The names of lighting artefacts: extraction and representation of Portuguese and Spanish terms in the archaeology of al-Andalus

Bruno Almeida* — **Rute Costa*** — **Christophe Roche****

* *NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Portugal*
** *Condillac Group, LISTIC, Université Savoie Mont-Blanc, France*

ABSTRACT. *This paper is focussed on the Portuguese and Spanish terms for lighting artefacts, which were extracted from a corpus on the archaeology of al-Andalus. The purpose of the work described in this paper is the creation of an ontology-based multilingual terminological resource. Domain knowledge is represented through OntoAndalus, an OWL ontology which uses DOLCE+DnS Ultralite as a foundation. Language-specific information are modelled through Lemon, the Lexicon Model for Ontologies, which is currently in development by a community group within the W3C. Lemon allows for the representation of grammatical and semantic information, most notably lexicosemantic relations between terms and their reference to ontology elements in OntoAndalus.*

RÉSUMÉ. *Cet article se concentre sur les termes des artefacts d'éclairage dans un corpus sur l'archéologie d'al-Andalus. Le but de ce travail est la création d'une ressource terminologique multilingue basée sur une ontologie. La connaissance du domaine est représentée par OntoAndalus, une ontologie OWL qui repose sur DOLCE + DnS Ultralite. Les informations spécifiques à la langue sont représentées par le modèle Lemon, le Modèle lexical pour les ontologies, lequel est en cours de développement par le W3C. Lemon permet la représentation d'informations grammaticales et sémantiques, notamment les relations lexicosémantiques entre les termes et sa référence aux éléments ontologiques dans OntoAndalus.*

KEYWORDS: *Terminology, Lighting artefacts, Archaeology of al-Andalus, Corpus analysis, Ontologies, Lexicon Model for Ontologies (Lemon).*

MOTS-CLÉS : *terminologie, artefacts d'éclairage, archeologie de l'al-Andalus, analyse de corpus, ontologies, Modèle lexical pour les ontologies (Lemon).*

TAL. Volume 60 – n° 3/2019, pages 113 à 137

1. Introduction

Corpora have had multiple research and practical applications in the past decades, from linguistics and terminology to digital humanities, natural language processing and knowledge representation. Relevant applications include the extraction of terms, contexts and lexicosemantic information in specialised domains (Bowker and Pearson, 2002; Melby, 2012; Meyer, 2001), topic modelling and macroanalysis in the humanities (Jockers, 2013; Meeks and Weingart, 2012) and ontology development (Hitzler *et al.*, 2010; Sure *et al.*, 2009).

This paper describes terminology work carried out in the archaeology of al-Andalus with the purpose of creating an ontology-based multilingual terminological resource in the domain. [1] The constitution of a bilingual comparative corpus allowed to pursue a two-folded methodology based on the double dimension of terminology. [2]

In a first moment, the corpus supported the development of OntoAndalus, an ontology of artefacts in al-Andalusian archaeology, based on the interpretation of textual and visual information. OntoAndalus is being developed with the purpose of constituting a language-independent layer to which terms in several languages may refer to in the proposed terminological resource. The case of lighting artefact concepts is presented in this paper.

At a later time, the corpus allowed to identify and extract Portuguese and Spanish terms used by domain specialists. The case of simple and complex terms denoting lighting artefacts is presented in this paper. The former were extracted by means of the Sketch Engine corpus manager and text analysis tool, which further allowed to study the more frequent collocational patterns involving complex terms for domestic lamps. [3] The terminologies in each language were subsequently organised in lexical networks by means of taxonomy and synonymy relations, and several comparisons were drawn between each lexical network. In this regard, the present paper describes the different conceptual motivations for naming domestic lamps in Portuguese and Spanish. An overview of the terms in each language denoting lighting artefact concepts in OntoAndalus is also provided, in order to facilitate future terminological harmonisation in the domain.

1. The archaeology of al-Andalus is an important subdomain of medieval archaeology in Portugal and Spain (Carvajal López, 2014; Covaneiro *et al.*, 2013). "Al-Andalus" is the Arabic name given to the Iberian Peninsula under Islamic rule during the Middle Ages.

2. Terminology is understood in this paper as an interdisciplinary domain concerned with knowledge and its expression, with the purpose of compiling, studying and presenting terms and concepts in specialised fields (NF ISO 704, 2009). As such, terminology integrates a linguistic dimension and a conceptual dimension. These dimensions constitute independent levels of analysis in terminology work: language-specific (i.e. terms and other linguistic expressions) and language-independent (i.e. concepts and other units of knowledge) (Costa, 2013; Roche, 2015; Santos and Costa, 2015).

3. Available from `https://www.sketchengine.eu/`.

Finally, the relationship between linguistic and conceptual information was modelled through Lemon, the Lexicon Model for Ontologies, which is under development by the W3C Ontology-Lexicon Community Group (Cimiano *et al.*, 2016). Lemon, and in particular its core component Ontolex, has been proposed for a number of projects involving linguistic linked open data and the design of web-based terminological resources (Bosque-Gil *et al.*, 2015; Cimiano *et al.*, 2015; Almeida *et al.*, 2016).

2. Background and motivation

The work presented in this paper is motivated by terminological issues noted by Portuguese and Spanish specialists in the archaeology of al-Andalus. Islamic presence in the Iberian Peninsula covered a period of nearly eight centuries (from 711 to 1492 A.D.), having left behind a wide range of materials, such as pottery, architectural fragments, weaponry, jewellery and glassware. The comparison and study of related finds is made possible through the classification of artefacts. Terminology is closely associated with archaeological classification, since artefact categories require terms for identification and communication purposes. The lack of terminology harmonisation has been noted as a hurdle for scholarly communication, while the development of terminological studies has been recognised as a means to acquire and organise knowledge in the domain (Torres *et al.*, 2003). Terminology work in Portugal was inspired by pioneering studies carried out in Spain since the late 1970's, which were focussed on languages such as Spanish, Catalan and Arabic (Coll Conesa *et al.*, 1988; Rosselló-Bordoy, 1991; Rosselló-Bordoy, 1978).

In Portugal, the need to revitalise Islamic pottery studies led to the creation of the CIGA research group.[4] One of the original purposes of this group was to create a database of the more representative instances of pottery artefacts from the Gharb al-Andalus, i.e. the western province of the al-Andalus. In order to facilitate this purpose, the specialists published a Portuguese terminology and classification of entities such as artefact types, shapes, manufacturing and decorative techniques (Bugalhão *et al.*, 2010).

The CIGA group and its ties to terminological studies in Spain are evidence of a need that goes beyond that of harmonising Portuguese terms. It is therefore thought that an ontology-based multilingual terminological resource could help overcome the communication issues noted by the specialists, as well as help furthering the acquisition of knowledge across multiple communities of practice in Portugal and Spain.

4. CIGA is a Portuguese acronym for *Cerâmica Islâmica do Gharb al-Ândalus* (Islamic Pottery of the Gharb al-Andalus). Available from `http://www.camertola.pt/info/ciga`.

3. Related work

In the past, the development of machine-readable terminological resources relied on reading texts and manually extracting information from them. Terminology work has since evolved towards automatic or semi-automatic extraction of linguistic information from corpora (e.g. term candidates, knowledge patterns, contexts of usage) by means of NLP tools, such as concordancers, corpus managers and other text analysis software (Cabré and Palatresi, 2013; Meyer, 2001; Costa, 2001).

On the other hand, computational terminologies have evolved from simple termbases to full-fledged knowledge-based resources which, besides providing information about the terms used in a specialised domain, are informative of the underlying conceptual structure of the domain itself (Meyer *et al.*, 1992; Nazarenko and Hamon, 2002; Faber *et al.*, 2014; Condamines, 2018). In more recent years, the development of Semantic Web and Linked Data technologies, as well as further research in applied ontology, led to ontology-based approaches to the creation of terminological resources. In the latter, a fundamental distinction is placed between the domain ontology, through which language-independent knowledge is modelled, and the lexical network(s) representing language-specific information about terms and other linguistic units (Roche, 2012).

While corpora are today paramount for terminology work, textual approaches to terminology often ignore (or explicitly reject) the distinction between knowledge and language as distinct levels of analysis, which may lead to several misunderstandings in multilingual terminology work, such as conflating language-specific relations at the term level (e.g. hyponymy, meronymy) with relations drawn at the concept level (e.g. subsumption, part-whole) (L'Homme, 2004; Condamines, 2018). On the contrary, the approach described in this paper explicitly distinguishes between the domain ontology and the Portuguese and Spanish networks of terms, while relating both linguistic and conceptual dimensions of terminology work in an effort towards building an ontology-based terminological resource in the Semantic Web. Here, the role of NLP tools is firmly placed in the linguistic dimension of terminology work, where they excel in extracting language-specific information about terms and other linguistic units, which can then be related to the domain ontology by means of a specific model (in this case, Lemon).

This brings us to the matter of related work in archaeology. To the best of our knowledge, there are no ontology-based terminological resources in our domain of interest, nor in the wider field of archaeological typologies of artefacts. With regard to ontology development, the CIDOC-CRM has become relevant for documenting archaeological data following the ARIADNE project and the initial proposal of the CRMarchaeo extension (Doerr, 2014). However, these initiatives remain focussed on archaeological excavation. This motivated the development of OntoAndalus, a domain ontology focussed on artefacts in al-Andalusian archaeology, which will be briefly described in the following section.

4. OntoAndalus: an ontology of artefacts in OWL

OntoAndalus is an ontology of relevant artefacts in the archaeology of al-Andalus. It was developed as part of a PhD thesis (Almeida, 2019) and is presently made available under a Creative Commons License (CC-BY-4.0). [5] The development process was based on the interpretation of selected texts from a Portuguese and Spanish specialised corpus, as well as English reference works on archaeology. The more specialised texts consist of Portuguese and Spanish conference papers, journal articles, theses and monographs on the description, classification and terminology of Islamic artefacts (mostly pottery). OntoAndalus is based on the so-called "functional form" criterion of classification of the artefacts, which is followed in the artefact typologies of the CIGA group and Rosselló-Bordoy (Bugalhão *et al.*, 2010; Rosselló-Bordoy, 1991; Rosselló-Bordoy, 1978).

OntoAndalus was developed using the Protégé ontology-editor. [6] Protégé integrates a host of tools and plugins which are invaluable for the modelling process and for visualisation purposes (e.g. plugins for generating conceptual graphs). OWL was chosen as a modelling language due to its relative simplicity and status as a W3C recommendation (W3C OWL Working Group, 2012). DOLCE+DnS Ultralite (DUL) was chosen as the foundational ontology for the development of OntoAndalus. DOLCE was one of the first notable top-level ontologies following the development of applied ontology as a research field (Guarino and Musen, 2005; Munn and Smith, 2008). [7] DUL is a streamlined version of DOLCE-Lite, the original translation of DOLCE into OWL, based on ontology design patterns (Gangemi, 2016). The latter enable the reuse of smaller ontological components in order to more efficiently solve recurrent modelling problems (e.g. physical objects, events). Besides streamlining the original translation of DOLCE into OWL, DUL also integrates the Descriptions and Situations ontology (DnS) for modelling social and cognitive entities (e.g. information objects). Another advantage of DUL lies in its complete availability in OWL format, including all classes and binary relations. [8]

As of the writing of this paper, OntoAndalus consists of 161 classes, 30 object properties and 135 individuals (excluding the elements already defined in DUL). The ontology includes 72 artefact types, which are organised in the following categories: lighting artefacts, tableware, kitchenware, domestic artefacts, recreational artefacts, ritual artefacts, agricultural artefacts, construction artefacts, artisanal artefacts, storage artefacts, transportation artefacts and artefact components.

5. OntoAndalus is the topic of a forthcoming paper (Almeida and Costa, 2019). The ontology is made available through `https://github.com/brunoalmeida81/OntoAndalus`.

6. Available from `https://protege.stanford.edu`.

7. "DOLCE" is an acronym for Descriptive Ontology for Linguistic and Cognitive Engineering (Masolo *et al.*, 2003). Although slightly outdated, Mascardi *et al.* (2007) provide a good overview and comparison of similar top-level ontologies.

8. Available from `http://www.ontologydesignpatterns.org/ont/dul/DUL.owl`.

4.1. *Modelling lighting artefacts*

In this section, the category of lighting artefacts will be described as an example of how artefact types are modelled in OntoAndalus. This category includes some of the more representative artefacts of the Islamic period in the Iberian Peninsula (Gómez Martínez, 2004).

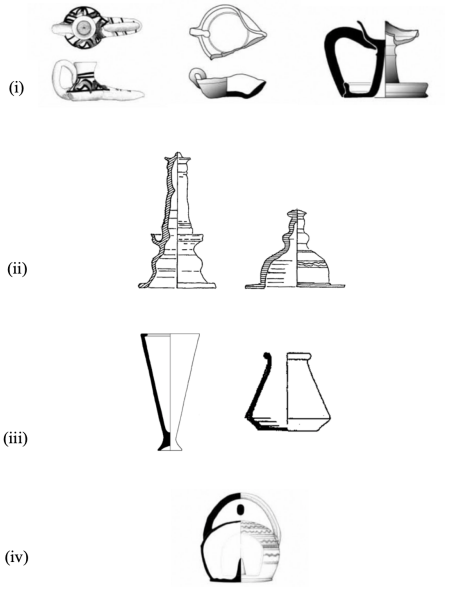

Figure 1. *Graphical representation of types of lighting artefacts.*

According to Gómez Martínez (2004), archaeologists usually acknowledge the existence of four kinds of lighting artefacts, denoted by the Spanish terms (i) *candil*; (ii) *policandela* (or *almenara*); (iii) *lamparilla*; and (iv) *fanal*. Figure 1 shows graphical representations of typical instances of each series. [9]

In order to maintain consistency with DUL and facilitate international communication, each class in OntoAndalus has an English identifier, namely: (i) `Lamp`; (ii) `MultipleLamp`; (iii) `StationaryLamp`; and (iv) `Lantern`. The formal definitions of these classes put forward in OntoAndalus adhere to the following pattern: **superordinate class + collection + function + part(s) or component(s)**. The following paragraphs explicate each class through natural-language definitions derived from formal definitions in the ontology, along with references to relevant texts in the corpus.

9. These illustrations are reproduced from Bugalhão *et al.* (2010, p. 471), Rosselló-Bordoy (1991, p. 174), Vallejo Triano and Escudero Aranda (1999, p. 165) and Gómez Martínez (2000, p. 433).

`Lamp` (*candil*$_{es}$) *Def.* Artefact for lighting in closed spaces composed by at least one spout and a single chamber for liquid fuel (Coll Conesa *et al.*, 1988; Gómez Martínez, 2004; Rosselló-Bordoy, 1991).

`Multiple lamp` (*almenara*$_{pt}$, *almenara*$_{es}$, *policandela*$_{es}$). *Def.* Artefact for stationary lighting in closed spaces composed by more than one chamber for liquid fuel unified by a structure (Gómez Martínez, 2004; Rosselló-Bordoy, 1991).

`Stationary lamp` (*lamparilla*$_{es}$). *Def.* Artefact for stationary lighting in closed spaces composed by a single chamber for liquid fuel (Vallejo Triano and Escudero Aranda, 1999).

`Lantern` (*fanal*$_{pt}$, *lanterna*$_{pt}$; *fanal*$_{es}$, *linterna*$_{es}$). *Def.* Artefact for lighting in open spaces composed by a single chamber for solid fuel (Bugalhão *et al.*, 2010; Gómez Martínez, 2004).

The `Lamp` class is the more complex part of the ontology with regard to lighting artefacts (figure 2). OntoAndalus includes four criteria of subdivision of `Lamp`: (i) vessel form, (ii) type of spout, (iii) inclusion of a discus or neck and (iv) inclusion of a tall foot. The multiple criteria of subdivision are represented through pairwise disjoint defined classes.

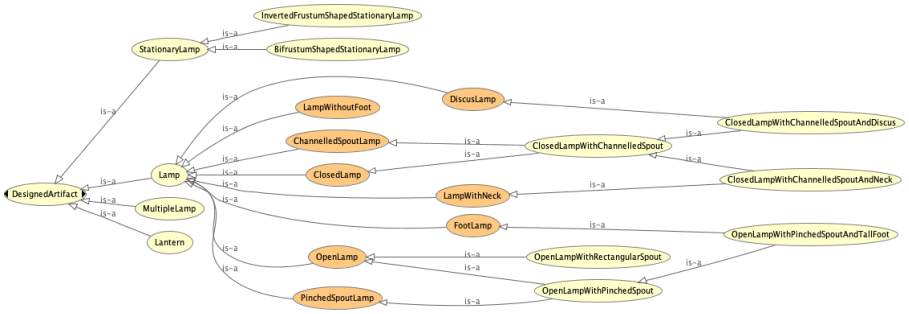

Figure 2. *Types of lighting artefacts in OntoAndalus.*

The vessel form (i.e. open or closed) is generally considered to be the more salient criterion for distinguishing between types of lamps (Bugalhão *et al.*, 2010; Gómez Martínez, 2004). This was chosen as the primary criterion of subdivision of the `Lamp` class. Therefore, only the disjoint classes `ClosedLamp` and `OpenLamp` are further subdivided in the asserted hierarchy of OntoAndalus, while the remaining classification can be inferred by a reasoner.

We have seen how artefact types can be modelled in OntoAndalus in order to constitute the language-independent layer of an ontology-based terminological resource. This includes formal descriptions and definitions of classes, which guide the drafting of natural-language definitions. The following section will describe the work carried out with regard to the extraction and representation of Portuguese and Spanish terms

from specialised corpora, as well as the relationship between conceptual and linguistic aspects.

5. Extracting and modelling information at the term level: the case of lighting artefacts

5.1. *Criteria for corpus design*

The purpose of constituting a corpus was twofold with regard to the work described in this paper. On the one hand, the corpus facilitated the understanding of the domain and assisted the modelling of OntoAndalus. On the other hand, the corpus was also paramount for compilling terminological and lexicosemantic information. The criteria for corpus design represented in table 1 were adopted based on Bowker and Pearson (2002) and Cabré (2008).

Domain	Pottery of the al-Andalus
Language	Portuguese, Spanish
Time period	1970 and later
Level of specialisation	Medium to high
Text integrity	Full texts
Medium and modality	Digitised written texts with visual information
Text genres and discourse types	Heterogenous

Table 1. *Criteria for corpus design.*

The texts constituting the corpus are primarily about al-Andalusian artefacts and were originally written in Portuguese or Spanish. [10] The time period is justified by the fact that the archaeology of al-Andalus only established itself as a domain from the 1970's onward. The corpus includes full texts with a medium to high level of specialisation, i.e. produced by domain specialists and intended for actual or future specialists of the domain. This implies the exclusion of works intended for the general audience, since the terminology used therein could be significantly different from that employed in specialised discourse (Bowker and Pearson, 2002).

Regarding the medium and modality, the corpus is formed by written texts with visual information. There are several reasons for the latter requirement: (i) the importance of the visual modality in the domain, (ii) facilitating the understanding of the

10. Following Costa (2001), a specialised text is understood here as a stable linguistic product resulting from the professional activity of experts within a specialised community. Specialised texts are, therefore, constrained by social and communicative conditions which determine, for example, their boundaries and internal structure.

domain (e.g. artefact kinds, vessel morphology) and (iii) future work (e.g. enrichment of a future terminological resource with visual information). [11]

Lastly, the corpus is heterogenous with regard to genres and discourses. It ranges over transdisciplinary discourses (e.g. archaeology, history) as well as multiple text genres (e.g. thesis, journal articles). The justification for this option lies in our focus on domain knowledge, which can be expressed across several genres and discourses.

5.2. *Structure and composition of the corpus*

A simple structure was devised for our corpus according to language and text genre. This was based on previous work carried out by Costa (2001). A review of the scholarly communication of the domain allowed to identify the more significant text genres, namely: (i) theses and dissertations, (ii) monographs, (iii) catalogues, (iv) articles in scientific journals, (v) papers presented in conferences and other events.

Although the proposed typology is methodologically useful, it does present several issues. For one, the same text (or a very similar version) may range over more than one genre. For example, the content of a thesis may be later published as a monograph or even as a series of monographs. This is the case of the doctoral thesis of Rosa Varela Gomes, entitled *Silves (Xelb): uma cidade do Gharb al-Andalus: arqueologia e história (séculos VIII-XIII)*. This thesis, originally presented in 2000, was later published in four volumes of the monograph series *Trabalhos de arqueologia*, which are included in the corpus (Gomes, 2003; Gomes, 2011; Gomes, 2006; Gomes, 2002).

A further difficulty was posed by museum and exhibition catalogues. These often include articles by several authors along with the actual catalogue description of the artefacts. For this reason, it was decided to include a separate category for catalogue articles in the structure of the corpus. The descriptions themselves were not included in the final typology, since we did not have access to digitised catalogues of artefacts in both languages. It is common, however, for other genres to include catalogues in a separate section. The discourse of catalogue descriptions is, therefore, present in the corpus. [12]

Tables 2 and 3 show the composition of the Portuguese and Spanish corpora. The data shown was gathered from Sketch Engine, the corpus manager used in the work carried out in this paper. Both corpora are comparable in terms of size: the Portuguese

11. The visual modality plays an important role in archaeology. While photography allows for a realistic depiction of an object, drawing is paramount in conveying selective (or *diagrammatic*) information about an artefact (Adkins and Adkins, 2009; Caballero Zoreda, 2006). With regard to terminology work, the importance of visual information in corpora design and terminological resources has already been noted (Prieto Velasco and Faber, 2012).

12. For example, the monograph of Rosselló-Bordoy (1978) about the classification and terminology of Majorcan pottery includes a catalogue in the final section. This work is available in an electronic format, which facilitated its inclusion in the corpus.

Text genre	**Texts (N)**	**Tokens (N)**	**Word-forms (N)**	**Word-forms (%)**
Theses, dissertations	6	324,817	242,597	36.20
Monographs	5	485,225	362,401	54.08
Journal articles	6	54,171	40,458	6.04
Catalogue articles	1	6,789	5,070	0.76
Conference papers	4	26,257	19,610	2.93
Total	22	897,259	670,136	≈100

Table 2. *Composition of the Portuguese corpus.*

Text genre	**Texts (N)**	**Tokens (N)**	**Word-forms (N)**	**Word-forms (%)**
Theses, dissertations	3	747,865	575,011	81.45
Monographs	1	76,037	58,462	8.28
Journal articles	7	66,684	51,271	7.26
Catalogue articles	1	3,977	3,057	0.43
Conference papers	4	23,604	18,148	2.57
Total	16	918,167	705,949	≈100

Table 3. *Composition of the Spanish corpus.*

corpus has over 670,000 word-forms while the Spanish corpus has over 705,000 word-forms. [13]

There is some asymmetry with regard to text genre: monographs are the more represented genre in the Portuguese corpus, while theses and dissertations are predominant in the Spanish corpus. Furthermore, all of the Portuguese texts included in the latter category are master's dissertations, while the Spanish texts are doctoral theses. These circumstances are compensated by the fact that four of the five monographs in the Portuguese corpus correspond to the doctoral thesis of Rosa Varela Gomes.

13. Sketch Engine distinguishes word-forms from tokens. The former are the several forms assumed by lexemes (i.e. the English verb "to go" has the word-forms "go", "went", "gone"), while the latter are instances of word-forms and punctuation symbols.

5.3. *Extraction of linguistic information from the corpora*

The NLP part of our work was carried out using Sketch Engine, a corpus manager and text analysis tool, following the constitution of a Portuguese and Spanish corpus of specialised texts.

The first stage of this consisted in extracting frequency lists of word-forms in both languages and selecting simple terms (i.e. with a single root) for our case study of lighting objects. Only words occurring more than once in a single text were considered for analysis.

A further stage consisted in extracting complex terms, which denote subtypes of the above-mentioned artefact types. The extraction was carried out based on the collocational strength between the head-word and its modifiers. A "collocate", in this context, describes any co-occurring words within a specified pattern. These collocates were automatically ordered by Sketch Engine according to their logDice score, which measures the collocational strength between two words in a corpus (Rychlý, 2008). LogDice has a theoretical maximum of 14, in which every instance of X in a corpus co-occurs with Y (and vice-versa). A score above 10 represents a strong collocation. Only collocates with a logDice score above 10 were considered for analysis.

The examples of complex terms presented in this paper adhere to the following patterns, which involve the more relevant collocates in the corpus (N = noun, P = preposition, A = adjective):

– `candil:N de:P N A?` (for both languages);

– `candeia:N de:P N A?` (only relevant in the case of Portuguese).

5.4. *The names of lighting artefacts in Portuguese*

Table 4 shows the Portuguese simple terms for lighting artefacts employed in the corpus.

The `Lamp` concept and its subordinates are paramount for the linguistic expression of lighting artefacts. There are, however, some issues regarding more generic lighting artefacts concepts in Portuguese. In one text of the corpus, *luminária* denotes lighting artefacts in general. Other generic terms are *lâmpada* and *candeeiro*, although the latter is much less relevant in the corpus. *Lâmpada* seems to approximate the `Lamp` concept more clearly, distinguishing it from *lanterna* and other terms denoting domestic or non-domestic lighting artefacts.

A dichotomy is established between closed and open lamps with, respectively, *candil* and *candeia*. This is in line with the terminological proposals put forward by several archaeologists (Bugalhão *et al.*, 2010; Torres *et al.*, 2003).

Term	Frequency	Texts
candil	551	15
lucerna	495	13
candeia	217	9
lamparina	68	8
luminária	67	1
lâmpada	10	2
lanterna	7	3
vela	5	1
tocha	3	1
fanal	2	2
candeeiro	2	1

Table 4. *Simple terms for lighting artefacts in Portuguese.*

The `candil:N de:P N A?` pattern provides an insight of how possible types of lamp are named in Portuguese, most notably the `ClosedLamp` concept. Table 5 shows the more significant noun collocates in the corpus.

Collocate	Frequency	logDice
disco	19	11.97
bico	17	11.3
depósito	7	10.45
pé	10	10.43

Table 5. *Noun collocates following* candil:N de:P *in the Portuguese corpus.*

As we can see, these collocates denote parts of the artefacts, represented by the `Discus` (*disco*), `Spout` (*bico*), `LampFuelChamber` (*depósito*) and `Foot` (*pé*) concepts in OntoAndalus. The collocates are present in the following complex terms in the corpus:

– *candil de disco impresso* (19 occurrences);

– *candil de bico* (17 occurrences);

– *candil de depósito aberto* (7 occurrences);

– *candil de pé alto* (10 occurrences).

Candil de disco impresso denotes an established type of lamp. [14] The second collocate, *bico*, occurs in the terms *candil de bico*, *candil de bico comprido* and *candil de bico curto*. These expressions are used to distinguish between closed lamps based on salient characteristics, namely the discus and spout.

14. This type, characterised by the discus surrounding the pouring hole (Zozaya, 1999), is represented in our ontology through the `DiscusLamp` concept.

With one exception, *depósito* and *pé* only collocate with *candil* when the *candil*/*candeia* dichotomy is not followed. These nouns take part in the terms *candil de depósito aberto* and *candil de pé alto*, respectively. The exception is a context where *candil de pé alto* is explicitly rejected in favour of *candeia de pé alto*.

The `candeia:N de:P N A?` pattern is equally informative in Portuguese. As we can see in table 6, the more relevant collocates also denote parts, which are represented by the `Foot` (*pé*) and `LampFuelChamber` (*depósito*/*câmara*) concepts in OntoAndalus. Contrary to the case of *candil*, *candeia* is used almost exclusively to denote an open lamp.

Collocate	**Frequency**	**logDice**
pé	25	12.05
depósito	10	11.41
câmara	4	10.2

Table 6. *Noun collocates following* candeia:N de:P *in the Portuguese corpus.*

These word-forms are employed in the following complex terms:

– *candeia de pé* (25 occurrences);

– *candeia de depósito aberto* (10 occurrences);

– *candeia de câmara aberta* (4 occurrences).

The first collocate, *pé*, occurs in the terms *candeia de pé* (5 occurrences) and *candeia de pé alto* (20 occurrences). These expressions are used indiscriminately in the corpus to refer to the same type of artefact, which is represented by the `FootLamp` concept in our ontology. This seems to indicate that the foot of these lamps is always 'tall' when compared to similarly-sized artefacts. A proposed term for the appendage of this type of lamp is *pé alto sobre prato de sustentação*, as opposed to *pé alto maciço*, which characterises other kinds of artefacts (Bugalhão *et al.*, 2010). *Candeia de pé* is, therefore, an abbreviation of longer and more precise terms.

The remaining collocates of *candeia* denote the fuel chamber of the artefacts. These motivate the seemingly redundant expressions *candeia de câmara aberta* and *candeia de depósito aberto*. Both expressions highlight the open fuel chamber as a distinctive quality of these artefacts.

5.5. *The names of lighting artefacts in Spanish*

The `Lamp` concept and its subordinates are also predominant naming-wise in Spanish (table 7). Contrary to Portuguese, *candil* is consensually used to denote both open and closed forms of al-Andalusian artefacts. *Lámpara* and *luminaria* are the more employed generic terms for lighting artefacts. The remaining terms apply to less studied or controversial artefact kinds (i.e. *policandela*, *fanal*, *linterna*, *lamparilla*, *almenara*) and artefacts emanating from the Roman period (i.e. *lucerna*).

Term	Frequency	Texts
candil	489	15
lámpara	25	5
fanal	22	2
lucerna	14	7
almenara	14	3
lamparilla	13	3
linterna	7	3
policandela	7	2
luminaria	3	1
candelabro	2	3

Table 7. *Simple terms for lighting artefacts in Spanish.*

As in the case of Portuguese, the `candil:N de:P N A?` pattern is of import for types of domestic lamps. As we can see in table 8, the noun collocates denote partitive concepts, namely `Spout` (*piquera*), `Foot` (*pie*), `Discus` (*disco*) and `LampFuelChamber` (*cazoleta* and *depósito*).

Collocate	Frequency	logDice
piquera	58	12.41
pie	32	11.76
disco	22	11.45
cazoleta	26	11.40
depósito	11	10.39

Table 8. *Noun collocates following* candil:N de:P *in the Spanish corpus.*

The more frequent complex terms containing these word-forms are the following:

– *candil de piquera* (57 occurrences);

– *candil de pie alto* (30 occurrences);

– *candil de disco impreso* (18 occurrences);

– *candil de cazoleta abierta* (9 occurrences);

– *candil de depósito abierto* (9 occurrences).

The first term denotes a thoroughly studied type of lamp in al-Andalusian archaeology (Zozaya, 2007). Although *piquera* may be used for denoting any kind of spout for holding a wick, it is often used for denoting the characteristic spout of closed lamps (i.e. the `ChannelledSpout` concept). The term *pellizco*, which only co-occurs once with *candil* in the corpus, denotes the spout that is typical of open lamps (i.e. the `PinchedSpout` concept). The term *piquera de pellizco*, with 18 occurrences in 4 texts of the corpus, is however a more precise denomination, since it makes clear the reference to a part (i.e. the pinched spout).

However, there does not seem to be any parallel for spouts applied to closed lamps. Instead, *piquera* is further modified by adjectives (e.g. *larga*, *corta*) or prepositional phrases (e.g. *de entronque suave*, *de quilla de barco*, *de bañera*), all of which denote sizes or shapes of the spout used in closed lamps.

While the spout is the more salient characteristic for referring to a closed lamp, the fuel chamber is predominant for reference to an open lamp. This is the more likely explanation for the fact that *candil de pellizco* and *candil de cazoleta/depósito cerrada(o)*, which follow the opposite motivation, are less relevant in the corpus.

5.6. *Representation and comparison of the Portuguese and Spanish terms*

The Portuguese and Spanish terms can be represented through lexical networks. The latter are prominent devices for representing language-specific information, which may be used for creating a concept-based terminological resource (Santos and Costa, 2015).

Figures 3 and 4 show the more relevant terms in Portuguese and Spanish. Terms motivated by the same criteria of subdivision (e.g. vessel form) are represented in the graphs through divided taxonomic arcs (e.g. *candil* and *candeia* in Portuguese).

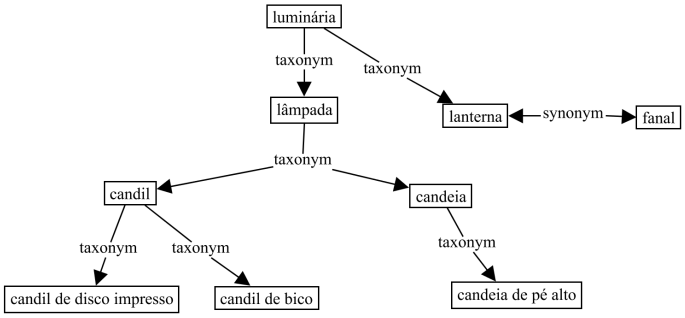

Figure 3. *Lexical network of lighting artefacts in Portuguese.*

The lexical relations employed in both networks are taxonomy, a specialisation of the hyponymy relation (Cruse, 1986), and synonymy. The latter is restricted here to what is assumed to be absolute synonymy. Including the relation of quasi-synonymy in the graphs would imply a much more complex network, since any two terms denoting closely related concepts could be considered as quasi-synonyms.

The generic terms *luminária*$_{pt}$/*luminaria*$_{es}$ and *lâmpada*$_{pt}$/*lámpara*$_{es}$ pose several difficulties. We have followed the assumption that the former are superordinates of the latter. While these terms are not directly relevant for the Islamic period, we have included them in the networks to clarify that *candil* can be seen as a subordinate term of *lâmpada*$_{pt}$/*lámpara*$_{es}$ in both languages. This leads us to argue that *lucerna*$_{pt}$/*lucerna*$_{es}$ are not synonyms of *candil*$_{pt}$/*candil*$_{es}$ in neither language. In-

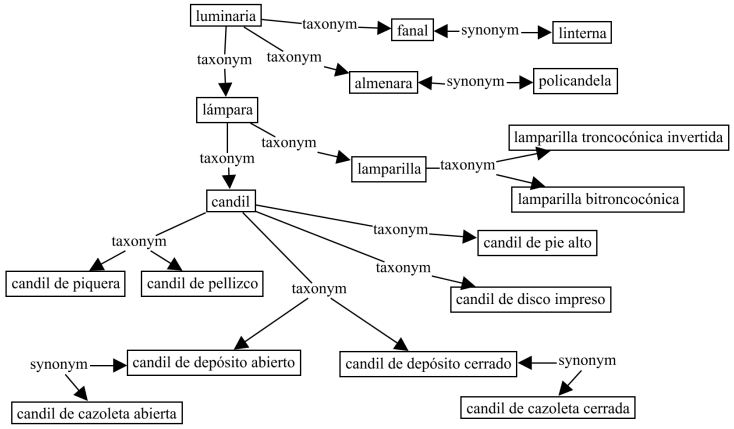

Figure 4. *Lexical network of lighting artefacts in Spanish.*

stead, they are subordinates of *lâmpada*$_{pt}$/*lámpara*$_{es}$: while *lucerna*$_{pt}$/*lucerna*$_{es}$ denote an artefact kind of the Classical period, *candil*$_{pt}$/*candil*$_{es}$ denote a related artefact kind of the Islamic period (as assumed by most specialists).

As we can see in the graphs, the situation is markedly different in both languages. A significant difference in the networks lies in the *candil*/*candeia* dichotomy in Portuguese. As pointed out before, *candil* is used in Spanish to denote both open and closed lamps, whereas in Portuguese it only denotes closed lamps.

A further difference lies in the use of criteria of subdivision, which is more apparent in Spanish. In Portuguese, the most obvious case is the *candil*/*candeia* dichotomy, which is based on the overall form of the vessels. The term *candil de bico* is motivated by the spout, but there are no converse terms based on this criterion in the corpus. Furthermore, its adequacy as a term is doubtful, since every kind of lamp in the domain should have a spout of some sort for holding the wick in place. *Candil de disco impresso*, on the other hand, is motivated by the portion of the chamber surrounding the orifice, which – in this kind of closed lamp – has a discus instead of a neck. Finally, the presence of an applied foot is the only productive criterion in the case of open lamps in Portuguese (i.e. *candeia de pé alto*).

In Spanish, the chamber is also an important criterion, as attested by *candil de depósito abierto*/*cerrado* and its respective synonyms *candil de cazoleta abierta*/*cerrada*. Other denominations are based on the spout, namely *candil de piquera* (57 occurrences) and *candil de pellizco* (1 occurrence). Furthermore, there is the already pointed out ambiguity in interpreting *piquera*, since *piquera de pellizco* is present in the corpus (cf. section 5.5). Finally, there are terms motivated by the presence or absence of a discus and applied foot: *candil de disco impreso* and *candil de pie alto*. A summary of the terms in both languages is shown in table 9.

Criteria of subdivision	**Portuguese term**	**Spanish term**
Chamber	*candil* *candeia*	*candil de depósito (cazoleta) cerrado(a)* *candil de depósito (cazoleta) abierto(a)*
Spout	*candil de bico* (?)	*candil de piquera* *candil de pellizco* (?)
Discus / neck	*candil de disco impresso*	*candil de disco impreso*
Foot	*candeia de pé alto*	*candil de pie alto*

Table 9. *Terms for lamps according to different criteria of subdivision.*

With regard to the other kinds of lighting artefacts, only the lantern is represented in both languages through the archaism *fanal*. The latter is, however, used more extensively in the Spanish corpus, while *lanterna* is the preferred denomination in the Portuguese corpus. *Almenara* and *policandela* are used interchangeably for denoting the same artefact kind in Spanish. Subtypes of the stationary frustum-shaped lamps are denoted in Spanish using adjective modifiers. The latter denote approximate geometrical shapes (i.e. *bitronconcónica* and *troncocónica invertida*).

5.7. *Relationship between linguistic and conceptual information*

We have shown how language-specific information can be represented by means of lexical networks. This brings into question the relationship between the lexical networks and OntoAndalus. Making this relationship explicit allows to contrast both languages with regard to the conceptualisation of the domain.

Table 10 summarises the relationship between the more established terms in the corpus and the concepts of lighting artefacts in OntoAndalus. The table highlights the asymmetry in both languages in expressing concepts from this section of the ontology.

Finding Portuguese denominations for the `Lamp` concept as well as for the artefact types not studied by the Portuguese archaeologists remain open questions. As shown in the Portuguese lexical network, the `Lamp` concept may remain unnamed in this language, since the generic term *lâmpada* is a suitable hypernym of *candil* and *candeia*. The second issue can be more problematic. A possible solution is using the terms *candeeiro* and *lamparina* for, respectively, the `MultipleLamp` and `StationaryLamp` concepts in our ontology, since both terms are already present in the corpus. Another possibility is to use the term *almenara* for the latter concept since the term exists in both languages (Gómez Martínez, 2004). It is, however, not represented in the Portuguese corpus.

Concept	Portuguese term	Spanish term
Lamp		*candil*
ClosedLamp	*candil*	*candil de depósito (cazoleta) cerrado(a)*
ClosedLampWithChannelled Spout		
ClosedLampWithChannelled SpoutandDiscus		
OpenLamp	*candeia*	*candil de depósito (cazoleta) abierto(a)*
OpenLampWithPinchedSpout		
OpenLampWithPinchedSpout andTallFoot		
OpenLampWithRectangular Spout		
ChannelledSpoutLamp	*candil de bico* (?)	*candil de piquera*
PinchedSpoutLamp		*candil de pellizco* (?)
FootLamp	*candeia de pé alto*	*candil de pie alto*
DiscusLamp	*candil de disco impresso*	*candil de disco impreso*
Lantern	*fanal* *lanterna*	*fanal* *linterna*
MultipleLamp	*almenara* *policandela*	
StationaryLamp	*lamparilla*	
BifrustumShaped StationaryLamp	*lamparilla bitroncocónica*	
InvertedFrustumShaped StationaryLamp	*lamparilla troncocónica invertida*	

Table 10. *Relationship between concepts of lighting artefacts and their terms.*

In this section, extracted term candidates in each language were described, which highlighted several inconsistencies in each lexical network in relation to the domain ontology. This is but one step of the overall process, since the extracted data would require expert validation before it can be included in a future terminological resource, in order to follow a quality-based approach to terminology management (Silva, 2014). Future work involving domain specialists will be carried out with regard to validation and terminology harmonisation, in which quantitative and/or qualitative methods, such as surveys or focus groups, will be employed.

5.8. *Modelling linguistic information with Lemon*

Lemon is an acronym for "Lexicon Model for Ontologies." The purpose of the model is to provide a linguistic grounding for computational ontologies. Most notably,

Lemon can be used to represent how ontology elements (e.g. classes, object properties, instances) are expressed in natural language.

Lemon consists of the following modules:

– *Ontolex*. It allows to establish an interface between a lexicon and an ontology;

– *Synsem*. It allows to represent information at the syntactic and semantic levels;

– *Decomp*. It allows to represent information on the decomposition of complex expressions;

– *Vartrans*. It allows to represent information regarding variation and translation;

– *Lime*. It allows to represent linguistic metadata.

Ontolex, the core module of Lemon, is especially important, since it allows to establish a relationship between a lexicon (or terminology) and an ontology. Lemon is structured around lexical entries, which are either single words, multiword expressions or affixes. A lexical entry is realised as a series of forms in a language. In Lemon, each entry needs to be linked to at least one form and, at most, to one canonical form. The latter is typically the lemma of a lexical entry in a dictionary. Each form may have written and/or phonetic representations.

There are several ways for relating a lexical entry to an ontology element. The relationship can be established directly through the `denotes/isDenotedBy` object properties. Another option is to establish a mediated link through the "lexical sense" construct. The latter allows to model the fact that a single lexical entry may have several distinct senses in a language, as traditionally represented in dictionaries. For example, "consumption" in English may be used in the everyday sense of "act of consuming" or in several specialised senses (e.g. in economics). Each of these senses may be linked to different ontology elements via the `reference/isReferenceOf` properties. Lemon also allows to represent pragmatic information regarding the usage of lexical entries. For example, the different senses of the French words *rivière* and *fleuve* may be clarified, although both of them can still point to the same ontology element, if this is deemed useful in support of a particular modelling decision.

Lemon further introduces the "lexical concept". The latter allows to model a unit of thought or collection of senses which are not directly represented in an ontology. A lexical concept may be associated to a lexical entry and to an ontology element via the available object properties. It may also be associated with a natural language definition through the `skos:definition` property. Finally, concept sets may be defined in order to organise a lexicon according to the concept (i.e. onomasiologically).

In the present case, language-specific information can be represented using Lemon while conceptual information is left in OntoAndalus. The former consists of information at the term level, including grammatical information and lexicosemantic relations. Conceptual information, on the other hand, pertains to domain knowledge. This approach allows to distinguish between the linguistic and conceptual dimensions of terminology work while still drawing relationships between each dimension.

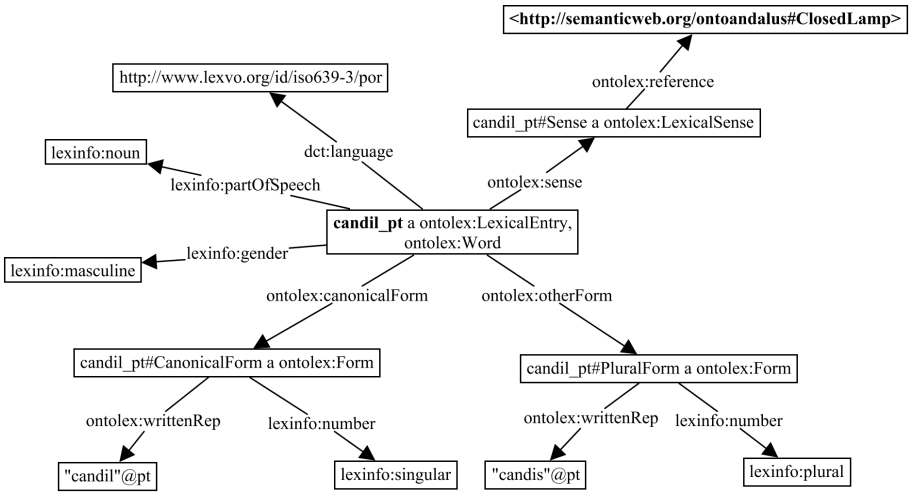

Figure 5. *Term entry in Lemon.*

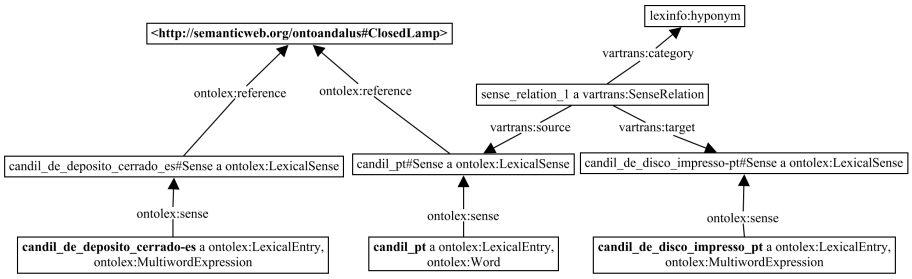

Figure 6. *Representing lexicosemantic relations in Lemon.*

Figure 5 shows a possible term entry for the Portuguese term *candil*. Relevant linguistic information includes the language, part of speech and grammatical gender. As recommended in Lemon, this information is represented via Dublin Core metadata and the LexInfo model. [15] Canonical and plural forms are provided along with their respective written representations.

The links between terms and the predicates of OntoAndalus are mediated by the lexical sense construct. This approach is required for the representation of lexicosemantic relations at the term level in each language. As we can see in figure 6, the equivalence between *candil*$_{pt}$ and *candil de depósito cerrado*$_{es}$ can be represented simply by pointing both lexical senses to the same class in OntoAndalus.

15. Available from, respectively, `http://dublincore.org/documents/dcmi-terms` and `https://www.lexinfo.net`.

Since LexInfo does not include the relation of taxonomy, it has been replaced in the graph above with the broader hyponymy relation. In Lemon, the relation is established by indicating the source and target terms as well as the respective category in the LexInfo model. In this example, the Portuguese term *candil de disco impresso* is asserted as a hyponym of *candil*.

The operationalisation of this data requires its expression through a suitable formalism. The following RDF code in Turtle syntax represents grammatical and semantic information about the Portuguese term *candil*:

```
@prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#> .
@prefix vartrans: <http://www.w3.org/ns/lemon/vartrans#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix lexinfo: <http://www.lexinfo.net/ontology/2.0/lexinfo#> .

:candil_pt a ontolex:LexicalEntry, ontolex:Word ;
    dct:language <http://www.lexvo.org/id/iso639-3/por> ;
    lexinfo:partOfSpeech lexinfo:noun ;
    lexinfo:gender lexinfo:masculine ;
    ontolex:canonicalForm :candil_pt#CanonicalForm ;
    ontolex:otherForm :candil_pt#PluralForm ;
    ontolex:sense :candil_pt#Sense .

:candil_pt#CanonicalForm a ontolex:Form ;
    ontolex:WrittenRep ‘‘candil’’@pt ;
    lexinfo:number lexinfo:singular .

:candil_pt#PluralForm a ontolex:Form ;
    ontolex:WrittenRep ‘‘candis’’@pt ;
    lexinfo:number lexinfo:plural .

:candil_pt#Sense a ontolex:LexicalSense ;
    ontolex:reference <http://semanticweb.org/ontoandalus#ClosedLamp> .

:senseRelation1 a vartrans:SenseRelation ;
    vartrans:source :candil_pt#Sense ;
    vartrans:target :candil_de_disco_impresso_pt#Sense ;
    vartrans:category lexinfo:hyponym .
```

As we can see, Lemon enables the representation of diverse information at the term level. Grammatical information includes the gender, part of speech and singular and plural forms of the term. Semantic information includes reference to a class in OntoAndalus as well as the hyponymy relation between the Portuguese terms *candil* and *candil de disco impresso*.

6. Conclusion

This paper presented our work towards multilingual terminological resource aimed at experts and students of the archaeology of al-Andalus. OntoAndalus provides a language-independent conceptualisation of the domain, which can be shared across multiple communities of practice, while the language-specific components can be represented with Lemon, a model for the linguistic grounding of computational ontologies. The constitution and subsequent analysis of the corpus with Sketch Engine was paramount for representing both language-specific and language-independent information.

The case of lighting artefacts was highlighted in this paper, starting with the conceptualisation of these artefact types in OntoAndalus and leading to the extraction and representation of their Portuguese and Spanish terms, with a special emphasis on complex terms derived from collocational patterns in the corpus. Lemon provides the necessary means for the representation of rich grammatical and semantic information on the terms, including lexicosemantic relations and reference to ontology elements. The approach described in this paper, therefore, is able to distinguish between the linguistic and conceptual dimensions of terminology work while drawing useful relationships between each dimension. Combining methods from NLP and ontology engineering allowed to better meet the needs of digital humanities research, in particular in the archaeology of al-Andalus. Such an approach helps the archaeologist to clearly differentiate matters pertaining to the terms used in each language from matters pertaining to domain knowledge, which may help guide future initiatives in terminology harmonisation as well as facilitate the dissemination of knowledge for research and educational purposes.

Acknowledgements

Research financed by Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia as part of the project Centro de Linguística da Universidade NOVA de Lisboa – UID/LIN/03213/2019.

7. References

Adkins L., Adkins R., *Archaeological illustration*, University Press, Cambridge, 2009.

Almeida B., Terminology and knowledge representation: ceramic artefacts of al-Andalus, Doctoral thesis, Universidade NOVA de Lisboa, 2019.

Almeida B., Costa R., "OntoAndalus: an ontology of Islamic artefacts for terminological purposes", 2019, Manuscript submitted for publication.

Almeida B., Roche C., Costa R., "Terminology and ontology development in the domain of Islamic archaeology", *in* H. Thomsen, A. Pareja-Lora, B. Madsen (eds), *Term bases and linguistic linked open data: TKE 2016*, Copenhagen Business School, Copenhagen, p. 147-156, 2016.

Bosque-Gil J., Gracia J., Aguado-de Cea G., Montiel-Ponsoda E., "Applying the OntoLex Model to a multilingual terminological resource", *in* F. Gandon, C. Guéret, S. Villata, J. Breslin, C. Faron-Zucker, A. Zimmermann (eds), *The Semantic Web: ESWC 2015 Satellite Events*, Springer International Publishing, p. 283-294, 2015.

Bowker L., Pearson J., *Working with specialized language : a practical guide to using corpora*, Routledge, London, 2002.

Bugalhão J., Catarino H., Cavaco S., Covaneiro J., Fernandes I. C., Gomes A., Gómez Martínez S., Gonçalves M. J., Grangé M., Inácio I., Lopes G., Santos C., "CIGA: projecto de sistematização para a cerâmica islâmica do Gharb al-Ândalus", *Xelb*, vol. 10, p. 455-476, 2010.

Caballero Zoreda L., "El dibujo arqueológico: notas sobre el registro gráfico en arqueología", *Papeles del partal*, vol. 3, p. 75-95, 2006.

Cabré M. T., "Constituer un corpus de textes de spécialité", *Cahier du CIEL*, vol. 2007-2008, p. 37-56, 2008.

Cabré M. T., Palatresi J., "Acquisition of terminological data from text: approaches", *in* R. H. Gouws, U. Heid, W. Schweickard, H. E. Wiegand (eds), *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume*, De Gruyter Mouton, Berlin, p. 1486-1497, 2013.

Carvajal López J. C., "The archaeology of al-Andalus: past, present and future", *Medieval Archaeology*, vol. 58, p. 318-339, 2014.

Cimiano P., McCrae J. P., Buitelaar P., Lexicon model for ontologies: community report, Technical report, Ontology-Lexicon Community Group, May, 2016.

Cimiano P., McCrae J. P., Rodríguez-Doncel V., Gornostay T., Gómez-Pérez A., Siemoneit B., Lagzdins A., "Linked terminologies: applying linked data principles to terminological resources", *in* I. Kosem, M. Jakubíček, J. Kallas, S. Krek (eds), *Electronic lexicography in the 21st century: linking lexical data in the digital age: Proceedings of the eLex 2015 conference*, Trojina, Ljubljana, p. 504-516, 2015.

Coll Conesa J., Martí Oltra J., Pascual Pacheco J., *Cerámica y cambio cultural: el tránsito de la Valencia islámica a la cristiana*, Dirección General de Bellas Artes y Archivos, Madrid, 1988.

Condamines A., "Terminological knowledge bases", *in* P. A. Fuertes-Olivera (ed.), *The Routledge Handbook of Lexicography*, Routledge, London, p. 335-350, 2018.

Costa R., Pressupostos teóricos e metodológicos para a extracção automática de unidades terminológicas multilexémicas, Doctoral thesis, Universidade Nova de Lisboa, 2001.

Costa R., "Terminology and specialised lexicography: two complementary domains", *Lexicographica*, vol. 29, n^{o} 1, p. 29-42, 2013.

Covaneiro J., Fernandes I. C., Gómez Martínez S., Gonçalves M. J., Inácio I., Santos C., Coelho C., Liberato M., Bugalhão J., Catarino H., Cavaco S., "Cerâmica islâmica em Portugal : 150 anos de investigação", *Arqueologia em Portugal : 150 anos*, Associação dos Arqueólogos Portugueses, Lisboa, p. 73-80, 2013.

Cruse D. A., *Lexical semantics*, Cambridge University Press, Cambridge, 1986.

Doerr M., "Tailoring the Conceptual Model to archaeological requirements", *ARIADNE: the way forward to digital archaeology in Europe*, ARIADNE, [S.l.], p. 65-74, 2014.

Faber P., Léon-Araúz P., Reimerink A., "Representing environmental knowledge in EcoLexicon", *in* E. Bárcena, T. Read, J. Arús (eds), *Languages for specific purposes in the digital era*, Springer, Cham, p. 267-301, 2014.

Gangemi A., "Dolce+D&S Ultralite and its main ontology design patterns", *in* P. Hitzler, A. Gangemi, A. Janowicz, A. A. Krisnadhi, V. Presutti (eds), *Ontology engineering with ontology design patterns: foundations and applications*, IOS Press, Amsterdam, p. 81-103, 2016.

Gomes R. V., *Silves (Xelb), uma cidade do Gharb Al-Andalus: território e cultura*, Instituto Português de Arqueologia, Lisboa, 2002.

Gomes R. V., *Silves (Xelb), uma cidade do Gharb Al-Andalus: a alcáçova*, Instituto Português de Arqueologia, Lisboa, 2003.

Gomes R. V., *Silves (Xelb), uma cidade do Gharb Al-Andalus: o núcleo urbano*, Instituto Português de Arqueologia, Lisboa, 2006.

Gomes R. V., *Silves (Xelb), uma cidade do Gharb Al-Andalus: a zona da Arrochela, espaços e quotidianos*, Instituto Português de Arqueologia, Lisboa, 2011.

Gómez Martínez S., "Contenedores de fuego en el Garb al-Andalus", *in* V. O. Jorge (ed.), *Actas do 3º Congresso de Arqueologia Peninsular*, vol. 7, ADECAP, Porto, p. 421-434, 2000.

Gómez Martínez S., La cerámica islámica de Mértola: producción y comercio, Doctoral thesis, Universidad Complutense de Madrid, 2004.

Guarino N., Musen M., "Applied ontology: focusing on content", *Applied ontology*, vol. 1, nº 1, p. 1-5, January, 2005.

Hitzler P., Krötzsch M., Rudolph S., *Foundations of semantic web technologies*, CRC Press, Boca Raton, 2010.

Jockers M. L., *Macroanalysis: digital methods and literary history*, University of Illinois Press, Urbana, Il, 2013.

L'Homme M. C., "A lexico-semantic approach to the structuring of terminology", *Proceedings of CompuTerm 2004*, COLING, Geneva, p. 7-13, 2004.

Mascardi V., Cordì V., Rosso P., "A comparison of upper ontologies", *Proceedings of WOA 2007*, Seneca Edizioni, Torino, p. 55-64, 2007.

Masolo C., Borgo S., Gangemi A., Guarino N., Oltramari A., Schneider L., WonderWeb deliverable D17: the WonderWeb library of foundational ontologies: preliminary report, Technical report, The WonderWeb Project, 2003.

Meeks E., Weingart S. B., "The digital humanities contribution to topic modeling", *Journal of Digital Humanities*, 2012.

Melby A. K., "Terminology in the age of multilingual corpora", *The Journal of Specialised Translation*, vol. 18, p. 7-29, July, 2012.

Meyer I., "Extracting knowledge-rich contexts for terminography: a conceptual and methodological framework", *in* D. Bourigault, C. Jacquemin, M. L'Homme (eds), *Recent advances in computational terminology*, John Benjamins, Amsterdam, p. 279-302, 2001.

Meyer I., Skuce D., Bowker L., Eck K., "Towards a new generation of terminological resources: an experiment in building a terminological knowledge base", *COLING '92 Proceedings of the 14th conference on Computational linguistics*, vol. 3, Association for Computational Linguistics, Stroudsburg, PA, p. 956-960, 1992.

Munn K., Smith B. (eds), *Applied ontology: an introduction*, Ontos, Heusenstamm, 2008.

Nazarenko A., Hamon T., “Structuration de terminologie : quels outils pour quelles pratiques?”, *TAL*, vol. 43, n° 1, p. 7-18, 2002.

NF ISO 704, *Terminology Work – Vocabulary*, AFNOR, La Plaine Saint-Denis, 2009.

Prieto Velasco J., Faber P., “Graphical information”, *in* P. Faber (ed.), *A cognitive linguistics view of terminology and specialized language*, De Gruyter, Berlin, p. 225-248, 2012.

Roche C., “Ontoterminology: how to unify terminology and ontology into a single paradigm”, *LREC 2012*, ELRA, Paris, p. 2626-2630, 2012.

Roche C., “Ontological definition”, *in* H. J. Kockaert, F. Steurs (eds), *Handbook of terminology: vol. 1*, John Benjamins, Amsterdam, p. 128-152, 2015.

Rosselló-Bordoy G., *Ensayo de sistematización de la cerámica árabe en Mallorca*, Institut d’Estudis Baleàrics, Palma de Mallorca, 1978.

Rosselló-Bordoy G., *El nombre de las cosas en al-Andalus: una propuesta de terminología cerámica*, Museo de Mallorca, Palma de Mallorca, 1991.

Rychlý P., “A lexicographer-friendly association score”, *in* P. Sojka, A. Horák (eds), *Proceedings of Recent Advances in Slavonic Natural Language Processing*, Masaryk University, Brno, p. 6-9, 2008.

Santos C., Costa R., “Domain specificity: semasiological and onomasiological knowledge representation”, *in* H. J. Kockaert, F. Steurs (eds), *Handbook of terminology*, vol. 1, John Benjamins, Amsterdam, p. 153-179, 2015.

Silva R., Gestão de terminologia pela qualidade: processos de validação, Doctoral thesis, Universidade Nova de Lisboa, 2014.

Sure Y., Staab S., Studer R., “Ontology engineering methodology”, *in* S. Staab, R. Studer (eds), *Handbook on ontologies*, 2nd ed edn, Springer, Berlin, p. 135-152, 2009.

Torres C., Gómez Martínez S., Ferreira M. B., “Os nomes da cerâmica medieval: inventário de termos”, *Actas das 3as Jornadas de Cerâmica Medieval e Pós-Medieval*, Câmara Municipal de Tondela, Tondela, p. 125-134, 2003.

Vallejo Triano A., Escudero Aranda J., “Aportaciones para una tipología de la cerámica común califal de Madinat al-Zahra”, *Arqueología y territorio Medieval*, vol. 6, p. 133-176, 1999.

W3C OWL Working Group, OWL 2 Web Ontology Language document overview: 2nd ed, Technical report, W3C OWL Working Group, December, 2012.

Zozaya J., “Una discusión recuperada: candiles musulmanes de disco impreso”, *Arqueología y territorio medieval*, vol. 6, p. 261-278, 1999.

Zozaya J., “Candiles de piquera”, *Tierras del olivo*, Fundación El Legado Andalusí, Granada, p. 125-135, 2007.