Modeling Semantic Anisomorphism in Multilingual Terminological Resources Using Knowledge Graphs

Giorgia Lodi[1,†], Eleonora Marzi[2,*,†] and Valentina Presutti[3,†]

[1]Institute of Cognitive Sciences and Technologies of CNR, Consiglio Nazionale delle Ricerche, Italy
[2]University of Chieti-Pescara "G. D'Annunzio", Italy
[3]Alma Mater Studiorum University of Bologna, Italy

Abstract

This paper addresses the problem of semantic anisomorphism in multilingual terminological resources through an RDF-based modeling approach. Drawing on a case study of a multilingual controlled vocabulary in the domain of architecture, it proposes a semantic framework designed to make explicit the asymmetries of conceptual relations across languages, while avoiding the erasure of culturally embedded distinctions in multilingual conceptual modeling. Rather than treating translations as simple formal equivalences, the proposed approach models them as qualified semantic relations that capture the conceptual stratification and domain-specific complexity of specialized knowledge.

Keywords

Semantic Anisomorphism, Knowledge Graphs, Multilingual Controlled Vocabulary, Architectural Terminology

1. Introduction and State of the Art

Technologies for the management of linguistic data enable large-scale data extraction, making it possible to generate extensive multilingual datasets with unprecedented speed and scope [1]. This expansion, however, is not merely quantitative: it reshapes the way linguistic and conceptual information is collected, structured, and reused across domains. In parallel with this trend, increasing attention is being paid to data quality, which has become a central concern in both research and applied contexts. In particular, ensuring complete and inclusive data representation [2], as well as an appropriate level of granularity in particular lexical or semantic gaps [3], is now essential for capturing linguistic variation, supporting interoperability, and enabling reliable downstream analyses [4]. Among the most relevant yet still underexplored aspects is the phenomenon of *semantic anisomorphism* [5]. Languages do not merely encode distinct linguistic systems; they also reflect different ways of structuring and interpreting reality. As a result, terms that appear to be straightforward equivalents in translation often cover partially overlapping semantic fields or convey language-specific cultural nuances. This phenomenon becomes especially evident in terminology, where the same domain of knowledge can be structured with varying degrees of granularity or according to different typological, functional, or historical parameters, depending on the disciplinary tradition involved. In such cases, translation does not necessarily imply full semantic equivalence; rather, it may correspond to relations of partial equivalence, conceptual overlap, or mere contextual translatability. Within this framework, it is crucial that the construction of multilingual terminological resources makes explicit the underlying semantic relations between the concepts conveyed by different languages.

Semantic anisomorphism in multilingual datasets thus refers to the lack of a strict one-to-one correspondence between meanings across languages. This phenomenon poses significant challenges for translation, data alignment, and the development of reliable computational models, particularly

5th International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2026, June 25-26, 2026, Zadar, Croatia.

*Corresponding author.
†These authors contributed equally.
giorgia.lodi@cnr.it (G. Lodi); eleonora.marzi@unich.it (E. Marzi); valentina.presutti@unibo.it (V. Presutti)
0009-0000-9032-4831 (G. Lodi); 0000-0002-4558-9483 (E. Marzi); 0000-0002-9380-5160 (V. Presutti)

in the context of artificial intelligence. Addressing such mismatches requires approaches capable of capturing fine-grained semantic distinctions, as well as modeling the cultural and conceptual dimensions embedded in linguistic expressions. This can be addressed with a multidimensional modeling approach, and knowledge graph-based data representation has proven to be particularly well suited for this purpose. It also enables interoperability while supporting linguistic and semantic inference. Crucially, such a modeling framework also provides the formal means to account for semantic anisomorphism, by allowing conceptual relations across languages to be represented in a non-reductive and non-isomorphic manner. The effectiveness of knowledge graph-based approaches in this scenario is widely demonstrated by the use of so-called controlled vocabularies, which are important for ensuring semantic consistency in terminology and knowledge representation and are typically defined using Semantic Web standards, such as RDF (Resource Description Framework)[1] and usually formalized through ontological models like SKOS (Simple Knowledge Organization System)[2]. Prominent examples include the *Art and Architecture Thesaurus* (AAT)[3], a controlled vocabulary dedicated to art, architecture, and material culture, which serves as a reference standard for semantic descriptions and cataloging in museums, archives, and libraries. Another notable case is *EuroVoc*[4], the multilingual thesaurus of the European Union, widely used in institutional and research contexts for indexing, metadata creation, and semantic interoperability. Within this framework, the discipline of terminology, to which controlled vocabularies belong, has increasingly embraced the paradigm of *digital terminology* [6], which positions technology not merely as a technical support, but as an integral component of the epistemological process. This perspective has engendered a growing body of research that explicitly addresses the interaction between terminological modeling, knowledge representation, and semantic variation between languages [7] [8]. Considering this background, the present contribution proposes an RDF-based modeling approach that explicitly represents conceptual non-symmetry, by integrating, into terminological resources, information concerning both the degree and the type of equivalence between concepts across different languages. Through such explicit formalization, the aim is to ensure the development of interoperable and theoretically grounded resources while avoiding reductive translation practices that tend to homogenize or oversimplify the conceptual complexity of specialized domains. The paper is structured as follows: Section 2 details our methodology, Section 3 illustrates a running example in the field of architecture, and Section 4 concludes the paper with some future work.

2. Methodology

To illustrate the proposed modeling approach, we draw on a use case of semantic anisomorphism within a controlled vocabulary for the architecture domain, developed as part of an ongoing collaboration funded by the Italian Ministry of Culture on the Italian ontology of constructions. More specifically, we model the relationship between the Italian term “pianta a croce latina immissa” and the French term “plan en croix latine”, highlighting the non-symmetric conceptual correspondence between the two linguistic designations. We collect terminological entries provided by domain experts that, though may take various forms, must include information as concept lexicalization and concept definition. It becomes possible to identify any preferred or alternative lexicalizations and the equivalence in languages through consultation of the domain-specific corpus that allows for the enrichment of the controlled vocabulary with real usage data, potentially revealing preferred or commonly used terms that are specific to a given period from a diachronic perspective. This is made possible by the conceptual framework of textual terminology [9], which, by adopting a view of the term as a general-lexicon word used in specialized contexts, provides access to specialized terminology through the consultation and analysis of corpora. For modeling, we used the OntoLex-Lemon ontology[5] [10], which is designed to enrich a given ontology by providing a linguistic anchoring, through the addition of lexical, morphological,

[1] https://www.w3.org/TR/rdf11-concepts/
[2] https://www.w3.org/2004/02/skos/
[3] https://www.getty.edu/research/tools/vocabularies/aat/index.html
[4] https://www.vocabularyserver.com/eurovoc/it/
[5] https://www.w3.org/2016/05/ontolex/

terminological and semantic information. The model follows the "semantics by reference" [11], allowing the introduction of lexicalized concepts that are not explicitly represented in the ontology. While the case study presented focuses on a specific instance of semantic anisomorphism, namely interlingual hypernymy, the modeling approach can be considered a generalizable framework, applicable to other cases requiring the explicit representation of similar semantic relationships.

Beyond ensuring a clear separation of levels, this structure enables a fine-grained alignment between linguistic data and domain knowledge. It becomes possible, for instance, to map multiple expressions onto the same concept while preserving their distinct semantic profiles, or to distinguish closely related concepts that may be lexicalized differently across languages. Such flexibility is particularly relevant in specialized domains like architecture, where terminological precision coexists with variation shaped by historical, cultural, and disciplinary factors. The articulation across three coordinated planes—formal, semantic, and conceptual—thus proves especially effective for addressing the complexity of the task. By organizing information in this layered way, the model supports interoperability between linguistic resources and ontological structures, facilitates the integration of new terms without disrupting existing relations, and provides a transparent framework for encoding both equivalence and divergence. This makes it not only technically robust, but also methodologically aligned with the needs of multilingual terminological modeling.

3. Modeling Semantic Anisomorphism: a case study

3.1. Linguistic overview of terms

Within the domain of architectural terminology, the Italian term "pianta a croce latina" denotes a plan of a religious building characterized by the intersection of the nave with the transept, in particular with a longitudinal axis that is more extended than the transverse one. In Italian, there are two stabilized hypernyms at the terminological level: "pianta a croce latina commissa" and "pianta a croce latina immissa", which differ in the position and projection of the transept with respect to the main body of the building, thus encoding a distinction that is both geometrical and typological. In the *immissa* type, the transept intersects the nave within its longitudinal body, whereas in the *commissa* type it is attached at the end of the nave, producing a more T-shaped configuration. The French term "plan en croix latine" - which is the equivalent of "pianta a croce latina" is terminological stable and conceptually robust, and covers the entire typological domain without requiring further specification in most contexts. However, there are no stable, autonomous lexicalized equivalents in French comparable to the Italian terms "pianta a croce latina immissa" and "pianta a croce latina commissa", as this distinction is not systematically terminologized and tends instead to be expressed through descriptive or context-dependent formulations. In specialized corpora, variant forms corresponding to "immissa" can nonetheless be observed, such as: "plan en croix latine à transept saillant" and "croix latine à transept débordant", while equivalents for "commissa" include: "plan en croix latine sans transept saillant", "croix latine en forme de T", and "croix latine à transept non débordant". These expressions, however, remain less fixed and display a lower degree of lexical stabilization. The semantic-lexical modeling approach we propose makes explicit the conceptual and terminological relations between Italian and French, by separating these two distinct levels while introducing a third one that mediates between them. The conceptual layer captures domain knowledge and category structures, while the linguistic layer encodes the language-specific expressions that realize these concepts. The meta-level makes it possible to explicitly qualify the nature of the correspondence between forms and concepts across languages, accounting for full matches as well as partial or asymmetric alignments.

3.2. Modeling of conceptual relations and lexical forms

The example of "pianta a croce latina" illustrates how a conceptual unit within the architectural domain is modeled as an *ontolex:LexicalConcept*, defined independently of language and serving as a shared semantic anchor. Within this framework, a lexical concept corresponds to a stable unit of

meaning that abstracts away from any particular linguistic realization, thereby functioning as a pivot for cross-linguistic alignment. The linguistic expressions in Italian and French are represented as *ontolex:LexicalEntry* instances, which group together a set of grammatically related forms and encode their lexical identity. Each lexical entry is associated with a specific language and is characterized by one or more *ontolex:Form* elements, which capture its surface realizations, including orthographic and morphological variants. In this way, the model distinguishes between the abstract lexical unit and its possible manifestations in actual usage. The link between lexical form and concept is mediated by *ontolex:LexicalSense*, which represents the specific meaning that a lexical entry assumes in a given context of use. A lexical sense thus functions as the interpretative bridge between linguistic expression and conceptual content, and, via *ontolex:reference*, it connects the lexical system to the corresponding conceptual node in the ontology. This three-level articulation allows for the formal representation of both convergences and asymmetries across languages, enabling a more precise treatment of equivalence, partial overlap, and structural divergence. By clearly separating conceptual structure, lexical realization, and sense-level mediation, the model provides a robust and flexible foundation for subsequent semantic and terminological modeling steps.

We report below the instantiation in RDF/Turtle, also graphically represented in Graffoo in Figure 1 (for the sake of readablity, in the Graffoo we do not report the second block of RDF triples).

```
PREFIX arco-catalogue: <https://w3id.org/arco/ontology/catalogue/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
PREFIX owl: <http://www.w3.org/2002/07/owl#>.
PREFIX skos: <http://www.w3.org/2004/02/skos#>.
PREFIX dct: <http://purl.org/dc/terms/>.
PREFIX ontolex: <http://www.w3.org/ns/lemon/ontolex#> .
PREFIX vartrans: <http://www.w3.org/ns/lemon/vartrans#> .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/1> a skos:Concept, ol:LexicalConcept ;
  ontolex:isEvokedBy <https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-entry/ita-1> ;
  ontolex:isEvokeBy <https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-entry/fra-1> ;
 skos:definition "pianta in cui navata e transetto hanno lunghezza differente e si intersecano ad angolo retto"@it .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-entry/ita-1> a ontolex:MultiwordExpression ;
  ontolex:sense <https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-sense/ita-1> ;
  ontolex:canonicalForm <https://w3id.org/arco/controlled-vocabulary/plain-layout/form/ita-1> ;
  dct:language <http://publications.europa.eu/resource/authority/language/ITA> .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-entry/fra-1> a ontolex:MultiwordExpression ;
  ontolex:sense <https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-sense/fra-1> ;
  ontolex:canonicalForm <https://w3id.org/arco/controlled-vocabulary/plain-layout/form/fra-1> ;
  dct:language <http://publications.europa.eu/resource/authority/language/FRA> .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-sense/ita-1> a ontolex:LexicalSense ;
  ontolex:reference <https://w3id.org/arco/controlled-vocabulary/plain-layout/1> ;
  ontolex:usage <https://w3id.org/arco/controlled-vocabulary/plan-layout/usage/architettura> .

<https://w3id.org/arco/controlled-vocabulary/plan-layout/usage/architettura> a rdfs:Resource ;
  rdfs:label "contesto architettura"@it .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/form/ita-1> a ontolex:Form ;
  ontolex:writtenRep "pianta a croce latina"@it .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/form/fra-1> a ontolex:Form ;
  ontolex:writtenRep "plan en croix latine"@fr .
```

The second-level concept related to “pianta a croce latina immissa” follows the same modeling approach as for the level-1 concept “pianta a croce latina”. We don’t repeat the prefix as the same of previous modelization.

```
<https://w3id.org/arco/controlled-vocabulary/plain-layout/11> a skos:Concept, ol:LexicalConcept ;
  ontolex:isEvokedBy <https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-entry/ita-11> ;
  skos:definition "tipo di pianta a croce latina in cui il transetto è posto all’estremità terminale
  della navata longitudinale"@it .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-entry/ita-11> a ontolex:MultiwordExpression ;
```

```
  ontolex:sense <https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-sense/ita-11> ;
  ontolex:canonicalForm <https://w3id.org/arco/controlled-vocabulary/plain-layout/form/ita-11> ;
  dct:language <http://publications.europa.eu/resource/authority/language/ITA> .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-sense/ita-11> a ontolex:LexicalSense ;
  ontolex:reference <https://w3id.org/arco/controlled-vocabulary/plain-layout/11> ;
  ontolex:usage <https://w3id.org/arco/controlled-vocabulary/plan-layout/usage/architettura> .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/form/ita-1> a ontolex:Form ;
  ontolex:writtenRep "pianta a croce latina commissa"@it .
```

For the purposes of modeling, translation is treated as a lexical-semantic relation (Figure 1), allowing us not only to distinguish between lexical and semantic aspects, which are not necessarily equivalent in translation, but also to provide annotations regarding the type of equivalence.

```
<https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-sense/ita-11>
  a ontolex:LexicalSense ;
  ontolex:reference <https://w3id.org/arco/controlled-vocabulary/plain-layout/11> ;
  ontolex:usage <https://w3id.org/arco/controlled-vocabulary/plan-layout/usage/architettura> .


<https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-sense/fra-1>
  a ontolex:LexicalSense ;
  <https://w3id.org/arco/controlled-vocabulary/plain-layout/1> ;
  ontolex:usage <https://w3id.org/arco/controlled-vocabulary/plan-layout/usage/architecture> .

<https://w3id.org/arco/controlled-vocabulary/plan-layout/usage/architettura> a rdfs:Resource ;
    rdfs:label "architecture"@fr .

<https://w3id.org/arco/controlled-vocabulary/plain-layout/translation/ita-11-fra-1> a vartrans:Translation ;
   vartrans:source <https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-sense/ita-11> ;
   #pianta a croce latina immissa
   vartrans:target <https://w3id.org/arco/controlled-vocabulary/plain-layout/lexical-sense/fra-1> ;
   vartrans:category lexinfo:hypernym ;
  rdfs:commnet "traduzione parziale: il termine francese è l'iperonimo del termine italiano 'immissa'"@it .
```

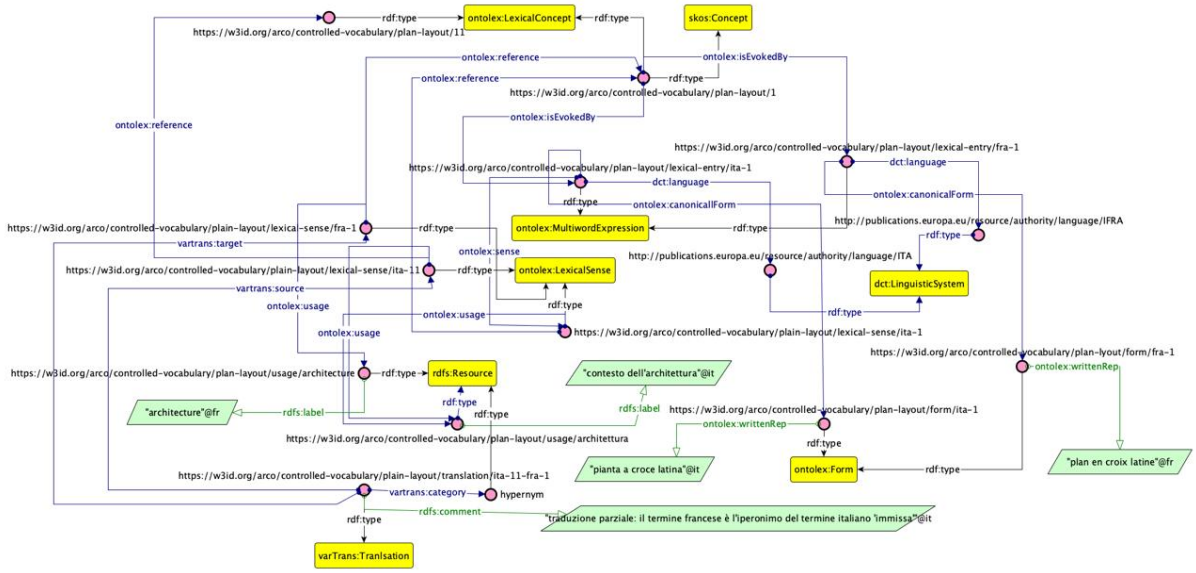

Figure 1: Modeling semantic anisomorphism of "pianta a croce immissa" and "plan en croix latine"

As previously noted, the Italian term "pianta a croce latina immissa" represents a hyponym within the conceptual hierarchy of "pianta a croce Latina". In French, although the equivalent hypernym concept ("plan en croix latine") exists, there is no stable lexicalized term corresponding to the Italian hyponym. To represent this specificity, the VarTrans module of OntoLex-Lemon was used. Specifically, the Italian

lexical sense associated with the hyponym is linked to its reference concept via *ontolex:reference*, while the corresponding French term is modeled as the sense of the hypernym. The translation relation between the two senses is thus formalized through an instance of vartrans:Translation, assigned with the category *lexinfo:hypernym*, which allows the degree of correspondence in translation to be made explicit: the French term corresponds to a more general concept than the Italian term, highlighting a loss of granularity in the interlingual mapping.

This modeling approach enables the formalization of several key aspects of multilingual terminology. First, it distinguishes concepts and their internal hierarchical relations within each language; second, it represents the anisomorphism nature of translations across languages, where the absence of a stabilized term in one language can be expressed as a hypernymic translation; and finally, it allows for the association of usage and contextual information, providing semantic richness useful for both specialist consultation and automatic integration into knowledge graphs.

4. Conclusion and future works

We have presented a case of data modeling for a multilingual terminological resource, focusing on the explicit representation of conceptual, lexical, and translational information, demonstrating its usefulness for modeling complex situations from a semantic perspective, such as semantic anisomorphism. The explicit articulation of these knowledge levels represents a significant contribution to the management of multilingual terminological resources, as it allows for: (i) preserving the original conceptual granularity despite linguistic asymmetries; (ii) supporting machine translation and terminological alignment systems that are more context-aware and sensitive to hierarchical relations; and (iii) providing a transparent representation of terminological choices, thereby promoting interoperability and reusability in semantic datasets. In summary, the proposed modeling approach highlights the possibility of treating translations not merely as formal equivalences, but as qualified semantic relations that reflect the complexity and conceptual stratification of specialized domains. We plan to use the result of this work as a possible knowledge base (RAG technique) for generative AI solutions so that to allow them to better understand the context of the specific language terms.

Acknowledgments

This research is partly supported by INFINITY: a EU Horizon Europe project under Grant Agreement No 101233051 and by the Central Institute of Cataloguing and Documentation of the Italian Ministry of Culture.

Declaration on Generative AI

During the preparation of this work, the author(s) used Mistral in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] S. Sun, K. Duh, CLIRMatrix: A massively large collection of bilingual and multilingual datasets for cross-lingual information retrieval, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 4160–4170. URL: https://aclanthology.org/2020.emnlp-main.340/. doi:10.18653/v1/2020.emnlp-main.340.

[2] F. Giunchiglia, G. Bella, N. C. Nair, Y. Chi, H. Xu, Representing interlingual meaning in lexical databases, 2023. URL: https://arxiv.org/abs/2301.09169. arXiv:2301.09169.

[3] M. Janssen, Multilingual lexical databases, lexical gaps, and simullda, International Journal of Lexicography 17 (2004) 137–154. URL: https://doi.org/10.1093/ijl/17.2.137. doi:10.1093/ijl/17.2.137. arXiv:https://academic.oup.com/ijl/article-pdf/17/2/137/2912751/ech045.pdf.

[4] J. Liu, I. Ouzzani, W. Li, L. Zhang, T. Ou, H. Bouamor, Z. Jin, M. Diab, Towards global ai inclusivity: A large-scale multilingual terminology dataset (gist), 2025. URL: https://arxiv.org/abs/2412.18367. arXiv:2412.18367.

[5] A. G. González-Jover, Meaning and anisomorphism in modern lexicography, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 12 (2006) 215–234. URL: https://www.jbe-platform.com/content/journals/10.1075/term.12.2.04gon. doi:https://doi.org/10.1075/term.12.2.04gon.

[6] F. Vezzani, Terminologie numérique : conception, représentation et gestion, Peter Lang, 2022. doi:10.3726/b19407.

[7] S. Piccini, M. Abrate, A. Bellandi, E. Giovannetti, Rappresentazione, costruzione e visualizzazione di risorse terminologiche diacroniche nell'era del web semantico, Firenze University Press, 2021, p. 125–139. URL: http://dx.doi.org/10.36253/978-88-5518-364-2.12. doi:10.36253/978-88-5518-364-2.12.

[8] S. Piccini, F. Vezzani, A. Bellandi, Tbx and 'lemon': What perspectives in terminology?, Digit. Scholarsh. Humanit. 38 (2023) 61–72. URL: https://api.semanticscholar.org/CorpusID:259809489.

[9] A. Condamines, A. Picton, Theoretical Approaches to Terminology, John Benjamins, 2021.

[10] J. P. McCrae, J. Bosque-Gil, J. Gracia, P. Buitelaar, P. Cimiano, The ontolex-lemon model: development and applications, in: Proceedings of the Electronic Lexicography of the 21st Century, Leiden, the Netherlands, 2017, pp. 19–21.

[11] J. Mccrae, G. A. de Cea, P. Buitelaar, P. Cimiano, T. Declerck, A. Gomez-Perez, J. Gracia, L. Hollink, E. Montiel-Ponsoda, D. Spohr, Interchanging lexical resources on the semantic web, Language Resources and Evaluation 46 (2012) 701–719. doi:10.1007/s10579-012-9182-3.