système Raisonné des Concepts pour Servir de Base à la Lexicographie*. Berlin: Akademie-Verlag.

Hirst, G. (2004). 'Ontology and the Lexicon'. In Staab and Suder (eds.). *Handbook on ontologies*. Berlin: Springer. 209-230.

Johansson, I. (2008). 'Bioinformatics and Biological Reality'. In Munn, K.; B. Smith (eds.). *Applied Ontology*. Frankfurt: Ontos/Verlag.

Lacy, L. (2005). *Owl: Representing Information Using the Web Ontology Language*. Victoria: Trafford.

Magnini, B.; Speranza, M. (2002). 'Merging Global and Specialized Linguistic Ontologies'. In *Proceedings of the Workshop Ontolex-2002 Ontologies and Lexical Knowledge Bases*, LREC-2002. 43-48. [on-line publication]. http://multiwordnet.fbk.eu/paper/ontomergeontolex.pdf [access date: 20 September 2009].

Martínez de Sousa, J. (1995). *Diccionario de Lexicografia Práctica*. Barcelona: Vox.

Moreira, A.; Alvarenga, L.; Oliveira, O. (2004). 'O nível do conhecimento e os instrumentos de representação: tesuarus e ontologias'. In data GramaZero – Revista de Ciência da Informação – vol. 5 (6). [on-line publication]. http://usuarios.cultura.com.br/eds/PDF/fasam.pdf [access date: 20 February 2009].

Munn, K.; B. Smith (eds.; 2008). *Applied Ontology*. Frankfurt: Ontos/Verlag.

Nickles et al. (2007). 'Ontologies across disciplines'. In Shalley, A. & D. Zaefferer (eds.). *Ontolinguistics*. Berlin: Mouton de Gruyter. 23-70.

Oltramari, A.; Vetere, G. (2008). 'Lexicon and Ontology Interplay in Senso Comune'. In *Proceedings of OntoLex 2008* (Hosted by Sixth international conference on Language Resources and Evaluation). Marrakech (Morocco). [on-line publication]. http://www.loacnr.it/Papers/lexicon_oltramari-vetere.pdf [access date: 07 July 09].

Smith, B. (2004). *Ontology: Philosophical and Computational*. [on-line publication]. http://ontology.buffalo.edu/smith/articles/ontologies.htm [access date: 31 October 2009].

Smith, B. (2008). 'New Desiderata for Biomedical Terminologies'. In Munn, K.; B. Smith (eds.). *Applied Ontology*. Frankfurt: Ontos/Verlag. 83-108.

Wielinga et al. (2001). 'From thesaurus to ontology'. In *International Conference On Knowledge Capture, Proceedings of the 1st international conference on Knowledge capture*. Victoria, Canada. 194-201. [on-line publication]. http://www.cs.vu.nl/~guus/papers/Wielinga01a.pdf [access date: 31 October 2009].