DEC Facile: An Intuitive Tool for Creating Combinatory Explanatory Dictionaries in the Semantic Web*

Abdelouahab Masbah[1], Andrea Bellandi[1] and Silvia Piccini[1,*]

[1] *Istituto di Linguistica Computazionale "A. Zampolli", via G. Moruzzi 1, 56122 Pisa, Italy*

Abstract

This paper presents *DEC Facile*, a tool specifically designed for the creation of computational Explanatory Combinatorial Dictionaries (ECDs). Powered by LexO-server and Lemon modules, *DEC Facile* integrates Linked Data principles for efficient lexical resource management. The tool's user-friendly interface supports tasks like dictionary search, sense editing, and lexical function linking, making it a valuable resource for lexicography, terminography, and computational linguistics.

Keywords

Explanatory Combinatorial Dictionary, Ontolex-Lemon, Semantic Web, Linked Data

1. Introduction

The aim of this paper is to introduce the first prototype of *DEC Facile*, a tool designed for the creation of computational Explanatory Combinatorial Dictionaries (ECDs). *DEC Facile* is specifically intended for users who already possess a computational resource structured according to the OntoLex-Lemon model [1] and wish to convert it into an ECD, adhering to the principles of the Explanatory Combinatorial Lexicology (ECL) [2].

ECL, developed by Mel'čuk and his collaborators within the framework of Meaning-Text Theory [3, 4, 5], has gained increasing recognition in general and computational linguistics due to its structured approach to describing the semantics and combinatorial properties of lexical units in a comprehensive and rigorous manner. In particular, the formalism of lexical functions—the core element of the theory—has attracted significant interest from specialists in computational linguistics, due to their effectiveness in supporting advanced Natural Language Processing tasks. Lexical functions help model systematic lexical relationships and play a crucial role in resolving syntactic and lexical ambiguities, identifying idiomatic equivalents in machine translation, and generating paraphrases, among other applications [e.g., 6, 7].

Given the strong theoretical foundations and practical advantages of the ECL model, several projects have implemented its principles to build large-scale digital resources. Notable examples include the *Diccionario de Colocaciones del Español* (DiCE), a web-based Spanish collocations dictionary developed by the Universidad de La Coruña [8], and the *Papillon Lexical Database Project* [9], which aims to create a multilingual French-English-Japanese lexical database using interlingual links, from which digital bilingual French-Japanese and Japanese-French dictionaries can be derived.

Beyond general computational lexicography, the ECL model has also been widely adopted in specialized lexicography and terminology (10, 11, 12, 13). At the *Observatoire de linguistique*

4th International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2025, June 19-20, 2025, Thessaloniki, Greece.

* Corresponding author.

mailto:abdelouahab.masbah@ilc.cnr.it (A. Masbah); andrea.bellandi@ilc.cnr.it (A. Bellandi); silvia.piccini@ilc.cnr.it (S. Piccini)

0009-0009-9692-0533 (A. Masbah); 0000-0002-1900-5616 (A. Bellandi); 0000-0002-2584-0191 (S. Piccini)

Sens-Texte (OLST) at the University of Montréal, several ECDs have been developed, including *DiCoEnviro* (environmental terminology), *DiCoInfo* (computer science terminology), and *DiCoLexitrans* (legal terminology).

Creating ECDs is a complex and time-consuming process. Previous attempts to develop editors for ECDs have been made [e. g., 14, 15], with the primary goal of assisting users in this challenging task. *DEC Facile* continues this effort while introducing a significant innovation. Specifically, it leverages semantic web technologies [16] and the Linked Open Data (LOD) paradigm [17] to facilitate the creation of ECDs that are natively structured for interoperability with other lexical resources and language technologies.

By adhering to the FAIR (Findable, Accessible, Interoperable, and Reusable) principles [18], *DEC Facile* ensures that the lexical data it produces can be integrated into existing linguistic infrastructures. For instance, resources generated with *DEC Facile* can be linked to external lexical databases such as BabelNet[2], WordNet[3], and Wikidata[4], thanks to its compatibility with the OntoLex-Lemon framework. This enables cross-resource querying and enhances the usability of ECDs in multilingual applications. Additionally, as we will see in detail in section 3, *DEC Facile* provides structured exports in RDF and JSON-LD formats, ensuring that its outputs can be used in natural language processing (NLP) pipelines, machine translation engines, and computational terminology management systems.

DEC Facile has been developed as part of a nationally funded research project titled "Old Words for a New World. Translating Christianity to Baltic Pagans", which includes the creation of an ECD focused on the religious lexicon extracted from the earliest Baltic catechisms as one of its primary objectives[5].

In the following sections, we will present the data model underlying *DEC Facile* (§2) and provide a brief overview of the tool's backend and frontend components (§3). Finally, we will draw conclusions (§4).

2. Data model behind the tool

The theoretical framework of the ECL, central to the development of *DEC Facile*, represents a cornerstone of modern lexicography and terminology. ECL's structured approach to modelling lexical relations, combinatorial properties, and semantic nuances offers a powerful tool not only for dictionary creation but also for broader applications such as computational linguistics, terminography, and specialized language analysis.

At its core, an ECD entry is organized into three distinct sections, as illustrated in Figure 1 by the Lithuanian entry *peržegnoti* "to bless; to cross oneself (reflexive form)". The first section (a) presents the definition(s) of the lexeme *L*, formulated using a predefined metalanguage, where the semantic actants [SemAs] introduced by *L* are explicitly identified as variables (X, Y, Z etc.).

---

[2] https://babelnet.org/

[3] https://wordnet.princeton.edu/

[4] https://www.wikidata.org/wiki/Wikidata:Main_Page

[5] The project, funded in 2023 and coordinated by Professor Pietro U. Dini, brings together historical linguistics, comparative religious studies, computational lexicography, and knowledge representation to achieve three interconnected objectives. First, it conducts a systematic contrastive analysis of the translation strategies used for religious vocabulary in the earliest Baltic (Old Lithuanian, Old Latvian and Old Prussian) Catechisms. Second, it develops a diachronic computational lexicon, formally describing key religious terminology according to the principles of Explanatory and Combinatorial Lexicology. Third, it examines the reception of fundamental Christian concepts in the Baltic region, tracing points of convergence and divergence with other European linguistic and cultural traditions, including Celtic and Slavic. These efforts come together in a formal ontology, where Pagan and Christian concepts are systematically represented. The Catechisms will be stored in a digital repository, where each religious term will be linked to its corresponding lexicon entry and ontology concept, ensuring semantic accessibility through advanced query mechanisms. For further details, see: https://www.ilc.cnr.it/en/progetti/ownw/.

The second section (b) describes the morphosyntactic realization of the SemAs, represented as a rectangular matrix, where columns correspond to the SemAs, and rows indicate the possible morphosyntactic realizations. Finally, the third section (c), which is pivotal in Mel'čuk's theory, focuses on lexical functions, which define the syntagmatic and paradigmatic relations that the lexeme *L* has with other dictionary entries.

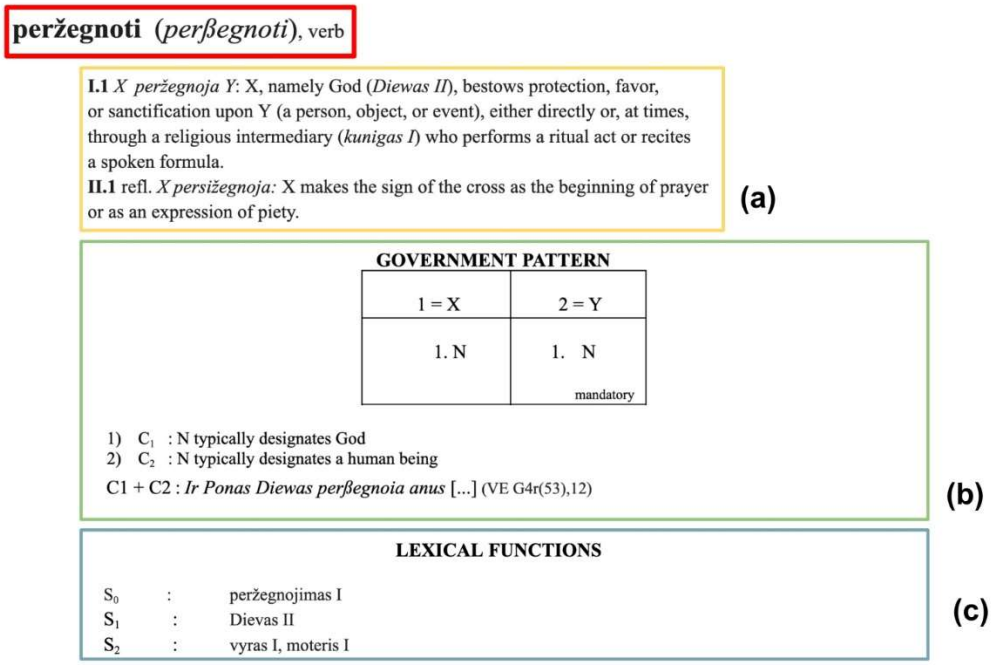

Figure 1: ECD entry for *peržegnoti*. (a) senses definition - (b) government pattern - (c) lexical functions description[6].

This theoretical structure is operationalized and implemented in a computationally actionable format through the OntoLex-Lemon model, developed by the W3C Ontology-Lexicon community group. The model is a suite of RDF vocabularies (referred to as modules) specifically designed to represent and interlink lexicons, terminologies, lexical-semantic resources and, more broadly, language resources in the Linguistic Linked Open Data (LLOD) cloud [19], in alignment with semantic web best practices and LOD principle. Building on prior lexical modeling efforts—including LingInfo [20], LexOnto [21], LIR [22], and LexInfo [23]—the model was first developed within the Monnet project, an initiative aimed at supporting multilingual, ontology-based knowledge representation. Since its initial release, OntoLex-Lemon has undergone continuous refinements and extensions to enhance its expressiveness and interoperability, leading to its widespread adoption in projects focused on multilingual lexicography, terminology management, and computational linguistics (e.g., [24, 25, 26]).

The OntoLex-Lemon modules cover various aspects of linguistic representation, including morphology, syntax-semantics mapping, variation, translation, and linguistic metadata.

In *DEC Facile*, the data model is primarily structured around three key modules: the lexicography module (Lexicog), the lexical functions module (Lexfom), and the syntax-semantics module (SynSem).

Designed to represent existing dictionaries and lexicographic resources as Linked Data, the lexicography module[7] extends the OntoLex core module (often referred to as OntoLex), enhancing its functionality to efficiently manage the structures and annotations commonly found

[6] The Government Pattern in Section (b) and the Lexical Functions in Section (c) refer to sense I.1
[7] https://www.w3.org/2019/09/lexicog/.

in lexicographic practices. Referring to Figure 2(a), a lexical entry (`LexicalEntry`) is defined as an entry of a lexicon (`Lexicon`) characterized by a set of senses (`LexicalSenses`) and a set of written forms (not depicted in Figure 2(a)). A dictionary (`LexicographicResource`), in turn, is composed of a set of dictionary entries (`Entry`), each corresponding to the description of a lexical entry (Figure 2(b)). In *DEC Facile*, the lexicography module is particularly useful for modelling the ordering of senses, as an ECD requires lexical senses to be hierarchically organized, when applicable, according to three levels that reflect the semantic distance between lexemes: large distances, indicated by Roman numerals (I, II, III, etc.); medium distances, indicated by Arabic numerals (1, 2, 3, etc.); and small distances, indicated by lowercase letters (a, b, c). In our example entry *peržegnoti*, as shown in Figure 1, only large distances can be found. To represent this ordering of senses in *Dec facile*, the `LexicographicComponent` class is used, which is a structural element that reflects the arrangement of descriptions provided in a lexicographic resource according to various aspects such as order, hierarchy, grouping, etc.

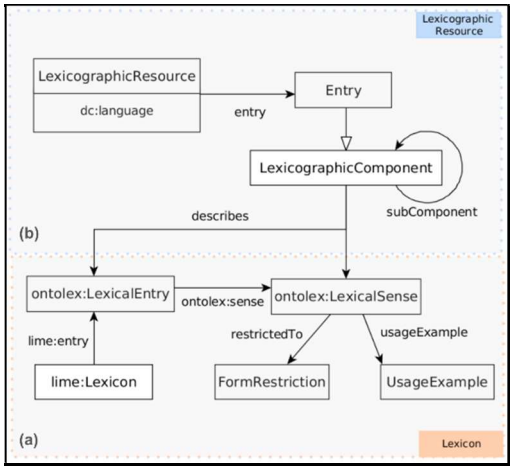

Figure 2: (a) The OntoLex-Lemon core - (b) The lexicography module.

Figure 3 illustrates a fragment of the encoding for the Lithuanian lexical entry *peržegnoti*. In lines 01 - 04, *peržegnoti* is defined as a verb with two lexical senses, whose definitions are provided in lines 20-23. Note that the two references to the senses of other dictionary entries, included in the definition of the first sense of *peržegnoti*, are formally encoded by the `rdfs:seeAlso property` (line 24). Lines 5-6 organize the lexical entry as a lexicographic component of the corresponding dictionary entry. Lines 7-19 define the ordering and the labelling (e.g., "I1", "II1") of the lexical senses. Each sense is linked to a lexicographic component via the `describes` property.

In order to model lexical functions among lexical senses, we used the LexFom [27] module. Based on the OntoLex elements, this module provides the necessary classes to implement lexical functions according to Mel'čuk's theory. These include distinctions such as whether a lexical function is simple or complex, syntagmatic or paradigmatic, as well as identifying the constituents of a lexical function, and other related aspects. Referring to Figure 3, line 25, the lexical function S_0 indicates the structural derivative formed through nominalization: *peržegnojimas* "blessing". Lines 26 and 27 introduce the lexical functions S_1 and S_2, which represent meaningful derivatives reflecting specific syntactic roles. S_1 corresponds to the typical designation of the Deep-Syntactic Actant I (e.g., *Dievas* "God"), broadly corresponding to the agent (or subject, from a syntactic perspective). S_2, on the other hand, corresponds to the typical designation of the Deep-Syntactic Actant II (e.g., *vyras* "man" or *moteris* "woman"), typically corresponding to the patient (the syntactic object) introduced by the verb *peržegnoti*.

```
:peržegnoti_lex_entry a ontolex:LexicalEntry ;
      lexinfo:partOfSpeech lexinfo:verb ;
      ontolex:sense :sense_I1, :sense_II1 ;
      rdfs:label "peržegnoti"@lit .

:peržegnoti_dict_entry a lexicog:Entry ;
      rdf:_1 :peržegnoti_entry_comp.

:peržegnoti_entry_comp a lexicog:LexicographicComponent ;
      lexicog:describes :peržegnoti_lex_entry ;
      rdf:_1 :sense_I_comp ;
      rdf:_2 :sense_II_comp .

:sense_I_comp a lexicog:LexicographicComponent ;
      rdfs:label "I" ;
      rdf:_1 :sense_I1_comp .

:sense_I1_comp a lexicog:LexicographicComponent ;
      rdfs:label "I1" ;
      lexicog:describes :sense_I1 .

:sense_II_comp a lexicog:LexicographicComponent ;
      rdfs:label "II" ;
      rdf:_1 :sense_II1_comp .

:sense_II1_comp a
lexicog:LexicographicComponent ;
      rdfs:label "II1" ;
      lexicog:describes :sense_II1 .

:sense_II1 a ontolex:LexicalSense ;
      skos:definition "refl. X persižegnoja:
      X makes the sign of  the cross as the
      beginn . . . "

:sense_I1 a ontolex:LexicalSense ;
      skos:definition "X  peržegnoja Y: X,
      namely God (Diewas II), bestows
      protection, favor, or . . . " ;
      rdfs:seeAlso :Diewas_sense_II ,
                   :kunigas_sense_I.

:sense_I1 lf:S0 :peržegnojimas_sense_I .

:sense_I1 lf:S1
            :Diewas_sense_II.

:sense_I1 lf:S2
            :vyras_sense_I, :moteris_sense_I.
```

Figure 3: Instantiation of the data model for *peržegnoti* entry.

3. The tool *Dec Facile*

In line with modern web application development practices, the architecture of the *DEC Facile* system is divided into a frontend (FE) and a backend (BE). The FE interface operates directly in the user's browser and is developed using Javascript, HTML, and CSS. Conversely, the BE component is implemented in Java and connects to a triple store to provide the services required by the FE.

BE is powered by LexO-server[8] [28], a free and open-source backend built on the semantic repository GraphDB. It is implemented as a collection of REST (Representational State Transfer) services based on the HTTP protocol and exchanges data in JSON format. The services adhere to the OpenAPI specification, which facilitates the description, production, consumption, and visualization of RESTful APIs in a machine-readable format. LexO-server originates from LexO-lite [29], a full-stack tool designed for editing OntoLex-Lemon resources. LexO-server manages both lexical and dictionary layers and has been enhanced to handle lexical functions.

As for the FE, the *DEC Facile* tool provides a streamlined interface for exploring and editing lexical and semantic information across multiple languages. As illustrated in Figure 4, the interface consists of several panels and functionalities, each designed to support the detailed construction and enrichment of an ECD[9].

Search Panel. This panel enables users to search for specific dictionary entries in multiple languages. The language selection is conveniently located next to the dictionary entry field, marked by the "@" icon, allowing users to toggle between languages (e.g., "lt" for Lithuanian). Results display dictionary entries alongside their part-of-speech (POS), forms, and associated

[8] The code is free and open source, available at the following link: https://github.com/andreabellandi/LexO-backend.
[9] The following link provides access to the *DecFacile* interface code: https://github.com/Abdoumasbah/DecFacile.

senses. These senses are hierarchically structured according to previously outlined principles, ensuring clarity and alignment with the linguistic framework.

Definition and Examples Panel. When a sense is selected, its corresponding definition and contextual examples are displayed in the right-hand panel. Users have the flexibility to modify the definition or add examples if they are missing in the source resource. This feature not only ensures adaptability but also allows for the enrichment of lexical data directly within the tool, supporting custom refinements and updates.

Lexical Functions Panel. This feature facilitates the linking of a selected sense from a dictionary entry to another sense through a specific lexical function (here e.g. S_0). In addition, users can click on the ad hoc tab (labeled LF) on the right-hand side to access a comprehensive list of all lexical functions managed by the LexFom module. Selecting a lexical function displays a detailed box with its definition and an illustrative example. This functionality is particularly useful for users unfamiliar with the linguistic theory underlying lexical functions, offering accessible explanations and practical examples. Consequently, it simplifies the construction of EDCs by providing both theoretical context and practical guidance.

Government Pattern Panel. Currently under development, this panel will enable users to specify the syntactic regime of a sense by defining the morphological realizations of the SemAs introduced by that sense.

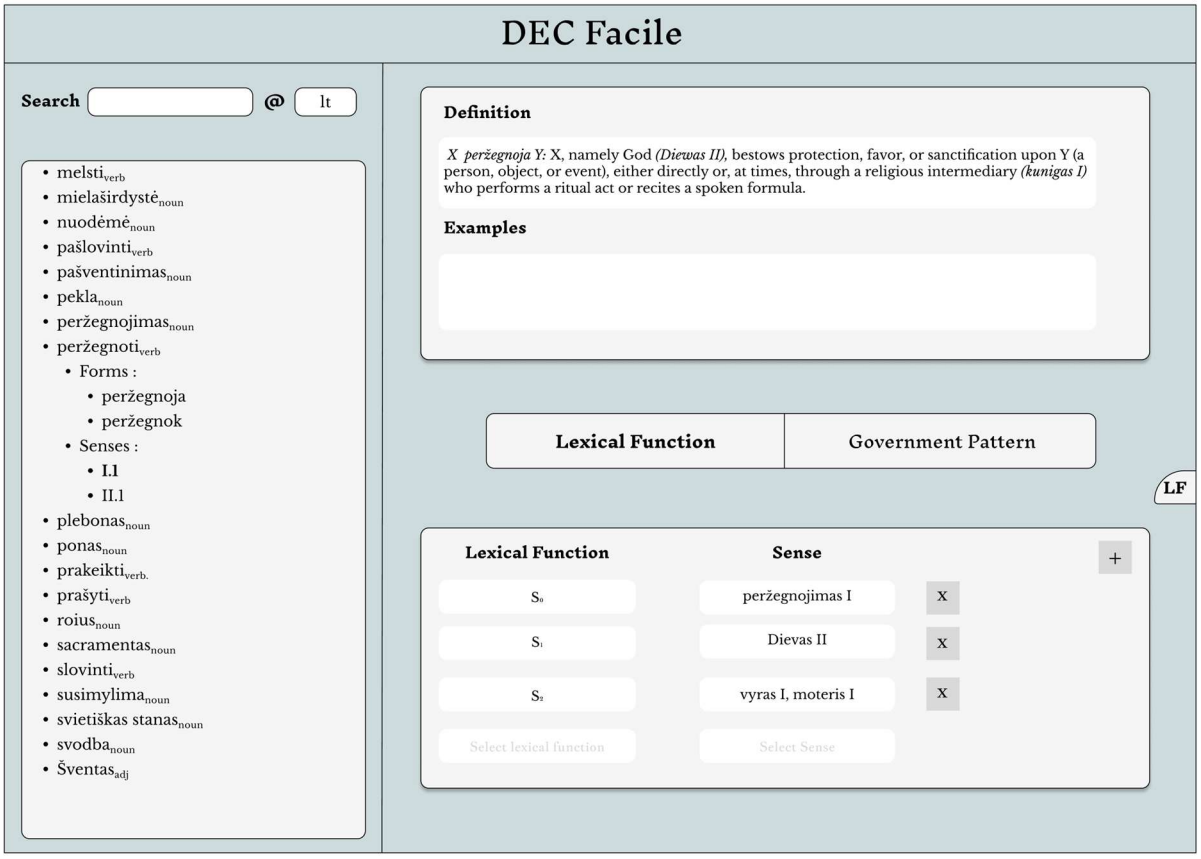

Figure 4: Mock-up of *DEC facile* GUI.

4. Conclusions

In this paper, we have presented an initial prototype of *DEC Facile*, a tool designed to assist users in constructing Explanatory Combinatorial Dictionaries, based on the theoretical framework of Explanatory Combinatorial Lexicology. Developed using semantic web technologies, the tool currently offers fundamental functionalities. Further development is planned, with a focus on

two main areas: enhancing the prototype to support terminology construction for the PRIN project "Old Words for a New World: Translating Christianity to Baltic Pagans" and implementing the Government Pattern using the SynSem module of OntoLex-Lemon. As *DEC Facile* evolves, its features will be refined through ongoing collaboration with experts in lexicography and terminography, who apply this theoretical framework in their work.

Acknowledgements

The research presented in this paper was conducted within the framework of the PRIN project "Old Words for a New World: Translating Christianity to Baltic Pagans", coordinated by Professor Pietro U. Dini and financed by the European Union – Next Generation EU, Mission 4 Component 1 CUP: B53D23014130006.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] P. Cimiano, J. P. McCrae, P. Buitelaar, Lexicon model for ontologies: Community report. W3C Ontology-Lexicon Community Group, 2016. URL: https://www.w3.org/2016/05/ontolex/.

[2] I. Mel'čuk, A. Clas, A. Polguère, Introduction à la lexicologie explicative et combinatoire, Duculot, Bruxelles, 1995.

[3] A. Žolkovskij, I. Mel'čuk, O vozmožnom metode i instrumentax semantičeskogo sinteza, Naučno-Texničeskaja Informacija 5 (1965) 23–28.

[4] A. Žolkovskij, I. Mel'čuk, O semantičeskom sinteze, Problemy kybernetiki 19 (1967) 177-238.

[5] A. Polguère, La théorie Sens-Texte, Dialangue : bulletin de linguistique 8-9 (1998) 9-30.

[6] L. Wanner (Ed.), Lexical Functions in Lexicography and Natural Language Processing, John Benjamins, Amsterdam/Philadelphia, 1996.

[7] A. Gelbukh, O. Kolesnikova, Lexical Functions and Their Applications, in: Gelbukh, A., Kolesnikova, O. (Eds.), Semantic Analysis of Verbal Collocations with Lexical Functions, volume 414 of Studies in Computational Intelligence, Springer, Berlin, Heidelberg, 2013, pp. 19-40. https://doi.org/10.1007/978-3-642-28771-8_2

[8] O. Vincze, E. Suárez, M. Ramos, An online collocation dictionary of Spanish, in: Proceedings of the 5th International Conference on Meaning-Text Theory Barcelona [Recurso electrónico], September 8-9, 2011, pp. 275-286.

[9] C. Boitet, M. Mangeot, G. Sérasset, The PAPILLON project: cooperatively building a multilingual lexical data-base to derive open source dictionaries & lexicons, in: Proceedings of the 2nd workshop on NLP and XML (NLPXML '02)- Volume 17, Association for Computational Linguistics, USA, 2002, pp. 1-3. doi: https://doi.org/10.3115/1118808.1118813.

[10] M.C. L'Homme, Fonctions lexicales pour représenter les relations sémantiques entre termes, Traitement automatique de la langue (TAL) 43/1 (2002) 19-41.

[11] M.C L'Homme, Lexical Semantics for Terminology: An introduction, Jhon Benjamins, Amsterdam/Philadelphia, 2020.

[12] M.C. L'Homme, J. Dancette, Modélisation des relations sémantiques dans un dictionnaire spécialisé bilingue, in : A. Clas, H. Awaiss, J. Hardanne (Eds.), L'éloge de la différence : la voix de l'autre. Actes des 6e Journées du réseau LTT, coll. Actualité scientifique, Paris, AUF, 2001, pp. 385-400.

[13] J. Dancette, Les relations lexico-sémantiques dans un dictionnaire spécialisé, in: Th. Szende (Ed.), Le français dans les dictionnaires bilingues, Champion/Slatkine, Paris et Genève, 2006, pp. 144-156.
[14] G. Sérasset, Un éditeur pour le dictionnaire explicatif et combinatoire du français contemporain, in : Journées lexique du PRC-CHM, Grenoble, France, 1996, pp.131-138.
[15] N. Gader, V. Lux-Pogodalla, A. Polguere, Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor, in: Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III), The COLING 2012 Organizing Committee, Mumbai, 2012, pp. 109–125.
[16] T. Berners-Lee, J.A. Hendler, O. Lassila, The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American CCLXXXIV/5 (2001) 34-43. doi: https://doi.org/10.1038/scientificamerican0501-34.
[17] Ch. Bizer, T. Heath, T. Berners-Lee, Linked data: The story so far, in: A. Sheth (Ed.), Semantic services, interoperability and web applications: emerging concepts, 1st ed., IGI Global, Hershey, Pennsylvania, 2011, pp. 205-227.
[18] M. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3, 2016.
[19] C. Chiarcos, S. Nordhoff, S. Hellmann (Eds.), Linked Data in Linguistics, Springer, Berlin, Heidelberg, 2012. doi: https://doi.org/10.1007/978-3-642-28249-2.
[20] P. Cimiano, P. Buitelaar, T. Declerck, A. Frank, S. Racioppa, M. Kiesel, R. Engel, M. Romanelli, D. Sonntag, B. Loos, V. Micelli, R. Porzel, LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies, in: Proceedings of the OntoLex 2006 - Ontologies and Lexical Resources Interfacing Ontologies and Lexical Resources for Semantic Web Technologies, Workshop at the LREC, Genoa, Italy, 27th May 2006, European Language Resources Association (ELRA), pp. 28-32.
[21] P. Cimiano, P. Haase, M. Herold, M. Mantel, P. Buitelaar, LexOnto: A Model for Ontology Lexicons for Ontology-based NLP, in: Proceedings of the OntoLex07 Workshop (held in conjunction with ISWC'07), Busan, Korea, 11 November 2007.
[22] E. Montiel-Ponsoda, G. Aguado de Cea, A. Gómez-Pérez, W. Peters, W., Modelling multilinguality in ontologies, in: Proceedings of the 21st International Conference on Computational Linguistics (COLING), 8-22 August 2008, Manchester, UK, pp. 67-70.
[23] P. Cimiano, P. Buitelaar, J. McCrae, M. Sintek, Lexinfo: A declarative model for the lexicon-ontology interface, Web Semantics: Science, Services and Agents on the World Wide Web 9 (1) (2011) 29-51.
[24] J. P. McCrae, J. Bosque-Gil, J. Gracia, P. Buitelaar, P. Cimiano, The Ontolex-Lemon model: development and applications, in: Proceedings of eLex 2017 conference, pp. 19-21.
[25] M. Mallia, M. Bandini, A. Bellandi, F. Murano, S. Piccini, L. Rigobianco, V. Quochi, DigItAnt: a platform for creating, linking and exploiting LOD lexica with heterogeneous resources, in: Proceedings of the 9th Workshop on Linked Data in Linguistics: Resources, Applications, Best Practices (LDL-2024)@LREC-COLING-2024, ELRA Language Resources Association (ELRA), pp. 55-65.
[26] S. Piccini, A. Bellandi, E. Giovannetti, A Model for Representing Diachronic Terminologies: the Saussure Case Study, DHQ: Digital Humanities Quarterly 15(3) (2021).
[27] A. Fonseca, F. Sadat, F. Lareau, Lexfom: a lexical functions ontology model, in: Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V), 2016.
[28] A. Bellandi, Building linked lexicography applications with LexO-server, Digital Scholarship in the Humanities 38.3 (2023) 937-952.

[29] A. Bellandi, LexO: an open-source system for managing OntoLex-Lemon resources, Language Resources & Evaluation 55 (2021), 1093–1126. doi: https://doi.org/10.1007/s10579-021-09546-4.