



#14

Terminologia e conhecimento arqueológico: contributos interdisciplinares para o estudo da cerâmica do al-Andalus¹

Bruno Almeida | NOVA CLUNL – Centro de Linguística da Universidade NOVA de
Lisboa | brunoalmeida@fcsb.unl.pt

¹ Investigação financiada por fundos nacionais através da FCT - Fundação para a Ciência e Tecnologia, como parte do projeto Centro de Linguística da Universidade NOVA de Lisboa – UID/LIN/03213/2020.

Resumo: O presente trabalho tem como objetivo divulgar investigação recente sobre terminologia cerâmica junto da comunidade de investigadores em história e arqueologia do al-Andalus. Em particular, pretende-se demonstrar as vantagens de uma abordagem baseada na dupla dimensão da terminologia — linguística e conceptual — no estabelecimento das bases para um futuro recurso terminológico na Web. No que diz respeito à área do conhecimento, é descrita uma ontologia de domínio — OntoAndalus — com ênfase na tipologia dos artefactos de iluminação. Relativamente à vertente linguística, são abordados tópicos como a constituição de corpora no domínio, a extração de termos e a sua representação computacional.

Palavras-chave: terminologia; ontologia; cerâmica; al-Andalus.

Resumen: Este trabajo tiene como objetivo dar a conocer investigaciones recientes sobre terminología cerámica a la comunidad de investigadores en historia y arqueología de al-Andalus. En particular, se pretende demostrar las ventajas de la doble dimensión de la terminología — lingüística y conceptual — al establecer las bases para un futuro recurso terminológico en la Web. La tipología de los artefactos de iluminación y sus designaciones en portugués y castellano serán resaltadas. En cuanto al área de conocimiento, se describe una ontología de dominio — OntoAndalus — con énfasis en la tipología de artefactos de iluminación. En cuanto al aspecto lingüístico, se abordarán temas como la constitución de corpus en el dominio, la extracción de términos y su representación computacional.

Palabras-clave: terminología; ontología; cerámica; al-Andalus.

1. Introdução

A palavra “terminologia” designa, no sentido mais corrente, o conjunto de termos e conceitos referentes a domínios especializados, sendo uma das características mais salientes do uso das línguas em contextos de especialidade. “Terminologia” designa ainda a área interdisciplinar baseada na recolha, organização, descrição e estudo de termos e conceitos. Neste âmbito, uma das principais atividades consiste na elaboração de recursos terminológicos (p. ex., glossários, vocabulários) de forma a facilitar a comunicação em domínios de especialidade, com vista à formação, atualização e/ou transferência de conhecimento (ISO 704, 2009; ISO 1087, 2019).

As seguintes secções descrevem o trabalho realizado com vista à criação de um recurso terminológico na Web sobre a cerâmica do al-Andalus, entendida como área de estudos da arqueologia islâmica peninsular. Este trabalho baseia-se na dupla dimensão da terminologia — linguística e conceptual —, uma abordagem teórica e metodológica assente na distinção entre um nível de análise linguístico (i.e., específico às línguas) e extralinguístico (i.e., relativo às diversas áreas do conhecimento) (Costa, 2013; Santos e Costa, 2015). O trabalho realizado no âmbito da dimensão conceptual centra-se na modelização da OntoAndalus, uma ontologia de artefactos cerâmicos do al-Andalus, com ênfase para a tipologia dos artefactos de iluminação. No que diz respeito à dimensão linguística, o trabalho realizado centra-se na constituição de um corpus de textos especializados e na subsequente extração e modelização de termos em português e castelhano.

2. Modelização da OntoAndalus

OntoAndalus² é uma ontologia de tipos de artefactos relevantes nos estudos de cerâmica de al-Andalus. A ontologia foi desenvolvida a partir da interpretação de um corpus bilingue (português e castelhano) de textos especializados no domínio, no qual as tipologias de artefactos propostas pelo grupo CIGA (Bugalhão *et al.*, 2010) em Portugal e por Rosselló-Bordoy (1978, 1991) em Espanha desempenharam um papel fundamental. Outras fontes de informação incluíram manuais e obras de referência em língua inglesa sobre arqueologia e análise de cerâmica, como Rice (1987). O desenvolvimento da ontologia é descrito detalhadamente em Almeida e Costa (2020).

A DOLCE Ultralite (DUL) foi seleccionada como ontologia de topo para o desenvolvimento da OntoAndalus após uma revisão da literatura sobre filosofia e ontologia aplicada. A DUL³ é uma versão simplificada da ontologia DOLCE para a Web Semântica, contando ainda com a adição da ontologia de Descrições e Situações (DnS) para modelizar objetos imateriais (p. ex., organizações, tarefas, conceitos mentais) (Gangemi, 2016). A reutilização da DUL apresenta uma série de vantagens, tais como a sua estabilidade, disponibilização livre e expressão em OWL (a linguagem de ontologias para a Web). Mais importante ainda, a DUL permite uma conceptualização rica do domínio, na qual conceitos importantes em arqueologia foram explicitados, por exemplo, artefacto, parte, qualidade e função.

² <https://github.com/brunoalmeida81/OntoAndalus>

³ http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite

A OntoAndalus foi desenvolvida no editor de ontologias Protégé⁴ tendo por base todos os elementos da ontologia DUL. Atualmente, a OntoAndalus consiste em 161 classes, 30 propriedades de objetos e 135 indivíduos (excluindo os elementos definidos pela DUL). A OntoAndalus inclui 72 tipos de artefactos, os quais são organizados nas seguintes coleções funcionais:

- Artefactos de iluminação
- Artefactos de mesa
- Artefactos de cozinha
- Artefactos domésticos
- Artefactos recreativos
- Artefactos rituais
- Artefactos agrícolas
- Artefactos de construção
- Artefactos artesanais
- Artefactos de armazenamento
- Artefactos de transporte
- Componentes de artefactos

2.1. O caso dos artefactos de iluminação

Nesta secção, a classe dos artefactos de iluminação é descrita como estudo de caso da modelização de tipos de artefactos na OntoAndalus. Esta tipologia abrange alguns dos artefactos mais representativos do período islâmico na Península Ibérica (Gómez-Martínez, 2004). Segundo Gómez Martínez, é reconhecida a existência de quatro séries de artefactos de iluminação, denotados pelos seguintes termos em castelhano: (i) “candil”; (ii) “policandela” (ou “almenara”); (iii) “lamparilla”; e (iv) “fanal” (ou “linterna”). A Fig. 1 mostra uma representação gráfica de exemplos típicos de cada série.

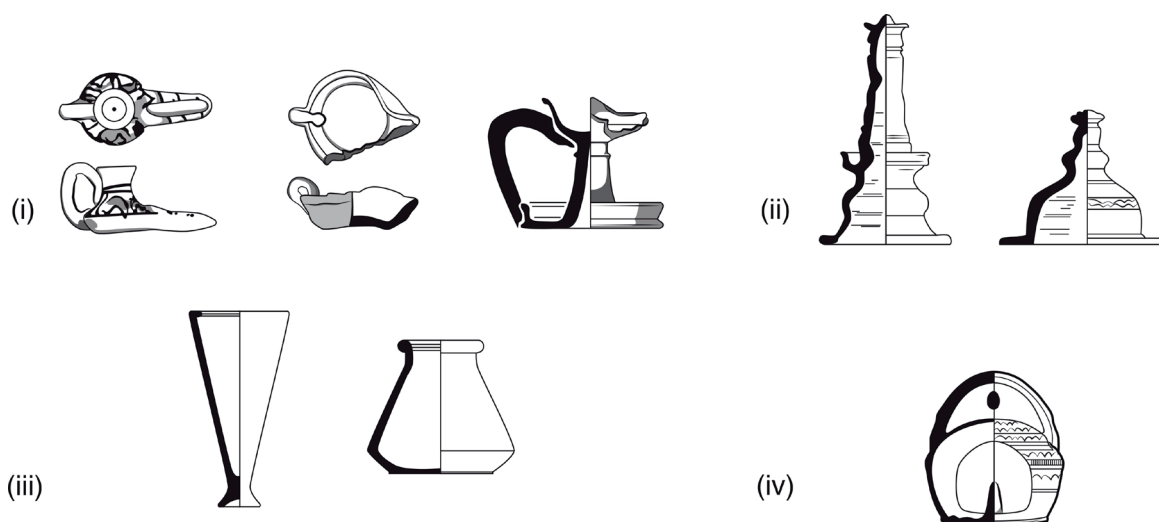


Fig. 1 – Séries de artefactos de iluminação na OntoAndalus.

⁴ <https://protege.stanford.edu/>

As referidas séries de artefactos, assim como os respetivos tipos e subtipos, são representados na OntoAndalus através de classes. Para manter a consistência com a DUL e facilitar a comunicação internacional, é atribuído a cada classe um identificador em inglês, a saber: (i) Lamp; (ii) MultipleLamp; (iii) StationaryLamp; e (iv) Lantern. As definições formais destas classes na OntoAndalus seguem o seguinte padrão: classe superordenada + coleção funcional + função + parte(s) ou componente(s). Os parágrafos seguintes definem cada classe em português tendo como base as descrições e definições formais da ontologia, juntamente com referências a textos relevantes do corpus.

Lamp (candil_{pt}/candeia_{pt}; candil_{es}): artefacto para iluminação em espaços fechados, composto por pelo menos um bico e um único depósito para combustível líquido (Coll Conesa, Martí Oltra e Pascual Pacheco, 1988; Gómez-Martínez, 2004; Rosselló-Bordoy, 1991).

MultipleLamp (almenara_{pt}, almenara_{es}, policandela_{es}): artefacto para iluminação estacionária em espaços fechados, composto por diversos depósitos de combustível líquido unificados por uma estrutura (Gómez-Martínez, 2004; Rosselló-Bordoy, 1991).

StationaryLamp (lamparilla_{es}): artefacto para iluminação estacionária em espaços fechados composto por um único depósito para combustível líquido (Vallejo Triano e Escudero Aranda, 1999).

Lantern (fanal_{pt}, lanterna_{pt}; fanal_{es}, linterna_{es}): artefacto para iluminação em espaços abertos composto por um único depósito para combustível sólido (Bugalhão *et al.*, 2010; Gómez-Martínez, 2004).

A classe Lamp é a parte mais complexa da ontologia no que diz respeito aos artefactos de iluminação (Fig. 2).

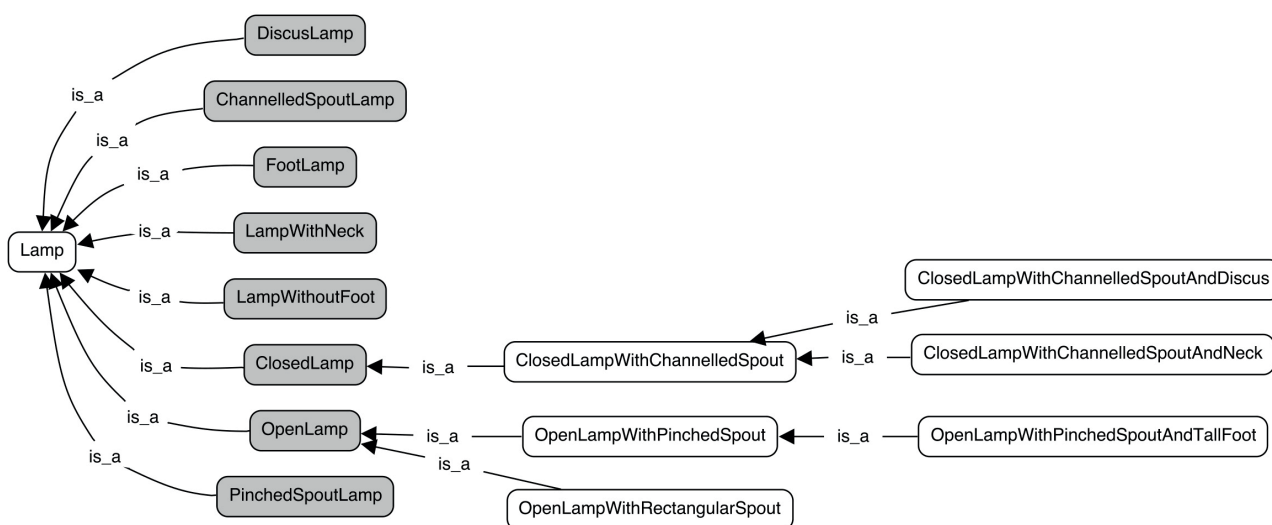


Fig. 2 – Tipos e subtipos da classe Lamp (“candil”/“candeia”).

A OntoAndalus inclui quatro critérios de subdivisão da referida classe: (i) forma (p. ex., aberta, fechada); (ii) tipo de bico (p. ex., pinçado, de canal); (iii) inclusão de um disco ou colo e (iv) inclusão de um pé alto. Os múltiplos critérios de subdivisão são representados por classes definidas e disjuntas par a par.

A forma é geralmente considerada o critério mais saliente para distinguir entre estes tipos de luminárias (Bugalhão *et al.*, 2010), tendo sido escolhido como o principal critério de subdivisão da classe Lamp. Assim, apenas as classes disjuntas ClosedLamp e OpenLamp são subdivididas na hierarquia declarada de OntoAndalus, sendo que a restante classificação pode ser inferida por um raciocinador.

Vimos como os tipos de artefactos podem ser modelados em OntoAndalus para constituir a camada extralinguística de um recurso terminológico. Esta camada inclui as descrições e definições formais de classes, as quais servem como base para a redação de definições em língua natural. A secção seguinte descreverá o trabalho realizado no que diz respeito à representação de termos portugueses e castelhanos extraídos do corpus de textos especializados.

3. Termos para tipos de artefactos em português e castelhano

3.1. Constituição de um corpus especializado em cerâmica do al-Andalus

A constituição de um corpus serviu dois objetivos em relação ao trabalho descrito neste artigo. Por um lado, o corpus facilitou a compreensão do domínio e auxiliou a modelização da OntoAndalus. Por outro lado, o corpus foi fundamental para a extração de informação de natureza linguística. Os critérios para constituição do corpus representados no Quadro 1 foram adotados com base em Bowker e Pearson (2002) e Cabré (2007).

Quadro 1

Critérios para constituição do corpus de textos especializados.

Domínio	Cerâmica do al-Andalus
Línguas	Português, Castelhana
Âmbito temporal	1970 -
Nível de especialização	Médio a alto
Integridade	Textos completos
Meio e modalidade	Textos escritos (em formato digital) com informação visual
Géneros de textos e discursos	Diversos

Os textos que constituem o corpus são principalmente sobre artefactos cerâmicos do al-Andalus e foram originalmente escritos em português ou castelhano. O âmbito temporal justifica-se pelo facto de a arqueologia peninsular islâmica se ter estabelecido como domínio a partir da década de 1970 (Covaneiro *et al.*, 2013). O corpus inclui textos completos com um nível médio a alto de especialização, ou seja, produzidos por — e destinados a — especialistas ou estudiosos do domínio. Tal critério implica a exclusão de obras destinadas ao público em geral, uma vez que a terminologia nelas empregada pode diferir significativamente da empregada nos textos especializados.

Em relação ao meio e à modalidade, o corpus é formado por textos escritos com informação visual. Existem várias razões para o último requisito: (i) importância da modalidade visual no domínio, (ii) auxílio na compreensão do domínio (p. ex., tipos de artefactos, morfologia de vasos) e (iii) trabalho futuro (p. ex., enriquecimento do futuro recurso terminológico com informação visual).

Por último, o corpus é heterogéneo no que diz respeito aos géneros e discursos. Ele abrange discursos transdisciplinares (p. ex., arqueologia, história), bem como vários géneros de texto (p. ex., teses, artigos de periódicos). Esta opção justifica-se com o nosso foco no conhecimento do domínio, o qual pode ser expresso em diversos géneros textuais e discursos.

No que diz respeito à estrutura do corpus, foi elaborada uma estrutura simples de acordo com a língua e o género de texto, baseada no trabalho de Costa (2001). A revisão preliminar da comunicação científica do domínio permitiu identificar os géneros textuais mais significativos, a saber: (i) teses e dissertações, (ii) monografias, (iii) artigos em catálogos, (iv) artigos em revistas científicas, (v) artigos em atas de conferências.

Embora a tipologia proposta seja útil em termos metodológicos, ela apresenta vários problemas. Por um lado, o mesmo texto (ou uma versão muito semelhante) pode abranger mais de um género. Por exemplo, o conteúdo de uma tese pode ser publicado posteriormente como monografia, ou mesmo como uma série de monografias. É o caso da tese de doutoramento de Rosa Varela Gomes. Esta tese, apresentada originalmente em 2000, foi posteriormente publicada em quatro volumes da série monográfica *Trabalhos de Arqueologia*, os quais integram o corpus (Gomes, 2002; 2003; 2006; 2011).

Outra dificuldade foi colocada pelos catálogos de museus e exposições. É frequente incluírem artigos de vários autores, juntamente com a descrição catalográfica dos artefactos. Por este motivo, foi decidido incluir uma categoria separada para artigos de catálogos na estrutura do corpus. As próprias descrições não foram incluídas na tipologia final, uma vez que não tivemos acesso a catálogos digitalizados de artefactos em ambas as línguas de trabalho. No entanto, é comum que outros géneros de textos incluam catálogos numa secção separada. O discurso das descrições catalográficas está, desta forma, representado no corpus.

Os Quadros 2 e 3 mostram a composição dos subcorpora em português e em castelhano. Os dados apresentados foram recolhidos no Sketch Engine, o gestor de corpora utilizado no trabalho desenvolvido neste artigo. Ambos os subcorpora são comparáveis em termos de dimensão: o corpus português tem mais de 670 000 formas de palavras, enquanto o corpus castelhano tem mais de 705 000 formas de palavras.

Existe alguma assimetria quanto ao género textual: as monografias são o género mais representado no corpus português, enquanto as teses e dissertações são predominantes no corpus castelhano. Para além disso, todos os textos em português incluídos nesta última categoria são dissertações de mestrado, enquanto os textos em castelhano são teses de doutoramento. Esta última assimetria é compensada pelo facto de quatro das cinco monografias do corpus português corresponderem à tese de doutoramento de Rosa Varela Gomes.

Quadro 2

Composição do subcorpus em português.

Tipo de texto	Textos (N)	Palavras (N)	Palavras (%)
Teses e Dissertações	6	242,597	36.20
Monografias	5	362,401	54.08
Artigos em revistas	6	40,458	6.04
Artigos em catálogos	1	5,070	0.76
Artigos em conferências	4	19,610	2.93
Total	22	670,136	≈ 100

Quadro 3

Composição do subcorpus em castelhano.

Tipo de texto	Textos (N)	Palavras (N)	Palavras (%)
Teses e Dissertações	3	575,011	81.45
Monografias	1	58,462	8.28
Artigos em revistas	7	51,271	7.26
Artigos em catálogos	1	3,057	0.43
Artigos em conferências	4	18,148	2.57
Total	16	705,949	≈ 100

3.2. Representação e comparação dos termos para artefactos de iluminação em português e castelhano

Uma das atividades mais importantes do nosso trabalho consistiu na extração de termos portugueses e castelhanos para artefactos de iluminação e na sua representação através de redes lexicais. No trabalho terminológico, estas redes são úteis para organizar e representar informação sobre termos e outras unidades linguísticas (Santos e Costa, 2015). O processo de extração dos termos do corpus é descrito em pormenor em Almeida, Costa e Roche (2019).

As Figs. 3 e 4 mostram os termos mais relevantes para tipos de artefactos de iluminação em português e castelhano. As relações lexicais usadas nestas redes são as de taxonomia (i.e., uma especialização da hiponímia entre termos que denotam tipos) e de sinonímia (Cruse, 1986). Os termos motivados pelo mesmo critério de subdivisão são representados por arcos taxonómicos divididos. Por exemplo, na Fig. 3 “candeia” e “candil” são ambos motivados pela forma dos artefactos (aberta ou fechada).

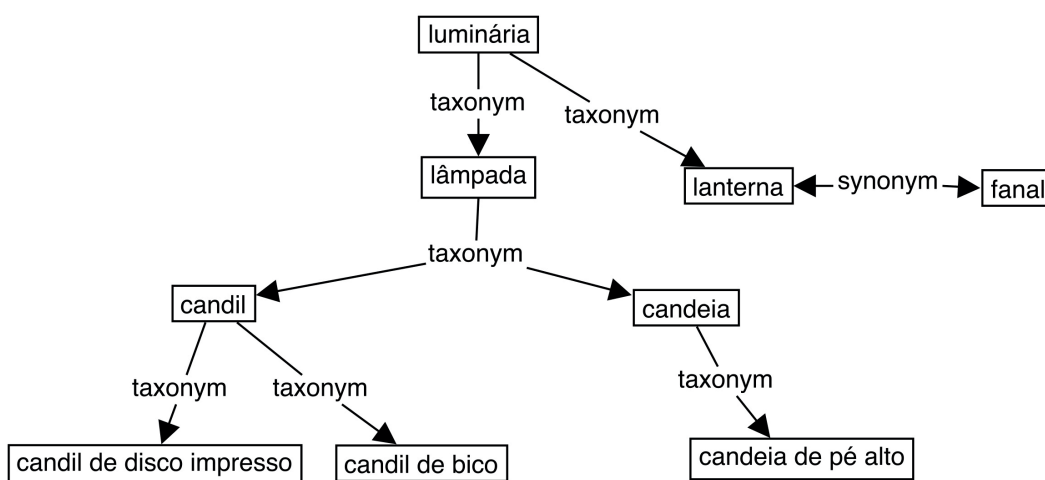


Fig. 3 – Termos para artefactos de iluminação em português.

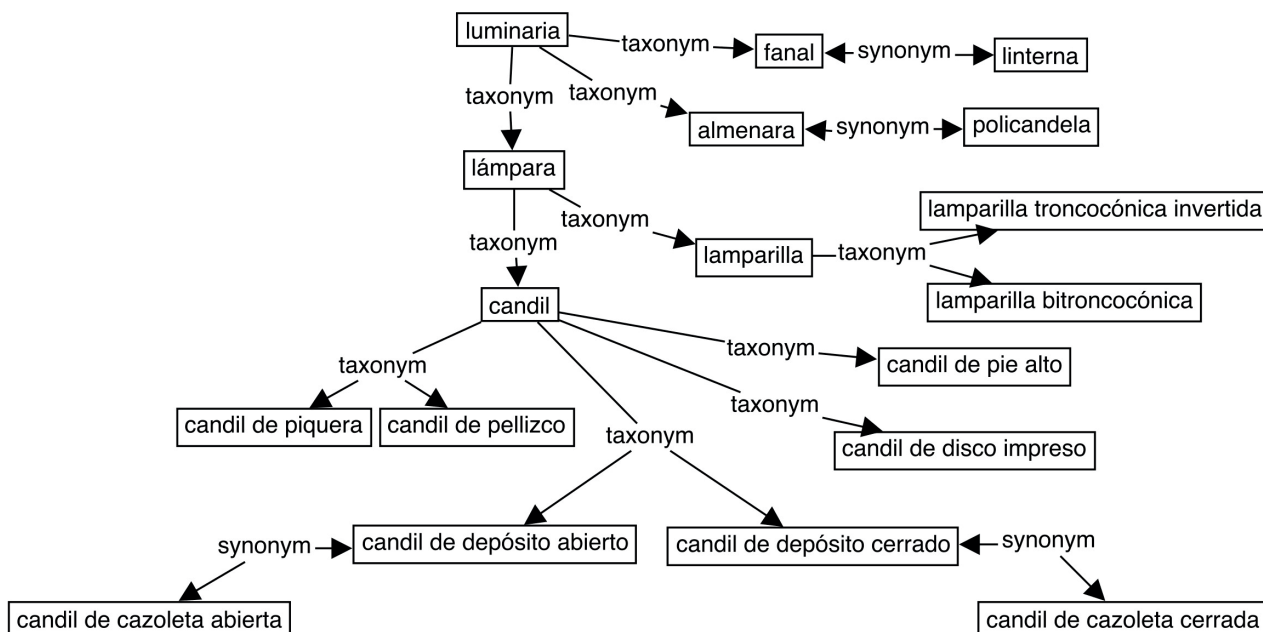


Fig. 4 – Termos para artefactos de iluminação em castelhano.

4. Integração de informação linguística e conceptual

A possibilidade de incluir predicados do Lemon diretamente na OntoAndalus permite uma representação detalhada da informação linguística, nomeadamente no que se refere a termos em diferentes línguas e nas relações léxico-semânticas entre termos. Isto envolve a utilização de elementos dos módulos *ontolex*, *lime* e *vartrans*, mas também da ontologia LexInfo, a qual fornece várias categorias para o Lemon (p. ex., a relação de hiponímia entre os sentidos das entradas lexicais).

A Fig. 5 mostra uma possível entrada para o termo português “candil” utilizando os elementos acima mencionados. Neste exemplo, a terminologia dos artefactos em português é declarada como uma subclasse de lime:Lexicon. As informações gramaticais sobre a entrada incluem a classe de palavras, género, forma canónica e forma plural. As informações semânticas incluem referência à classe ClosedLamp na OntoAndalus.

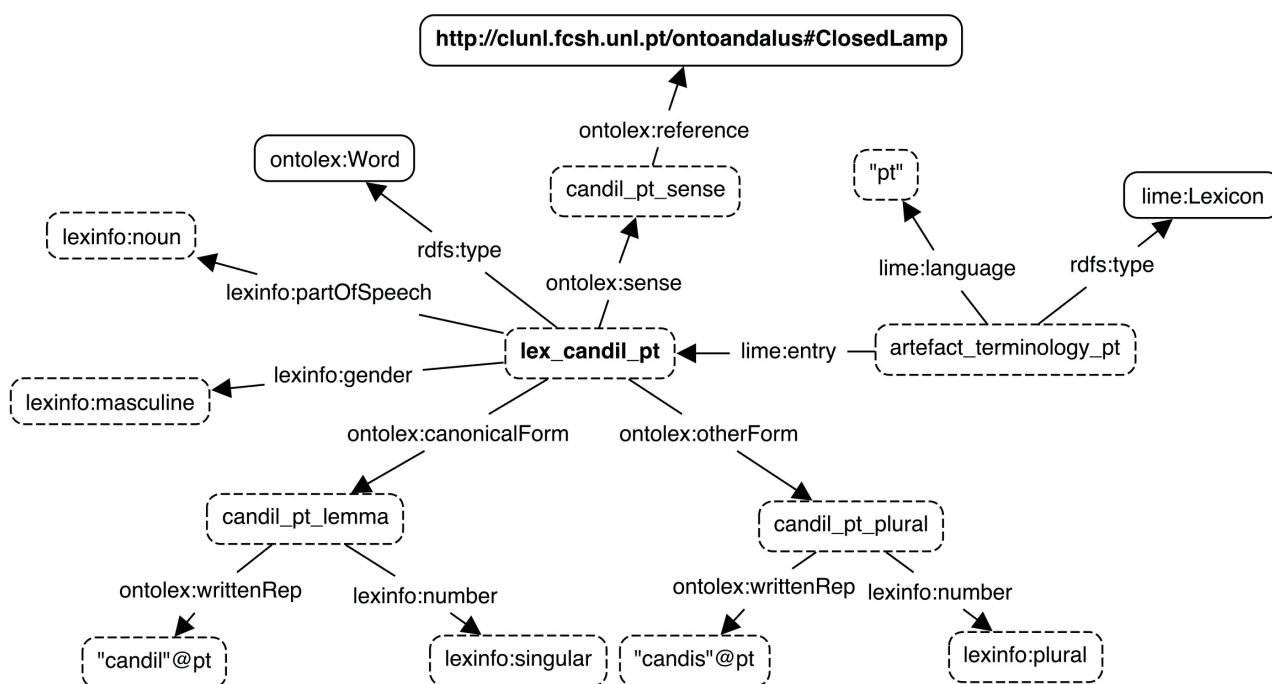


Fig. 5 – Representação de informação gramática e semântica relativa ao termo português “candil”.

No que diz respeito às relações léxico-semânticas, a Fig. 6 mostra um exemplo da relação de equivalência entre termos, assim como um exemplo da relação de hiponímia. A relação de equivalência é definida entre o termo português “candil” e o termo castelhano “candil de depósito cerrado”. Tal é feito apontando a referência dos sentidos de ambos os termos para o mesmo elemento ontológico, neste caso a classe ClosedLamp na OntoAndalus. Utilizando as propriedades do objeto relevantes no módulo *vartrans*, a relação de hiponímia é estabelecida entre os termos portugueses “candil” e “candil de disco impresso”. Neste exemplo, a relação de sentido é reificada, ou seja, representada como um indivíduo, o que permite identificar a sua fonte (o sentido de “candil”), alvo (o sentido de “candil de disco impresso”) e categoria (hiponímia).

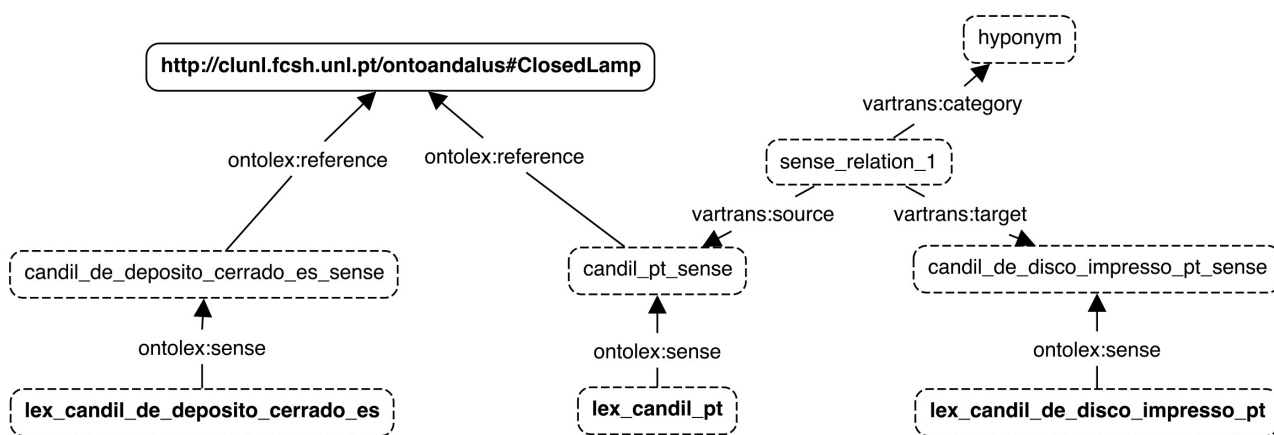


Fig. 6 - Relações léxico-semânticas entre termos.

O Lemon permite, desta forma, representar diversas informações sobre os termos, incluindo informações gramaticais (p. ex., classe de palavras, género gramatical) e informações semânticas (p. ex., relações léxico-semânticas, referência a elementos ontológicos).

5. Conclusões

Com este artigo, pretendemos dar uma visão geral do trabalho realizado para elaborar um recurso terminológico em arqueologia islâmica. O conhecimento do domínio foi representado através da OntoAndalus, uma ontologia centrada nos tipos de artefactos relevantes para os estudos de cerâmica do al-Andalus. A OntoAndalus reutiliza a DOLCE Ultralite, o que permite uma representação detalhada do conhecimento do domínio, como foi demonstrado pelo caso dos artefactos de iluminação. Os termos para artefactos em português e castelhano foram extraídos de um corpus especializado e organizados em redes lexicais. Subsequentemente, o Lemon foi utilizado para representar informações sobre os termos e sua relação com os conceitos relevantes da ontologia de domínio. O Lemon demonstrou ser capaz de representar informações gramaticais e semânticas sobre os termos, sendo um modelo promissor para aplicações no trabalho terminológico.

Bibliografia citada

ALMEIDA, Bruno; COSTA, Rute (2020) – OntoAndalus: an ontology of Islamic artefacts for terminological purposes. *Semantic Web Journal*. Pre-press, 1–17. doi: 10.3233/SW-200387.

ALMEIDA, Bruno; COSTA, Rute; ROCHE, Christophe (2019) – The names of lighting artefacts: extraction and representation of Portuguese and Spanish terms in the archaeology of al-Andalus. *Revue TAL*. 60:3, pp. 113–137.

BOWKER, Lynne; PEARSON, Jennifer (2002) – *Working with specialized language : a practical guide to using corpora*. London: Routledge.

BUGALHÃO, Jacinta; CATARINO, Helena; CAVACO, Sandra; COVANEIRO, Jaquelina; FERNANDES, Isabel Cristina; GOMES, Ana; GÓMEZ-MARTÍNEZ, Susana; GONÇALVES, Maria José; GRANGÉ, Mathieu; INÁCIO, Isabel; LOPES, Gonçalo; SANTOS, Constança dos (2010) – CIGA: Projecto de sistematização para a cerâmica islâmica do Gharb al-Ándalus. In *Actas do 7º Encontro de Arqueologia do Algarve* (Xelb, 10). Silves: Câmara Municipal de Silves, pp. 455-476.

COLL CONESA, Jaime; MARTÍ OLTRA, Javier; PASCUAL PACHECO, Josefa (1988) – *Cerámica y cambio cultural: el tránsito de la Valencia islámica a la cristiana*. Madrid: Dirección General de Bellas Artes y Archivos.

COSTA, Rute (2001) – *Pressupostos teóricos e metodológicos para a extracção automática de unidades terminológicas multiléxicas*. [S.l.]: Universidade Nova de Lisboa, Tese de doutoramento. Disponível em: <http://hdl.handle.net/10362/12067>

COSTA, Rute (2013) – Terminology and specialised lexicography: two complementary domains. *Lexicographica*. 29:1, pp. 29-42.

COVANEIRO, Jaquelina; FERNANDES, Isabel Cristina; GÓMEZ-MARTÍNEZ, Susana; GONÇALVES, Maria José; INÁCIO, Isabel; SANTOS, Constança dos; COELHO, Catarina; LIBERATO, Marco; BUGALHÃO, Jacinta; CATARINO, Helena; CAVACO, Sandra (2013) – Cerâmica islâmica em Portugal: 150 anos de investigação. In ARNAUD, José Morais; MARTINS, Andrea; NEVES, César, eds. – *Arqueologia em Portugal – 150 Anos*. Lisboa: Associação dos Arqueólogos Portugueses, pp. 73-80.

CRUSE, David Alan (1986) – *Lexical semantics*. Cambridge: Cambridge University Press.

GANGEMI, Aldo (2016) – Dolce+D&S Ultralite and its main ontology design patterns. HITZLER, Pascal; GANGEMI, Aldo; JANOWICZ, Krzysztof; KRIZNADHI, Adila; PRESUTTI, Valentina, eds. – *Ontology engineering with ontology design patterns: foundations and applications* [Em linha]. Amsterdam: IOS Press, pp. 81-103.

GOMES, Rosa Varela (2002) – *Silves (Xelb), uma cidade do Gharb al-Andalus: território e cultura*. Lisboa: Instituto Português de Arqueologia.

GOMES, Rosa Varela (2003) – *Silves (Xelb), uma cidade do Gharb al-Andalus: a alcáçova*. Lisboa: Instituto Português de Arqueologia.

GOMES, Rosa Varela (2006) – *Silves (Xelb), uma cidade do Gharb al-Andalus: o núcleo urbano*. Lisboa: Instituto Português de Arqueologia.

GOMES, Rosa Varela (2011) – *Silves (Xelb), uma cidade do Gharb al-Andalus: a zona da Arrochela, espaços e quotidianos*. Lisboa: Instituto Português de Arqueologia.

GÓMEZ-MARTÍNEZ, Susana (2004) – *La cerámica islámica de Mértola: producción y comercio*. [S.l.]: Universidad Complutense de Madrid. Tese de doutoramento.

ISO 704 – *Terminology work – Principles and methods*. Geneva: ISO, 2009.

ISO 1087 – *Terminology work and terminology science – Vocabulary*. Geneva: ISO, 2019.

RICE, Prudence (1987) – *Pottery analysis: a sourcebook*. Chicago, IL: The University of Chicago Press.

ROSSELLÓ-BORDOY, Guillermo (1978) – *Ensayo de sistematización de la cerámica árabe en Mallorca*. Palma de Mallorca: Institut d'Estudis Balearics.

ROSSELLÓ-BORDOY, Guillermo (1991) – *El nombre de las cosas en al-Andalus: una propuesta de terminología cerámica*. Palma de Mallorca: Museo de Mallorca.

SANTOS, Cláudia; COSTA, Rute (2015) – Domain specificity: semasiological and onomasiological knowledge representation. In KOCKAERT, Hendrik J.; STEURS, Frieda, eds. – *Handbook of terminology* [Em linha]. Amsterdam: John Benjamins, v. 1. pp. 153-179.

VALLEJO TRIANO, Antonio; ESCUDERO ARANDA, José (1999) – Aportaciones para una tipología de la cerámica común califal de Madinat al-Zahra. *Arqueología y Territorio Medieval*. 6, pp. 133-176. doi: 10.17561/aytm.v6i0.1530.