ENCO: Conceptualising a Multilingual Corpus-Based Terminological Database for Environmental Economics

Alberte Fernández Castro
European Master in Lexicography (EMLex) student
E-mail: alberte.fernandez.castro@gmail.com

Abstract

This paper conceptualizes an online multilingual corpus-based terminological database for the domain of environmental economics named ENCO. Drawing inspiration from prior works, ENCO aims to be versatile and user-friendly, serving a wide range of professionals. The database will feature three primary functions: translation, understanding, and communication. These functions encompass informative queries, providing comprehensive understanding with examples and extra-linguistic information, and terminology management, offering tools for generating and editing lists and repertoires. ENCO will adopt an ontology-based organization, encompassing both semantic relations and organization at different concept levels. This approach ensures a comprehensive and nuanced representation of terminological data, aiming to address the needs of a diverse range of users. The theoretical proposal lays the groundwork for future implementation, detailing the step-by-step planning process and addressing potential challenges.

Keywords: Multilingual terminological database; environmental economics; terminology management; corpus

1. Introduction

Environmental economics has gained importance in recent years due to the challenges that the new climate situation has triggered. Taking the environment into consideration when carrying out any economic activity has turned the conception of the world in financial terms upside down. Due to the growing popularity of the subject, the number of publications and various forms of media concerning this domain has increased, and with it, the terminology associated with environmental economics has become prominent in the technical language used.

As an interdisciplinary field, environmental economics draws from other disciplines such as environmental science, economics or finance and shares terms with these other areas of knowledge (Bretschger & Pittel 2020). This interdisciplinary nature makes the handling and study of its terminology particularly interesting (Kirchherr, Reike, and Hekkert 2017). This work will focus on the compilation and description of a database for environ-

mental economics (henceforth ENCO) that will be oriented to different purposes and addressed to a wide group of users.

Throughout the paper the theoretical foundations of ENCO, the different features of the database and its functionalities will be described. Being a theoretical approach, the different challenges and problems encountered will be specified together with the solutions designed to overcome them. Regarding the nature of ENCO as product itself, this conceptualisation will be based on both theoretical frameworks of terminology and lexicography following a mixed approach, with the main objective of optimising the theoretical foundations from which the database is conceived and extracting the relevant ideas from each theory.

2. Theoretical Frameworks

2.1 Previous and Current Terminological Work for the Domain of Environmental Economics

The domain of environmental economics does not feature a great deal of terminological work carried out in recent years. Due to its interdisciplinary character and its close relationship with other disciplines such as environmental sciences, ecology, economics and political science, terminological work concerning these other fields of knowledge overshadowed environmental economics and condemned it to be understood as a subfield.

In terms of lexicographical and terminological production, different works related to the aforementioned fields have been published. Here are examples such as, *Dictionary of Ecology and Environmental Science***,** *Dictionary of Ecology and Environmental Science* (Ahad & Ferdous 2019), *Oxford Dictionary of Ecology* (Allaby 2010), *Dictionary of Environmental Science and Technology* (Porteus 2008), *Dictionary of Environmental and Climate Change Law* (N.A. Robinson 2013), *Dicionário de Direito Ambiental: terminologia das leis do meio ambiente* (Krieger 1998), *A Dictionary of Climate Change and the Environment: Economics, Science and Policy* (Grafton et al. 2012) and *Dictionary of Ecological Economics: Terms for the New Millennium* (Haddad & Solomon 2023). All these works touch in one way or another on the domain of environmental economics, although in no case are they intended to deal with this field in its entirety. On the other side, some works such as *Dictionary of Environmental Economics* (Perelet, Mason, Markandya, & Taylor 2001), indeed do a good job compiling knowledge from the domain in its entirety, but the repertoire of lexicographical work for this area is still very limited in comparison with other related fields of study. The limited number of works exclusively dedicated to this domain highlight the need for further research in compiling a comprehensive and dedicated database that addresses the diverse terminological nuances, complexities and main topics within this realm.

2.1.1 State of Environmental Economics Terminology in Current Termbases

For the conception of ENCO, a preliminary study was carried out to identify and understand the current state of the terminology in the field of environmental economics within the available terminological databases (term banks). For that purpose, four terminological

databases (two general and 2 specialized ones) where selected to test how terms of the domain of environmental economics are registered and displayed in their term records.

- The chosen terminological databases are:
- Termium PLUS® The Government of Canada's terminology and linguistic data bank
- IATE Interactive Terminology for Europe
- AGROVOC Multilingual Thesaurus
- GEMET General Environmental Multilingual Thesaurus

The following results provide a small insight into the different ways these terms are included in the selected database and highlight once more the dispersion regarding their categorization.

Term occurrences of the lemma carbon tax			
	Number of records	Field/domain	Own term record
TERMIUM	11 records	Environmental economics	Yes
IATE	5 records	Environment, taxation, finance	Yes
AGROVOC	0 records	None	No
GEMET	0 records	None	No

Table 1: *Occurrences of carbon tax*

Term occurrences of the lemma (environmental) externality			
	Number of records	Field/domain	Own term record
TERMIUM	14 records	Environmental economics, economics, finances, economic planning	Yes
IATE	14 records	Environment, taxation, finance	Yes
AGROVOC	0 records	None	No
GEMET	1 record	None	No

Table 2: *Occurrences of (environmental) externality*

Based on the results obtained from these explorations, the low and irregular presence of the category "environmental economics" can be confirmed. This shows the under-representation of this field in the current termbase landscape and confirms the need to explore the creation of a specialised database for this domain.

2.2 Frame-Based Terminology Approach

Both the structure and conceptualization of ENCO draw heavily from MiliMarco (Faber & León-Araúz 2019). ENCO relies on the categorization and classification of terms, considering their underlying semantic implications and their behaviour within the semantic space of texts. This approach facilitates the creation of a network that characterizes the features of frame-based terminology (FBT) (Faber 2009). In this theoretical framework, terms are portrayed as semantic networks with a strong focus on use of the corpus to define and explore these relations (Faber & León-Araúz 2019: 5). FBT also employs a top-down and a bottom-up approach, combining both extraction methods using multilingual corpora and the usage of specialized dictionaries and termbases to complement it (Faber Benítez 2009: 124).

3. Design of ENCO

The conception of ENCO will draw both from lexicographical and terminological theory to ensure that all the users need are met. Following Dobrina (2015), different steps have been carried out to plan this termbase.

3.1 User Needs and Target Groups

For addressing user needs and defining the target group for this study, ENCO will focus on 3 main functions: translation, understanding and communication. Faber (2009: 109) emphasizes the addition of extra-linguistic information alongside the language-equivalent with the objective of setting up the possible options for a translation. For an effective communication, as highlighted by Cabezas-García and León-Araúz (2023), it is important the use of precise and contextually relevant terminology in translation contexts. This consideration becomes crucial in ensuring clarity and comprehension within the specified user group. Additionally, a term bank that serve as model for ENCO is the TriMED (Vezzani, Di Nunzio, and Henrot 2018), which facilitates both translation and communication from experts to laypeople. It is used as a valuable resource for enhancing understanding and comprehension. Recognizing and integrating these insights into ENCO will contribute to a more accurate and useful representation of the terminology of this field.

In the conceptualization of a dictionary (in this case termbase) as a utility product (Gouws 2003: 17), its value is linked to its utility to the following target groups: the primary target group is the one composed by translators and professionals specialised in fields such as economics, accounting, auditing, law, and engineering, reflecting the utility of the dictionary for experts in various domains not only for translation tasks but also other matters that require deep understanding of the terminology of environmental economics. By leveraging FBT, this project ensures that terms are not only translated accurately but also contextualized, enhancing both the precision and usability of the termbase for translators. This methodology will facilitate the translation process by offering detailed conceptual structures and relations, thereby supporting translators in producing more coherent and contextually appropriate translations.

As secondary target group, politicians, heads of state, journalists, and individuals in public positions would find it useful for consulting and addressing language difficulties. Another target group would be composed by linguists and students that stand to benefit from ENCO for academic purposes. This broad audience, to which this database is intended, em-

phasizes its versatility as a practical and indispensable tool for a wide array of individuals with distinct language needs and professional backgrounds.

3.2 Corpus Compilation, Curation & Creation of the Candidate Term List

As for the nature of ENCO as a multilingual terminological database, the process of corpus compilation involves an exhaustive and meticulous selection of texts that are representative of the domain of environmental economics. For this project a corpus of around 5M words was compiled from different kinds of documentation that directly or indirectly deal with the topic of environmental economics. The content of the corpus is made of resources that fall into the licence of Open Access.[1]

The ENCO Corpus was loaded into the corpus management software Sketch Engine (SkE) (Kilgarriff et al. 2004). Once the full corpus is uploaded, the next step is using the "Keyword Extraction" tool to retrieve both single – and multiword terms. SkE provides to the user a simple and intuitive way to create candidate lemma list and export them to different formats for their later curation. The outcome of the process was extracted in .csv and a list of 200 terms was following a frequency criterium.

An initial list of 200 terms is created, encompassing both single – and multi-word terms, sorted by frequency. Further refinement was achieved through manual curation, resulting in a curated list of 173 lemma candidates.[2] To enhance the organization and conceptual grouping of terms, an initial clustering process was carried out. This allows to form coherent groups based on semantic relationships among the terms. Subsequently, using Protégé (Musen 2015), a manual organization of term into an ontology was carried out. This step aimed to ensure accuracy and coherence in the representation of terminological relationships within the ontology.

3.3 Definitions

The creation of definitions was a challenge from the beginning. Initially, the intention was to follow the methodology described in Chambó & León-Araúz (2023: 591) using definitional KPs, but this turned out not to be of much help as the texts compiled in the ENCO Corpus are mostly addressed to experts, so that no explanations of the terms used are included. Considering this as an ongoing project, for now definitions would be crafted based on keywords in context (KWIC) and the skill of the terminologist. To achieve this, a frame-based approach based on the methodology described in Durán-Muñoz (2016) will be applied. These approach focuses on the compilation of an ontology-based conceptual classification, applying it to definitional templates that are flexible enough to register the necessary information and capture the hierarchical and non-hierarchical relations between concepts. The final methodology concerning this part of the ENCO conception is yet to be determined.

1 Due to limitations in accessing papers, specifically financial constraints and subscription barriers, I was unable to use a more diverse range of sources beyond Open Access materials.

2 As this is only a theoretical work, it is important to note that if the research progresses, a field expert must validate the term list and assess the entire process to ensure accuracy and comprehensiveness.

4. Structure and Features of ENCO

4.1 Microstructure

As stated by Bowker (2018), the creation of this terminological database will benefit both from terminological and lexicographical theory. In this case, the layout of the term record would follow a more dictionary-like entry, ensuring a quick understanding of the information presented and a user-friendly access to all the features of ENCO. As in the microstructure of the term records within the database, various elements are presented to offer a detailed examination of each term. The items presented in the term records of ENCO are the following (Wiegand et al. 2010; 2017; 2020a; 2020b; 2020c; 2020d):

- Lemmazeichengestaltangabe [LZGA] (Item given the form of the lemma sign) (WLWF Vol. 3 p. 215).
- (Übersetzungs)äquivalentangabe [ÄA] (Item giving an equivalent) (WLWF Vol. 2 p. 547)
- Definitionsangabe [Def.A] (Item giving the definition) (WLWF Vol. 2 p. 14)
- Explizite Wortartangabe [WAA.expl] (Explicit item giving the part-of-speech) (WLWF Vol. 4-5 p. 113)
- Corpusbeispielangabe [CorBeiA] (Item giving a corpus example) (WLWF Vol. 1 p. 750)
- Diatechnische Markierungsangabe [MarkA.diatech] (Item indicating technical labelling) (WLWF Vol. 2 p. 54)

This microstructural level ensures that the term records are comprehensive, providing users with a wealth of information for each entry in the database presented in a consistent way (Gouws & Prinsloo 2005: 115). The information organized at the microstructural level (items) will be organized

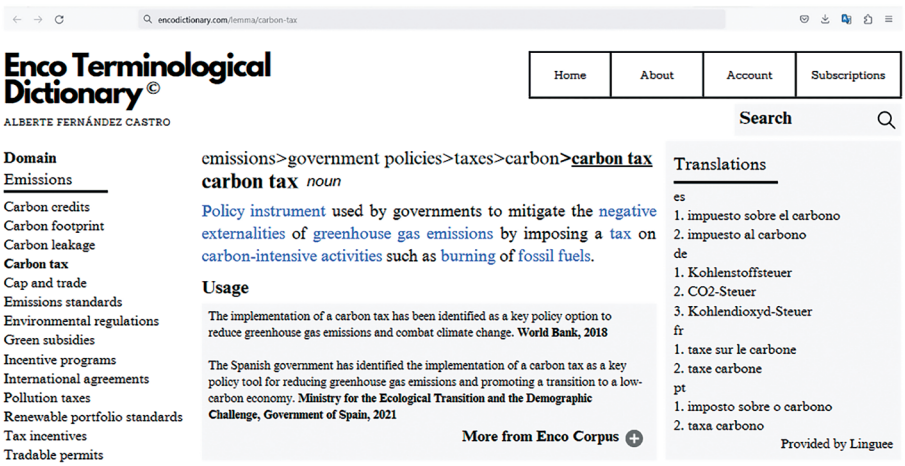

Figure 1: *Example of a term record*

displaying a series of structural indicators that will lead the user to other items and search areas (Gouws & Prinsloo 2005: 116). These structural indicators belong to the article structure (in this case term-record structure). To sum up, the term-record will present a series of items that will provide information to the user and will be organized and accessible through the structural indicators (Gouws 2014: 161).

In image 1 one, a prototype of the layout of a term record is presented. Different search areas can be clearly distinguished from one another with a proper space dedicated to present a certain kind of information. Items are denoted with both typographical and non-typographical structural indicators.

4.1.1 Term Record Structure and Search Areas

The term record is based on a dictionary-like structure. In contraposition to the traditional article format (term record) divided in comments, as stated in Gouws (2014: 167), the online environment facilitates alternative methods of organizing and accessing data, moving away from reliance on comments as integral components of articles. Prioritizing the allocation of lexicographic data to easily accessible search areas takes precedence over strictly adhering to a division on comments on form and semantics.

Clearly distinguished search areas can be spotted: on the central side there is the term with an item that indicates the part of speech, just above is presented the position of the term in a hierarchy, this is complemented by the area on the left which shows a more general level of classification into groups. The classification and inclusion of the term in a group is based on the aforementioned ontology and will allow the users to navigate through related terms and located themselves in the lexical environment the surrounds the current term they are consulting. At the bottom, examples directly extracted from the ENCO corpus are provided and, in the right side, the translation in different languages. The definition of the term, placed in the centre of the term record, presents clickable items that corresponds with other terms in the database. These features allow the user to employ different pathways to navigate through ENCO.

4.2 Ontology-Based Access Structure

“The access structure determines the search route a user must follow to reach an entry in a dictionary” (Gouws & Prinsloo 2005: 165). In this case, the users have at their disposal different ways to access the term-records based on their preferences and needs. ENCO provides resources such as a search bar (image 1) and a direct access to the ontology that underlies under the integrated system. The ontology will structure the content of ENCO corresponding to a concept-based organisation. It will be based directly on the ENCO Corpus composed of texts from the domain of environmental economics (Moreno and Pérez 2000: 1). The classes present in the ontology will allow the ENCO content to be grouped in such a way that the terms are hierarchically classified (Roussey et al. 2018: 223) and linked following semantic relations such as “causes”, “is part of”, “is a subfield of” (Faber & León 2019: 17; Condamines 2018: 339). The significance of these knowledge structures extends to professionals such as translators and technical writers, allowing them to go in depth into the terms and understand how they behave in specialized text, leading to a better result in

their tasks. The access to an ontology composed by semantic relations, not only enriches the understanding of terms but also enhances the ability to navigate and comprehend the behaviour of the environmental economics terms in specialized texts.

The initial clustering of the primary term list will serve as the base for developing the classes and the relations between terms that will be featured in ENCO. As already stated in 3.3., the ontology would be primarily developed using the software Protégé for a further integration in the whole system of ENCO.

4.3 Knowledge Graph

The presence of a knowledge graph such in WIPO Pearl, EcoLexicon (Gil-Berrozpe, León-Araúz & Faber 2019) and BabelNet (Navigli & Ponzetto 2012) is considered. The main objective of this feature is to visualize in a dynamic way the relations between the terms integrated in ENCO, therefore ensuring a deeper understanding of the terminology of environmental economics by the users (Hogan 2021: 2). This way, the user will have access to the broad semantic network thar lies behind the terminology of the domain displayed in a visual and user-friendly interface that will allow the navigation through the content of ENCO. This knowledge graph will be based on an ontology (Hogan et al. 2021: 2024) that would display a hierarchically organized set of semantic relations with the chance to directly access the term records.

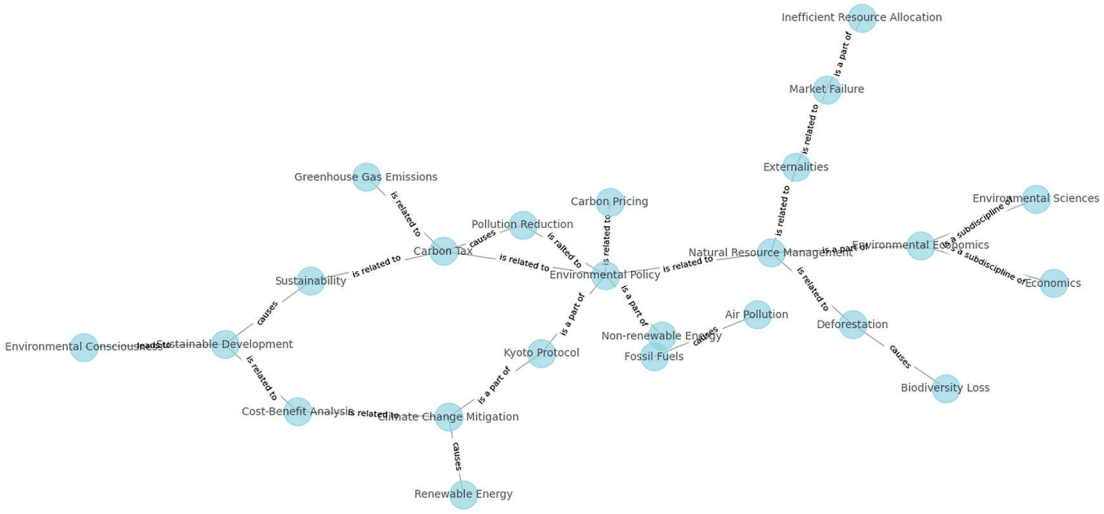

Figure 2: *Example of a knowledge graph*

4.4 Compatibility with CAT Tools

Users have the chance to further employ ENCO by terms and their corresponding translations to personalized lists. This feature allows translators and other professionals to curate their individualized glossaries within the system. Also, these user-generated lists can be

exported in .csv format, facilitating seamless integration into Computer-Aided Translation (CAT) tools like MemoQ and SDL Trados. This functionality empowers translators to extend their workflow efficiency by incorporating customized terminology repositories derived from the ENCO database, aligning with the specific linguistic demands of their projects.

5. Conclusion, Challenges and Further Research

In this paper the conceptualisation of the ENCO terminological database was described. Both the theoretical foundations and the main features of ENCO terminological database were described, with the main objective of this overview being to lay the foundations for a real project on a larger scale. Different challenges and limitations were described in the paper. Starting with the more theoretical part, one of the main challenges, which is still to be solved, was to conceive the nature of ENCO as a terminological database or a dictionary, since this product presents characteristics of both, potentially raising some doubt in its classification. Focusing on the practical side, the creation of an integrated system would be a major challenge involving advanced technical knowledge. Also, certain aspects such as the creation of definitions and the organisation of the ontology (relations, classes...) will have to be reviewed in order to establish a clear methodology. In conclusion, ENCO is a solid project that lays the foundations for future work that will contribute to the field of environmental economics and the study of its terminology. Furthermore, the integration of all the functions described in the paper will result in a useful product that will serve as a tool for different functions and different user groups.

References

Ahad, A., & Ferdous, A. (2019). *Dictionary of Ecology and Environmental Science***.** Himachal Publication, Dhaka.

Allaby, M. (Ed.). (2010). A *Dictionary of Ecology***.** Oxford Quick Reference.

Art, H.W. (1993). *Dictionary of Ecology and Environmental Science (Henry Holt Reference Book)* (1st ed.). Henry Holt & Co.

Bowker, L. (2018). Terminology and lexicography. In Fuertes-Oliveira, *The Routledge Handbook of Lexicography* (pp. 138-151). Routledge.

Bretschger L, Pittel K. Twenty Key Challenges in Environmental and Resource Economics. Environ Resour Econ (Dordr). 2020;77(4):725-750. doi: 10.1007/s10640-020-00516-y. Epub 2020 Oct 16. PMID: 33082623; PMCID: PMC7561503.

Cabezas-García, M. & León-Araúz, P. (2023). Term and concept variation in climate change communication. *The Translator***.** 1-21. 10.1080/13556509.2023.2182168.

Chambó, S., & León-Araúz, P. (2023). Operationalising and Representing Conceptual Variation for a Corpus-driven Encyclopaedia. In *Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference. Brno, 27–29 June 2023.* Brno: Lexical Computing CZ s.r.o.

Condamines, A. (2018). Terminological knowledge bases. In Fuertes-Oliveira, *The Routledge Handbook of Lexicography* (pp. 335-349). Routledge.

Dobrina, C. (2015). Getting to the Core of a Terminological Project. In H.J. Kockaert & F. Steurs (Eds.), *Handbook of Terminology* (pp. 180–202). Amsterdam: John Benjamins.

Durán-Muñoz, I. (2016). Producing frame-based definitions: A case study. Terminology. *International Journal of Theoretical and Applied Issues in Specialized Communication*, 22(2), 223-249.

Faber Benítez, P (2009). The cognitive shift in terminology and specialized translation. *MonTI. Monografías De Traducción E Interpretación*, (1), 107-134. https://doi.org/10.6035/MonTI.2009.1.5

Faber, P., & León-Aráuz, P. (2019). Frame-based terminology applied to military science: transforming a glossary into a knowledge resource. *Lexicography*, 6(2), 105-131.

Fuertes-Olivera, P. A. (2017). *The Routledge Handbook of Lexicography*. Routledge.

Gil-Berrozpe, J.C., León-Araúz, P. & Faber, P. (2019). Ontological Knowledge Enhancement in EcoLexicon. In *Proceedings of the eLex 2019 conference: Electronic lexicography in the 21st century*, edited by Kosem, I., Zingano-Kuhn, T., Correia, M., Ferreira, J.P., Jansen, M., Pereira, I., Kallas, J., Jakubíček, M., Krek, S. & Tiberius, C., pages 177-197. Brno: Lexical Computing CZ, s.r.o.

Gouws, R. (2003). Milestones in metalexicography. *XVII International Congress of Linguists: Prague: 24-29 July 2003: State-of-the-art paper: Lexicology and Lexicography***.**

Gouws, R. H. (2014). Article structures: moving from printed to e-dictionaries. *Lexikos***,** 24(1), 155-177.

Gouws, R.H. & Prinsloo, D.J. (2005). *Principles and Practice of South African Lexicography***.** Stellenbosch: SunMedia.

Grafton, R.Q., Nelson, H.W., Lambie, N.R., & Wyrwoll, P.R. (2012). *A Dictionary of Climate Change and the Environment: Economics, Science, and Policy.* Edward Elgar Publishing. ISBN: 978-1-84980-387-8.

Haddad, B.M., & Solomon, B.D. (Eds.). (2023). *Dictionary of Ecological Economics: Terms for the New Millennium.* Edward Elgar Publishing. ISBN: 978-1-78897-490-5.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. D., Gutierrez, C., Labra Gayo, J. E., Kirrane, S., Neumaier, S., Polleres, A., Navigli, R., Ngonga Ngomo, A.-C., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys (Csur), 54*(4), 1-37.

Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). The Sketch Engine. In *Proceedings of the 11th EURALEX International Congress* (pp. 105-116).

Kirchherr, J., Reike, D., & Hekkert, M. (2017). Conceptualizing the circular economy: An analysis of 114 definitions. *Resources, Conservation and Recycling, 127*, 221-232. https://doi.org/10.1016/j.resconrec.2017.09.005

Krieger, M.G. (1998). *Dicionário de Direito Ambiental: Terminologia das Leis do Meio Ambiente.* Editora da Universidade, Universidade Federal do Rio Grande do Sul.

Moreno, A. and Pérez, C. (2000). Reusing the Mikrokosmos Ontology for Concept-based Multilingual Terminology Databases**.** *In Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, Athens, Greece. European Language Resources Association (ELRA).

Musen, M.A. (2015) *The Protégé project: A look back and a look forward*. AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4). DOI: 10.1145/2557001.25757003.

Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence***,** 193, 217-250.

Perelet, R., Mason, P., Markandya, A., & Taylor, T. (2001). *Dictionary of Environmental Economics* (1st ed.). Routledge. https://doi.org/10.4324/9781849774253

Porteous, A. (2008). *Dictionary of Environmental Science and Technology*. John Wiley & Sons.

Robinson, N.A., Xi, W., Harmon, L., & Wegmueller, S. (2013). *Dictionary of Environmental and Climate Change Law.* Edward Elgar Publishing.

Roussey, C, Hernandez, N & Zargayouna, H. (2018). Domain ontologies. In Fuertes-Oliveira, *The Routledge Handbook of Lexicography* (pp. 217-234). Routledge.

Vezzani, F., Di Nunzio, G. M., & Henrot, G. (2018). TriMED: A Multilingual Terminological Database. **In** *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)***.**

Wiegand, H., Gouws, R., Kammerer, M., Mann, M. & Wolski, W. (2020). *Band 3 I – U*. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110621716

Wiegand, H., Gouws, R., Kammerer, M., Mann, M. & Wolski, W. (2020). *Band 4 V – Z*; Nachträge und Gesamtregister A – H. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110621754

Wiegand, H., Gouws, R., Kammerer, M., Mann, M. & Wolski, W. (2020). *Band 3 I – U*. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110621716

Wiegand, H., Gouws, R., Kammerer, M., Mann, M. & Wolski, W. (2020). *Band 4 V – Z*; Nachträge und Gesamtregister A – H. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110621754

Wiegand, H.E., Gouws, R., Kammerer, M., Mann, M., Wolski, M. (2010). *Band 1 A – C*. Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110226119

Wiegand, H.E., Gouws, R., Kammerer, M., Mann, M., Wolski, M. (2017). *Band 2 D – H*. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110341027