Towards a terminological knowledge base on Islamic archaeology: linguistic and conceptual aspects

Bruno Almeida*, Rute Costa*

*NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Portugal
{brunoalmeida,rute.costa}@fcsh.unl.pt
http://clunl.fcsh.unl.pt

Abstract. This paper describes work carried out for the OntoAndalus project at NOVA CLUNL. This project aims at establishing the foundations for a terminological knowledge base (TKB) on Islamic archaeology. The main part of the work carried out so far is the development of an ontology of artefact types in al-Andalusian pottery studies, which was done by reusing the DOLCE+DnS Ultralite foundational ontology. Subsequently, Portuguese and Spanish terms for artefact types were extracted from a corpus of specialised texts and represented by means of lexical networks. The Lexicon Model for Ontologies (Lemon), recently developed by a W3C Community Group, was put forward as a promising framework for integrating language-specific and language-independent information in a future TKB on the domain. The possibility of aligning OntoAndalus with Lemon is also discussed in this paper.

1 Introduction

Developing multilingual terminological resources in the Semantic Web and integrating them in the Linked Open Data Cloud have motivated closer ties between terminology work and ontology development. The recognition of applied ontology as a domain in its own right and the acknowledgement of computational ontologies as an integral part of the Semantic Web have brought new light to the foundations of terminology as a discipline (Durán-Muñoz and Bautista-Zambrana, 2013; Roche, 2012; Santos and Costa, 2015; Temmerman and Kerremans, 2003). Furthermore, the development of ontologies representing shared domain knowledge has been featured in several recent terminological projects, from healthcare to cultural heritage (Brin-Henry et al., 2019; Roche et al., 2019).

The notion of a knowledge-based and computational terminology is, however, far from new, having an important precedent in the notion of 'terminological knowledge base' (TKB) put forward in the 1990's (Meyer et al., 1992). The development of such resources is often complex and relies on interdisciplinary work involving, for example, terminologists/linguists, domain experts and computer scientists. TKBs, however, have a number of advantages, such as the possibility of drafting natural language definitions based on formal descriptions of classes and other ontological predicates, which should lead to more consistent terminological resources. Furthermore, TKBs may provide richer conceptual and linguistic information about specialised domains and their terms in multiple languages.

This paper describes work carried out in the context of the OntoAndalus project, with the purpose of establishing the foundations of a TKB in a subdomain of Islamic archaeology, namely the pottery artefacts of al-Andalus. This work was originally motivated by terminological issues in the community of Portuguese and Spanish archaeologists working on this domain, particularly with respect to the harmonisation of terms denoting artefact types in pottery studies (Bugalhão et al., 2010; Gómez Martínez, 2004; Torres et al., 2003). In order to have a TKB interoperable in the Semantic Web, it was decided to develop a domain ontology focussing on the relevant artefact types, which was eventually named OntoAndalus. [1] Further work was carried out with regard to the extraction and representation of artefact designations in Portuguese and Spanish, as well as on their relationship with the relevant concepts of the domain, already present in OntoAndalus. Lemon, the Lexicon Model for Ontologies developed by the W3C Ontology-Lexica Community Group (2016), was adopted for this purpose, since – as will be shown – it allows to represent rich semantic and grammatical information about terms and their relationship to ontology predicates.

2 Overview of OntoAndalus

OntoAndalus is an ontology of relevant artefact types in pottery studies of al-Andalus. The ontology was developed based on the interpretation of a bilingual (Portuguese and Spanish) corpus of specialised texts in the domain, in which the artefact typologies put forward by the CIGA group (Bugalhão et al., 2010) in Portugal and by Rosselló-Bordoy (1978, 1991) in Spain were paramount. Other sources of information included English textbooks and other reference works on archaeology and pottery analysis, such as Rice (1987).

DOLCE+DnS Ultralite (DUL) was selected as a foundational ontology for the development of OntoAndalus following a review of literature on philosophy and applied ontology (Almeida, 2019). DUL is a version of DOLCE for the Semantic Web with the addition of the Descriptions and Situations ontology (DnS) for modelling non-physical objects (e.g. organisations, tasks, mental concepts) (Gangemi, 2016). DUL as a number of practical advantages, such as its stability, completeness and being made available in OWL format. More importantly, DUL allowed for a rich conceptualisation of the domain, in which important concepts in archaeology were sufficiently explicated (e.g. artefact, part, quality, function) (Almeida and Costa, 2019).

OntoAndalus was developed in Protégé (Musen, 2015) by importing the DUL ontology through the `owl:imports` property. At this time, OntoAndalus consists of 161 classes, 30 object properties and 135 individuals (excluding elements in the DUL namespace). OntoAndalus includes 72 artefact types, which are organised in the functional collections shown in Fig. 1.

In the following section, lighting artefacts will be outlined as a case study of how artefact types can be modelled in OntoAndalus.

2.1 Modelling artefact types in OntoAndalus

Lighting artefacts include some of the more representative objects in the archaeology of al-Andalus. According to Gómez Martínez (2004), lighting artefacts may fall within four pottery

1. OntoAndalus is made available through `https://github.com/brunoalmeida81/OntoAndalus`.

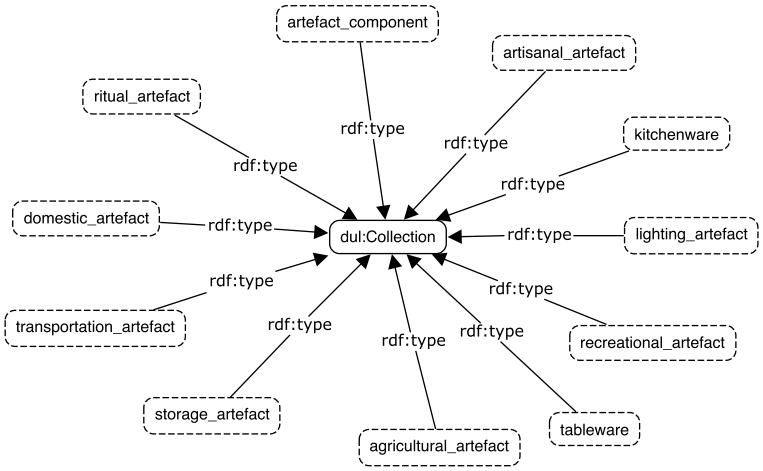

FIG. 1 – *Functional collections declared in OntoAndalus. Dashed boxes denote individuals and solid boxes denote classes.*

types (or *series*), which are denoted by the following Spanish terms: (i) *candil*; (ii) *almenara* or *policandela*; (iii) *lamparilla*; and (iv) *fanal* or *linterna* (Fig. 2).

In OntoAndalus, the above-mentioned pottery types were represented as classes subsumed by `dul:DesignedArtifact`. The following English identifiers were attributed to each class: (i) `Lamp`; (ii) `MultipleLamp`; (iii) `StationaryLamp`; and (iv) `Lantern`. Each class was then described according to the following pattern: **superordinate class + collection + function + part(s) or component(s)**. This pattern was also the basis for drafting natural language definitions, which were associated to each class via the `skos:definition` property. The formal and natural language definitions or descriptions of each class are the following:

Lamp (*candil*). *Def.* Artefact for lighting in closed spaces composed of at least one spout and a single chamber for liquid fuel.

```
Lamp rdfs:subClassOf dul:DesignedArtifact

Lamp rdfs:subClassOf dul:isMemberOf some lighting_artefact

Lamp rdfs:subClassOf hasFunction some containing_fire_for
_lighting_in_closed_spaces_using_liquid_fuel

Lamp rdfs:subClassOf hasComponent exactly 1 LampFuelChamber

Lamp rdfs:subClassOf hasComponent min 1 Spout
```

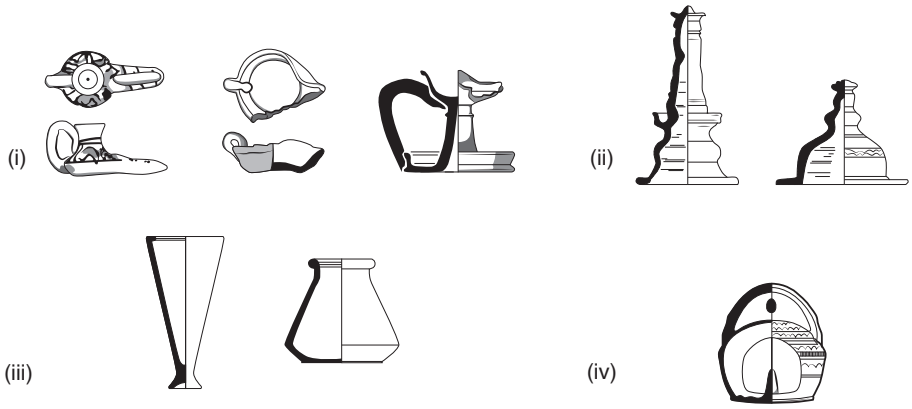

FIG. 2 – *Lighting artefact types. Vector illustrations by Mariana Tavares.*

Explanation: The function of the *candil* was that of domestic lighting. It was typically composed of a fuel chamber, a single spout for holding a wick and a handle. There are, however, objects in this series which were designed to be suspended and, as a consequence, do not have any handles. Furthermore, less typical instances of this type have more than one spout (Gómez Martínez, 2004; Rosselló-Bordoy, 1991).

MultipleLamp (*almenara* or *policandela*). *Def.* Artefact for stationary lighting in closed spaces composed of more than one chamber for liquid fuel unified by a structure.

```
MultipleLamp rdfs:subClassOf dul:DesignedArtifact

MultipleLamp rdfs:subClassOf dul:isMemberOf some lighting
_artefact

MultipleLamp rdfs:subClassOf hasFunction some containing
_fire_for_stationary_lighting_in_closed_spaces_using
_liquid_fuel

MultipleLamp rdfs:subClassOf hasComponent min 2
LampFuelChamber

MultipleLamp rdfs:subClassOf hasComponent some
MultipleLampStructure
```

Explanation: Not much is known on this artefact type, whose instances would have consisted of several fuel chambers unified by a base for stationary or, possibly, suspended lighting (Gómez Martínez, 2004; Rosselló-Bordoy, 1991).

StationaryLamp (*lamparilla*). *Def.* Artefact for stationary lighting in closed spaces composed of a single chamber for liquid fuel.

```
StationaryLamp rdfs:subClassOf dul:DesignedArtifact

StationaryLamp rdfs:subClassOf dul:isMemberOf some lighting
_artefact

StationaryLamp rdfs:subClassOf hasFunction some containing
_fire_for_stationary_lighting_in_closed_spaces_using
_liquid_fuel

StationaryLamp rdfs:subClassOf hasComponent exactly 1
LampFuelChamber
```

Explanation: This possible lighting artefact type would have been left on a table stand or on a discoidal plate, which could have held several instances (Gómez Martínez, 2004; Vallejo Triano and Escudero Aranda, 1999). In OntoAndalus, the `StationaryLamp` class is further divided according to shape: either inverted frustum or bifrustum-shaped.

Lantern (*fanal* or *linterna*). *Def.* Artefact for lighting in open spaces composed of a single chamber for solid fuel.

```
Lantern rdfs:subClassOf dul:DesignedArtifact

Lantern rdfs:subClassOf dul:isMemberOf some lighting
_artefact

Lantern rdfs:subClassOf hasFunction some containing
_fire_for_lighting_in_open_spaces_using
_solid_fuel

Lantern rdfs:subClassOf hasComponent exactly 1
LanternFuelChamber
```

Explanation: The function of this artefact was to provide a light source outdoors by using a solid fuel, such as wax. Its typical features include having a closed shape, globular body and a single handle, which are based on isolated findings (Bugalhão et al., 2010; Gómez Martínez, 2004).

Since the `Lamp` class corresponds to the most widely studied type of lighting artefacts in the archaeology of al-Andalus, the ontology puts forward a classification of possible subtypes according to multiple criteria of subdivision, namely: (i) vessel form (e.g. open, closed); (ii) type of spout (e.g. channelled, pinched); (iii) inclusion of a discus or neck; and (iv) inclusion of a tall foot. Relevant classes were defined based on these criteria and were made pairwise

disjoint. Since the more salient criterion is that of vessel form (i.e. open or closed), this was chosen as the primary criterion for subdividing the `Lamp` class (Bugalhão et al., 2010; Gómez Martínez, 2004). Only the disjoint classes `ClosedLamp` and `OpenLamp` are further subdivided in the asserted hierarchy of OntoAndalus. The full classification can be inferred by a reasoner (Figs. 3 and 4)

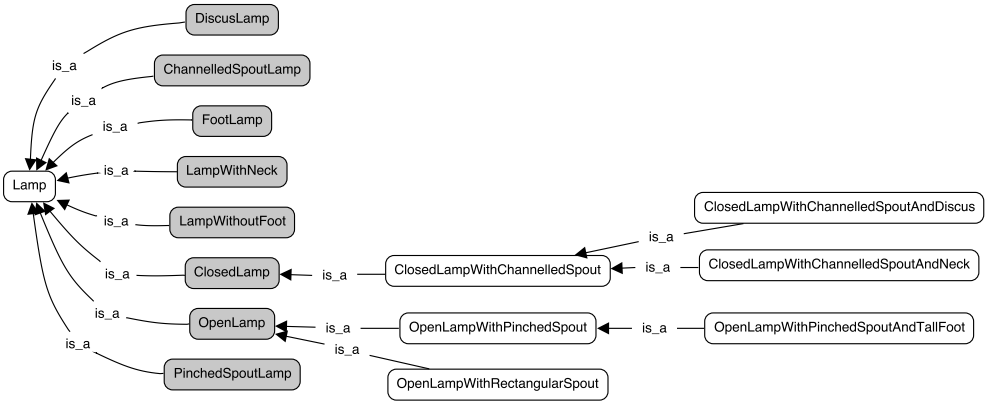

FIG. 3 – *Asserted hierarchy of* `Lamp`*. Gray boxes indicate defined classes.*

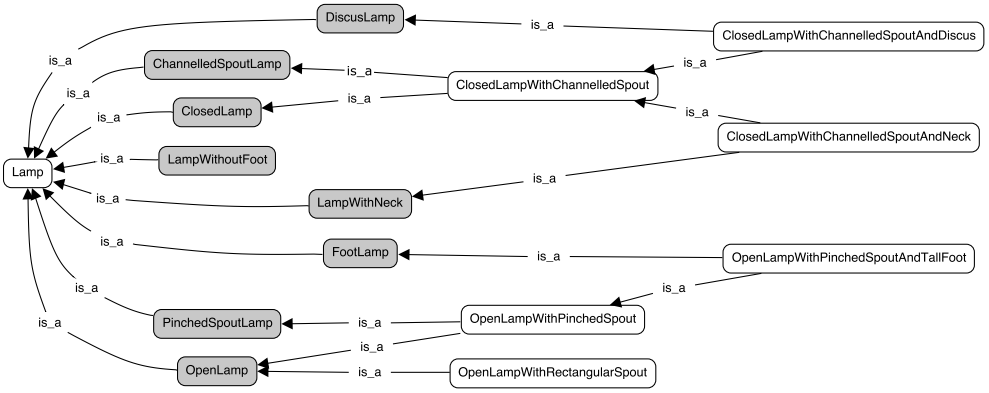

FIG. 4 – *Inferred hierarchy of* `Lamp`*.*

3 Representation of artefact terms

One of the more important steps in our work consisted of representing the extracted Portuguese and Spanish terms for lighting artefacts by means of lexical networks. [2] In terminology work, these networks are considered to be useful for organising and representing language-specific information about terms and other linguistic units (Santos and Costa, 2015).

Figs. 5 and 6 show the more relevant terms for lighting artefact types in Portuguese and Spanish. The lexical relations used in these networks are those of taxonomy (i.e. a specialisation of hyponymy between terms denoting types) and synonymy (Cruse, 1986). Terms arising from the same criterion of subdivision are represented through divided taxonomic arcs, for example *candil* and *candeia* in Fig. 5, which are both motivated by vessel form.

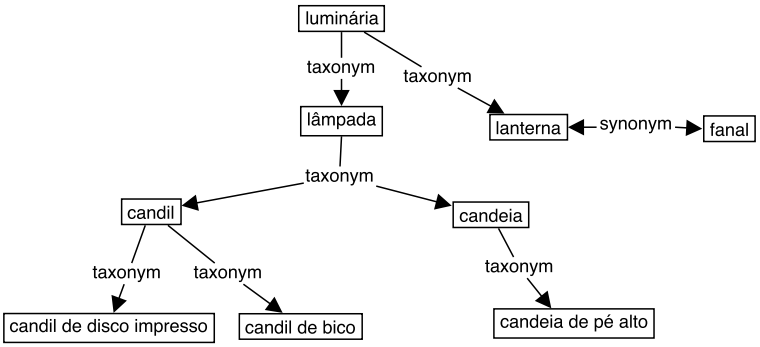

FIG. 5 – *Lighting artefacts terms in Portuguese.*

The question remains of how to relate information about terms with language-independent information about the concepts of the domain in a multilingual resource. A promising model for achieving this is Lemon, which will be briefly described in the following section.

3.1 Integrating linguistic and conceptual information with Lemon

Lemon is an acronym for 'Lexicon Model for Ontologies'. The model has been in development over the past few years (W3C Ontology-Lexica Community Group, 2016), and has been made available through several OWL files. [3]

The purpose of Lemon is to provide linguistic grounding for ontologies, which is often nonexistent or limited to labels represented by means of the `rdfs:label` annotation property. In order to achieve this purpose, Lemon makes use of the following core modules: [4]

— *Ontolex*. For relating a lexicon with an ontology;

— *Synsem*. For representing information at the syntactic and semantic levels;

2. A thorough description of the corpus as well as of the tools and methods used for terminology extraction is provided in a forthcoming article (Almeida et al., 2019).

3. `https://github.com/ontolex/ontolex/`.

4. Recently, Lemon has been extended by means of lexicog, the lexicographic module of Lemon, which works in conjunction with the modules presented here (particularly ontolex) for describing dictionaries and similar resources (Bosque-Gil et al., 2019).

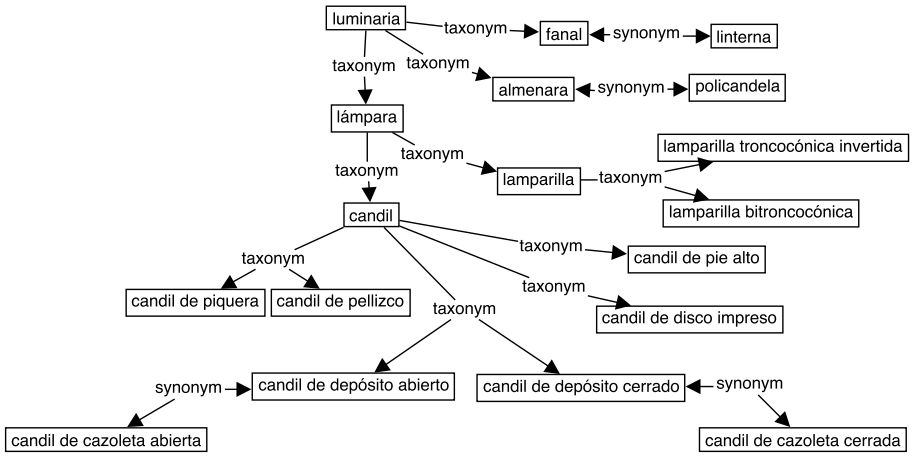

FIG. 6 – *Lighting artefacts terms in Spanish.*

— *Decomp*. For representing information on the decomposition of multiword expressions;
— *Vartrans*. For representing information regarding variation and translation;
— *Lime*. For representing linguistic metadata.

Ontolex is the primary module of Lemon, since it allows to draw relations between lexical entries (i.e. single words, multiword expressions and affixes) and ontology elements (e.g. classes, object properties, individuals). A lexical entry can be realised as a series of forms in a language, including its canonical form – typically the lemma – and other forms (e.g. the plural of nominals). Each form may have several written and/or phonetic representations.

For our present purposes, the relationship between lexical entries and ontology elements can be drawn directly, via the `denotes`/`isDenotedBy` object properties or indirectly through instances of the `LexicalSense` class.[5] The latter can be linked to instances of *LexicalEntry* by means of the `sense`/`isSenseOf` object properties and, subsequently, to ontology elements through the `reference`/`isReferenceOf` object properties. While the direct method is considerably simpler, using the `LexicalSense` class and the previously mentioned object properties allows to model lexicosemantic relations (e.g. taxonomy, synonymy), which may hold between instances of `LexicalSense`.

3.2 Aligning OntoAndalus with Lemon

The possibility of including predicates from Lemon directly in OntoAndalus would allow for a detailed representation of linguistic information, namely with regard to terms in different languages and their lexicosemantic relations. This would involve importing predicates from

5. There is also the possibility of using the `LexicalConcept` class and its associated object properties (i.e. `evokes`/`isEvokedBy` and `concept`/`isConceptOf`), but this is not relevant to our purposes, since OntoAndalus already includes the relevant concepts of our domain of interest.

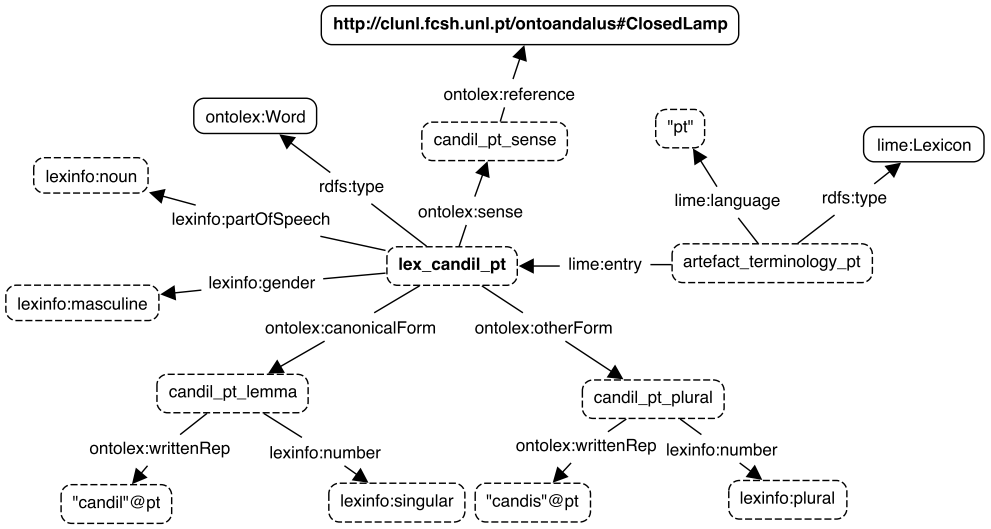

FIG. 7 – *Entry for the Portuguese term* candil.

(at least) the ontolex, lime and vartrans modules, but also from the lexinfo ontology, which provides several categories for Lemon (e.g. the `hyponym` relation between lexical senses). [6]

Fig. 7 shows a possible entry for the Portuguese term *candil* using predicates from the above-mentioned namespaces. In this example, the terminology of artefacts in Portuguese is declared as an instance of `lime:Lexicon`. Grammatical information about the entry includes its part of speech, gender, canonical and plural forms. Finally, semantic information includes reference to the `ClosedLamp` class within OntoAndalus.

With regard to lexicosemantic relations, Fig. 8 shows an instance of the equivalence relation between terms in different languages, as well as an instance of the hyponymy relation. The former is drawn between the Portuguese term *candil* and the Spanish term *candil de depósito cerrado*. This is done by simply pointing the reference of the senses of both terms to the same ontology element, namely the `ClosedLamp` class. Using the relevant object properties in the vartrans module, the hyponymy relation is established between the Portuguese terms *candil* and *candil de disco impresso*. [7] In this example, the sense relation is reified, i.e. represented as an individual, which allows to identify its source (the sense of *candil*), target (the sense of *candil de disco impresso*) and category (hyponymy).

Lemon, therefore, allows to represent diverse information about the terms, including grammatical information (e.g. part of speech, gender) and semantic information (e.g. lexicosemantic relations, reference to ontology elements).

6. `https://lexinfo.net`.

7. Since the more specialised taxonomy relation is not present in lexinfo, hyponymy was used instead in this example.

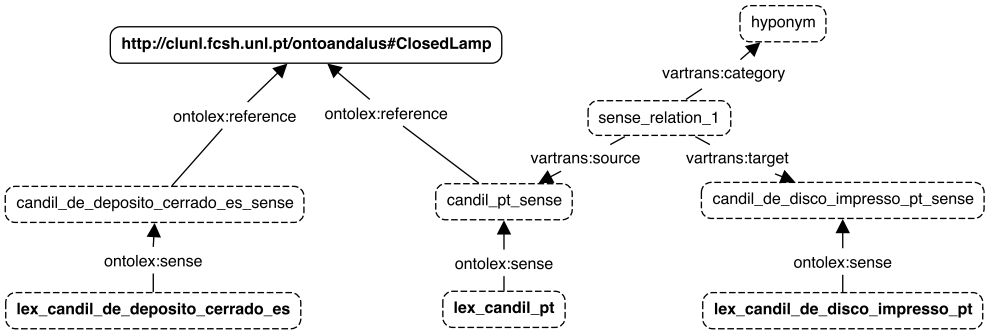

FIG. 8 – *Representing lexicosemantic relations with Lemon.*

4 Conclusion

This paper provided an overview of work carried out for establishing the foundations of a terminological knowledge base in Islamic archaeology. Domain knowledge was represented by means of OntoAndalus, which focusses on relevant artefact types in al-Andalusian pottery studies. OntoAndalus reuses DOLCE+DnS Ultralite, which allows for a detailed representation of domain knowledge, which was exemplified with the case study of lighting artefacts. Information regarding the artefact designations in Portuguese and Spanish was then extracted from a specialised corpus and the Lexicon Model for Ontologies (Lemon) was used in order to represent information about the terms and their relation to the relevant concepts within the domain ontology. Lemon was shown to be capable of representing rich grammatical and semantic information about the terms, and remains a promising model for terminographic applications.

Acknowledgement. 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/03213/2020.

References

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

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

Almeida, B., R. Costa, and C. Roche (2019). The names of lighting artefacts: extraction and representation of Portuguese and Spanish terms in the archaeology of al-Andalus. *Revue TAL 60*(3), 113–137.

Bosque-Gil, J., J. Gracia, J. P. McCrae, P. Cimiano, S. Stolk, F. Khan, K. Depuydt, J. Does, F. Frontini, and I. Kernerman (2019). The OntoLex Lemon Lexicography Module: final community report. Technical report, Ontology-Lexicon Community Group.

Brin-Henry, F., R. Costa, and S. Desprès (2019). TemPO: towards a conceptualisation of pathology in speech and language therapy. In *Proceedings of the TALN-RECITAL 2019 conference*, Toulouse, pp. 69–80.

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

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

Durán-Muñoz, I. and M. Bautista-Zambrana (2013). Applying ontologies to terminology: advantages and disadvantages. *Hermes 51*, 65–77.

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

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

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

Musen, M. A. (2015). The Protégé project: a look back and a look forward. *AI Matters 1*(4), 4–12.

Rice, P. (1987). *Pottery analysis: a sourcebook*. Chicago, Il: The University of Chicago Press.

Roche, C. (2012). Ontoterminology: how to unify terminology and ontology into a single paradigm. In *LREC 2012*, Madrid, pp. 2626–2630.

Roche, C., R. Costa, S. Carvalho, and B. Almeida (2019). Knowledge-based terminological e-dictionaries: the EndoTerm and al-Andalus pottery projects. *Terminology 25*(2), 262–294.

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

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

Santos, C. and R. Costa (2015). Domain specificity: semasiological and onomasiological knowledge representation. In H. J. Kockaert and F. Steurs (Eds.), *Handbook of terminology*, Volume 1, pp. 153–179. Amsterdam: John Benjamins.

Temmerman, R. and K. Kerremans (2003). Termontography: ontology building and the sociocognitive approach to terminology description. In *Proceedings of CIL17*, Prague. Matfyzpress.

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

Vallejo Triano, A. and J. Escudero Aranda (1999). Aportaciones para una tipología de la

cerámica común califal de Madinat al-Zahra. *Arqueología y territorio Medieval 6*, 133–176.

W3C Ontology-Lexica Community Group (2016). Lexicon model for ontologies: community report. Technical report, W3C Ontology-Lexica Community Group.

Résumé

Cet article décrit le travail effectué dans le cadre du projet OntaAndalus développé au NOVA CLUNL. Ce projet a pour objectif de jeter les bases d'une base de connaissances terminologiques (BCT) concernant l'archéologie islamique. La partie centrale de cette recherche a été le développement d'une ontologie des types d'artefacts de la poterie de l'al-Andalus qui a été réalisée en réutilisant l'ontologie DOLCE + DnS Ultralite. Par la suite, les termes en portugais et en espagnol désignant les types d'artefacts ont été extraits d'un corpus de textes spécialisés et représentés par des réseaux lexicaux. Le Lexicon Model for Ontologies (Lemon), récemment développé par un groupe communautaire du W3C, est présenté comme un cadre prometteur pour l'intégration d'informations linguistiques et extralinguistiques dans une future BCT du domaine. La possibilité d'aligner OntoAndalus et Lemon est abordée dans cet article.