Qf Lingüístics

ojs.uv.es/index.php/qfilologia/index

Per a citar aquest article: Calañas Continente, José Antonio & Domínguez Vázquez, María José. 2023. "About Sustainable and Socially Responsible Lexicography: State of the Art". *Quaderns de Filologia: Estudis Lingüístics* XXVIII: 9-19.

doi: 10.7203/QF.28.27659

About Sustainable and Socially Responsible Lexicography: State of the Art

Lexicografía sostenible y socialmente responsable: Estado de la cuestión

José Antonio Calañas Continente
Universitat de València
j.antonio.calanas@uv.es

María José Domínguez Vázquez
Universidade de Santiago de Compostela
majo.dominguez@usc.es

1. Introduction

Systems for storing and managing lexical information have recently experienced a huge qualitative leap parallel to improvements in both computer technology and speed at spreading information. This context certainly allows for science to broaden and widen the spectrum of research objects, and consistent with this framework, lexicography, defined as "total of all activities directed at the preparation of a lexicographic reference work" (WLWF 3/2020: 224), evolves as well and opens up its research field to include new values and functions which clearly reflect upon items about which society is concerned, especially those related to sustainability and social responsibility. Both concepts are especially relevant in lexicography, since it is a scientific and cultural practice (Wiegand, 1983: 38). Accordingly, we agree with Geoffrey Williams (2016: 79) as he points out this fact:

> Lexicographers not only study language for what it is, the central tool for communication, but also provide the means by which a language, and its underlying cultural values, may be taught and given full value within a society.

Lexicographic work requires training and a set of skills, both intellectual and IT-related, among others. Therefore lexicographers –and not machines– bear full responsibility for the quality of the information (Kouassi, 2022: 174) and are also responsible for ensuring that this information corresponds to reality, taking into account its impact on the target group and society in general. This is called social responsibility (see section 2).

Lexicographic work needs to rely on new technologies, on the reuse of open data, on the interconnection of resources (which are easily manageable) as a way of ensuring their sustainability. Sustainability, though, has a wider reach (see section 3) and should also be understood in a broad sense, since the lexicographic work also requires servers and installations, staff, group managing, workflow organization, and last, but not least, adequate funding.

Those two aspects, sustainability and social responsibility, build the core of the articles in this issue of *Quaderns de Filologia – Estudis lingüístics*. Since both are increasingly relevant in society, the authors in this volume provide an overview of the most current orientations of lexicographic work, a task lexicographers set for themselves to account for changing values in society. Throughout the following pages, an overview of the state of the art in both sustainability and social responsibility is provided.

2. Sustainable Lexicography

Sustainability is a concept which applies mostly to both the rational use of natural resources and the reduction of the *human footprint* in the environment while optimizing (i. e., minimizing) the usage of applications and devices. Thus, a sustainable lexicography usually focuses on optimizing the many resources already available, doing this by generating synergies and interconnections which allow for feedback on effectiveness-proven systems. This section reviews the different lexicography-related approaches to sustainability.

2.1 *Environmental Sustainability*

The emergence of digital dictionaries or lexicographic tools leaving printed materials behind can be regarded as the beginning of this trend of creating resources while preserving the environment. A sustainable digital lexicography must rely on efficient data management, on servers complying with environmental certifications, as well as on technologies which enable a reduction

of the carbon print. Avoiding the usage of new storage devices, servers, etc. contributes as well to reducing electromagnetic pollution, thus improving working conditions for the people involved in linguistic research.

Environmental sustainability is also dealt with in the production of printed dictionaries, where the usage of recycled paper, environment-friendly production facilities, etc. has played a crucial role for some time already. Reducing or even avoiding the production of printed products (which become quickly outdated, thus requiring regular updates to be printed) result in more sustainable lexicography.

2.2 *Economic Sustainability*

Nowadays, a dramatic change in the editorial market has been brought about by the newer information technologies. Aside from purely economic questions, it is relevant to point out that lexicographic products (not only printed ones) require updating to deal with social evolutions and the subsequent evolution of vocabulary. At the same time, there are still gaps in terms of different resources: an obvious case when considering minority languages with scarce resources. The relationship between funding and lexicography is described by Colman (2016: 141) as follows:

> Unfortunately, financers of dictionaries are more and more reluctant to invest in long-term dictionary projects, as in a market-driven society such projects are often considered to be time-consuming and expensive. So, in lexicography it is not the natural (language) resources that are depleted but the financial resources. To sustain lexicography, lexicographers will need to convince funders that their investments are not a waste of time and money and that it is possible to optimize the workflow through responsible use of materials, products and financial resources. Reuse of materials and products in the lexicographic context can be achieved by reuse of the content of existing dictionaries.

Therefore, economic sustainability in lexicography seems to go hand in hand with the reuse of data and the interoperability of existing resources and data, as well as the use of techniques and strategies of natural language processing and artificial intelligence that make it possible to reduce time and resources in the optimization and elaboration of resources, without this implying a deterioration in quality (more about this aspect in section 2.3).

Another approach to economic sustainability focuses on the effective management of funds, whether public or private. It is essential, therefore, to stress the idea that

> behind the data, there is always lexicographical practice that calls on the skills and knowledge of lexicographers. Raising the profile means making clear to funders and evaluators that there is more to lexicography than meets the eye and that it deserves both recognition and funding (Williams, 2016: 87).

2.3 *Sustainability of Hard Products and Digital Data Management*

Lexicographic practices which optimize the many resources that already exist by generating synergies and interconnections that allow the feedback of systems of proven effectiveness find a place here; for example, applying already existing ontologies, e. g., those of WordNet, in new lexicographic tools enables those tools to take advantage of WordNet without the need to use computer storage space for such an extensive corpus, as well as to improve and evolve both the starting and the target resources.

Different clusters, infrastructures and strategic actions exemplify this shift in technological sustainability:

- European Network of e-Lexikography - Enel (https://www.elexicography.eu/): This lexicography network was a pioneer in working together for a broader and more systematic exchange of knowledge, as well as for the purpose of establishing common standards and solutions. It was a COST action with European funding.
- ELEXIS - European Lexicographic Infrastructure[1]: Another example of a commitment to sustainability from an institutional perspective is the creation of infrastructures such as ELEXIS. This infrastructure focuses on central issues in dictionary production such as:
  - The change in the lexicographic landscape in which dictionary bases are integrated as part of other resources, applications, or devices. Thus, the reuse and interconnection of existing data.
  - Difficulties in maintaining electronic resources and servers, as well as in updating the data contained in dictionaries.

[1] European Lexicographic Infrastructure | ELEXIS Project | Fact Sheet | H2020 | CORDIS | European Commission (europa.eu).

  - Difficulties related to the different data encoding systems, which often makes it impossible to reuse these data and therefore incompatible and impossible to integrate them to develop new resources.

  On their website (https://elex.is/tools-and-services/), ELEXIS offers resources either to create new dictionaries from scratch or to enrich existing ones, while always pointing at taking advantage of what is already available. Tools to convert, create, link, edit, enrich, and publish dictionaries are offered together with a schematic of how those tools interact to complete lexicographic projects.
- DARIAH: The Digital Research Infrastructure for the Arts and Humanities (DARIAH) aims to enhance and support digital research and teaching in the arts and humanities (DARIAH EU openaire.eu). DARIAH, which became ERIC (European Research Infrastructure Consortium) in 2014, develops, maintains, and operates this infrastructure. It aims at enhancing sustainable development in digitized research and education by bringing together research with new technologies and long-standing research with technology and technological advances, while focusing on a new society and its needs. Like ELEXIS, DARIAH offers different resources on its website (https://www.dariah.eu/tools-services/tools-and-services/), although in this case with a broader spectrum, that of the Digital Humanities. DARIAH is also an example of Open Science.
- CLARIN is a digital infrastructure that offers data, tools, and services to support research based on linguistic resources. It provides access to digital linguistic data. CLARIN's linguistic resources can be found in repositories (https://www.clarin.eu/content/data) as well as being accessible through the Virtual Observatory of Languages (https://vlo.clarin.eu/?1).
- INTELE (Infraestructura de Tecnologías del Lenguaje): It is a research network originally established by Spanish researchers. It mainly aims to "reduce the digital divide and promote new multidisciplinary lines of research in humanities and social sciences, participating in a digital transformation of the same with the help of language technologies" (https://ixa2.si.ehu.eus/intele/home).

Therefore, sustainability in lexicography involves creating common standards, interconnecting existing data and resources, and developing data conversion tools that allow their exploitation for the development of new

resources. This also implies economic sustainability, an increased productivity or use of human and economic resources, therefore, a lower investment in time and effort by working in a coordinated manner to achieve a common goal. In short, the possibility of reusing data from existing tools and their use in external applications leads us to a kind of *circular lexicographic economy* in which it is possible to make optimal use of what already exists and give a new meaning to the effort made previously. In the framework of sustainable lexicography, therefore, the concepts of interoperability and feedback, as well as collaborative lexicography, are central. The confluence of all these factors together with the digital revolution opens the horizon for the design and development of tools, resources, and dictionaries hitherto unimaginable. They are even more unimaginable if we consider the irruption of Artificial Intelligence in the development of products such as ChatGPT.

3. Socially Responsible Lexicography

In his 1971 study about *Business in contemporary society*, Johnson explains what he terms as "lexicographic view of social responsibility". In this definition, the goals of the enterprise, like those of the consumer, are ranked in order of importance and targets are assessed for each goal. These target levels are shaped by a variety of factors, but the most important are the firm's experience with these goals and the past performance of similar enterprises; individuals and organizations generally want to do at least as well as others in similar circumstances. Johnson said that "lexicographic utility theory suggests that strongly profit-motivated firms may engage in socially responsible behaviour. Once they attain their profit targets, they act as if social responsibility were an important goal- even though it isn't" (1971: 75).

Lexicography is a cultural practice whose main purpose is the creation of dictionaries and other resources. Dictionaries and lexical information systems document and preserve linguistic and social knowledge, meet the information needs of individuals, groups and nations, and serve to promote linguistic and cultural understanding and exchange. Databases, dictionaries and lexical information systems are repositories of high-quality knowledge that are considered reliable data sources for modern information societies.

In this sense, a socially responsible lexicography:

1. Brings together the past and the future. Thus, dictionaries, databases, etc. store knowledge of a society that must be comprehensible for future

generations. Therefore, it must document languages and linguistic varieties of societies and epochs, as well as social and cultural knowledge.

2. Forces lexicographers to reflect on the role of dictionaries, portals, or information systems (Villa-Vigoni-Theses, 2018) in bringing about changes in the values of society (Müller-Spitzer, 2022), a society that is digital, multilingual and increasingly interconnected. Dictionaries or lexical information systems and lexicographic products are no longer those static works that needed decades to renew their contents; updates can nowadays be made practically in real time. In such a scenario, lexicography plays a fundamental role in the transmission of new social roles and values.
3. Must reflect on the sources from which it collects data and the quality of the products it offers, as well as on the social and cultural impact of the content it transmits, among others.
4. Must meet the information needs of individuals, groups and nations and thus promote their understanding and linguistic-cultural exchange. To this end, it must promote dialogue with the users of the resources in favor of their usefulness and suitability to the recipients and their needs. In this sense, it is important to promote collaborative lexicography and, in short, a lexicography understood as an Open Science.
5. Has a social responsibility in the necessary integral and pluralistic description of the linguistic reality and in the consideration of intercultural diversity. It must avoid the perpetuation of stereotypes and prejudices. In this respect, we are called upon to reflect. Müller-Spitzer & Rüdiger (2022) point out that there are clearly stereotyped statements about men and women in the entries of modern corpus-based German dictionaries and indicate that "the representation of gender in dictionaries thus seems to be caught between language use and lexicographic-moral responsibility". In this sense, studies have been conducted on the *Contemporary Chinese Dictionary* (Hu, Xu & Hao, 2019), or Spanish dictionaries (Fuertes-Olivera & Tarp, 2022), therefore the awareness in realization with inclusive language is increasing.
6. Must be accurate in description, but also equitable and fair in terms of attention to the linguistic, cultural, religious, ethnic, etc. diversity of different communities[2]. Therefore, it must contribute to a communica-

[2] Note in this regard the critique of Crowley (2010: 1): "While lexicographers have often seen themselves as 'harmless drudges', scholars have increasingly pointed out that there are many aspects of lexicography that are inherently political".

tion that not only is effective, but also fair. It requires constant updating to avoid stereotypes or errors, but also to reflect changes in society.

7. Must support the development of standards and tools that allow their reuse by other communities with fewer resources or that are in a minority situation. Thus, in this way, fostering an awareness in favor of the preservation of diversity.

In short, under socially responsible lexicography we understand the delicate function of lexicography in favor of a complete and pluralistic description of reality, as well as in the awareness of new gender roles, new and old ideologies, and cultural differences and similarities. In turn, a socially responsible lexicography requires the participation of the user in the design of the resources, that is, a lexicography for the citizens by relying on the citizens. It is, consequently, a social lexicography from two different, yet complementary, perspectives: on the one hand, the social responsibility of lexicography in terms of the transmission of content, on the other, the citizen lexicography developed by those to whom the resources are addressed.

4. This volume

The structure of the volume progresses from issues related to lexicography in a new global and interconnected society, through socially responsible lexicography, to sustainable lexicography in a global and interconnected world, although all the articles show a clear interrelation between both the central axes of the volume. The compiled studies provide a clear view of different approaches to the qualifiers "socially responsible" and "sustainable".

The volume opens with the work of **M.ª Concepción Maldonado González**, "Prehistoria y ciencia ficción en lexicografía" (Prehistory and Science Fiction in Lexicography), which summarizes an overview of traditional commercial lexicography, addressing, in turn, the digital revolution, as well as the new tasks to be undertaken by lexicography as a discipline. This is defined by the author as a circular lexicography, in the sense of sustainable and socially responsible.

The concepts of socially responsible lexicography and glotopolitics are intertwined in the study by **Susana Rodríguez Barcia**, "Principios para una lexicografía socialmente responsable desde la perspectiva glotopolítica" (Principles for a socially responsible lexicography from a glotopolitical perspec-

tive). With the aim of deepening the concept of a socially responsible lexicography, the author proposes four central axes that any dictionary must fulfill to be framed within this typology: neutrality, identification of lexicographic sources, accessibility and updating.

The study by **Irene Renau**, "A corpus-based study of semantic neology of the Covid-19 pandemic", also focuses on the issue of updating the data provided by lexicography. Her work, which pursues the objective of analyzing semantic neology in the period of the Covid pandemic, is clearly related to the social responsibility of the resources when it comes to satisfying the consultation needs of the people who handle them. The description of methods, strategies and resources applied in this research point to an important feature of current and future lexicography: its sustainability.

Euphemisms are linked to different social and cultural aspects and are therefore relevant in terms of their social and communicative function. **Dolores Torres Medina** deals with this subject in her article "El eufemismo en diccionarios actuales español-inglés" (Euphemisms in current Spanish-English dictionaries). The author focuses on its marking and coding in Spanish-English bilingual lexicography and on the necessary contextualized information of the lexicon. The study analyzes the different markings of the lexicon related to death, physical appearance, social inequalities or illness, among others.

Two papers address specific resource proposals that focus on the relationship between lexicography and society:

An example of linguistics and lexicography for and with citizens is the *Lingua viva* platform, presented by **María Álvarez de la Granja** and **Elisa Fernández Rei**. After pointing out the central characteristics of this platform, the authors describe the *Tes a palabra!* dictionary production system. This free, networked, intuitive and freely accessible system allows the individual or collective-collaborative design of different lexicographic products with diverse functionalities and uses. The different fields of application of *Tes a palabra!* are an example of a socially responsible, but also sustainable, lexicography.

The connection between collaborative lexicographic resources and teaching/learning is another example of a lexicography oriented to society, especially if we consider the fact that language students are central addressees of lexicographic resources. In this sense, **Iván Arias Arias** and **Pau Bertomeu Pi** propose the use of collaborative lexicography resources such as *Wiktionary* for teaching Linguistics at the University. The authors propose different didactic

activities related to semantics and lexicography, also addressing issues such as minority languages, neologisms or pejorative or discriminatory lexicon.

The studies by Carlos Valcárcel and Carlos Periñán-Pascual deal with the concept of sustainability and the reuse of resources and data.

Also framed in principles of collaborative lexicography, "Reusing the structure of the *PortLex* dictionary to create a contrastive dictionary of Romance languages: a proposal" by **Carlos Valcárcel** suggests the creation of the online dictionary of Romance languages –*DiCoRom*– from a previous collaborative dictionary, *PortLex*. The author details how and what can be reused for the development of this new dictionary, which has Interlingua as its pivot language. The main purpose of *DiCoRom* is the learning and contrastive analysis of Romance languages, but not only: practicing metalinguistic and intercultural competences and raising awareness of linguistic diversity and the importance of preserving it are central axes of the resource.

As its title indicates, the contribution of **Carlos Periñán-Pascual**, "Un uso sostenible de WordNet en la inteligencia artificial", focuses on the sustainability of this lexical database and the role of artificial intelligence. The description of WordNet from the perspective of sustainability in Periñán-Pascual's work deals with three central criteria: extensibility, interoperability, and reusability. The study also points out the central role of this lexical database in the development of NLP applications, its resurgence in the last decade thanks to neural networks and predictive methods, and its importance for the optimization and development of new multilingual resources, among others.

In summary, all the compiled articles present different approaches to the concepts of sustainable and socially responsible lexicography, enriching each other and bringing coherence to the volume.

As editors, we would like to thank the authors for their contributions, as well as the entire team of reviewers of this volume, which is thematically linked to the project "Automatic and Sustainable Multilingual Semantic Tagger" (ESMAS-ES+; PID2022-137170OB-I00 funded through MCIN/AEI// FEDER "A way to make Europe").

References

Colman, Lut. 2016. Sustainable lexicography: Where to go from here with the ANW (Algemeen Nederlands Woordenboek, An Online General Language Dictionary of Contemporary Dutch). *International Journal of Lexicography* 29(2): 139-155.

Crowley, Terry. 2010. The Socially Responsible Lexicographer in Oceania. *Journal of Multilingual and Multicultural Development* 20(1): 1-12, doi: 10.1080/01434639908666366

Fuertes-Olivera, Pedro A. & Tarp, Sven. 2022. Critical Lexicography at Work: Reflections and Proposals for Eliminating Gender Bias in General Dictionaries of Spanish. *Lexikos* 32(2): 105-132. http://www.scielo.org.za/pdf/lexikos/v32nspe/05.pdf

Gouws, Rufus. 2020.: Metalexicographic models for multilingual online dictionaries in emerging e-societies. In Domínguez; Mirazo, & Valcárcel (eds.) *Studies on Multilingual Lexicography*. Berlin / Boston: de Gruyter. Lexicographica. Series Maior, vol. 157, 29-46.

Hu, Huilian; Xu, Hai, & Hao, Junjie. 2019. An SFL approach to gender ideology in the sentence examples in the Contemporary Chinese Dictionary. *Lingua* 220: 17-30.

Johnson, Harold L. 1971. *Business in contemporary society: Framework and issues*. Belmont, CA: Wadsworth.

Kouassi, Konan. 2022. Mensch-Maschine-Interaktion im lexikographischen Prozess zu lexikalischen Informationssystemen. In Klosa-Kückelhaus, Annette; Engelberg, Stefan; Möhrs, Christine, & Storjohann, Petra (eds.) *Dictionaries and Society. Proceedings of the XX EURALEX International Congress*. Mannheim: IDS-Verlag, 172-180. https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202022/EURALEX2022_Pr_p172-180_Kouassi.pdf

Müller-Spitzer, Carolin & Rüdiger, Jan Oliver. 2022. The influence of the corpus based on the representation of gender stereotypes in the dictionary. A case study for corpus-based dictionaries of German. In Klosa-Kückelhaus, Annette; Engelberg, Stefan; Möhrs, Christine, & Storjohann, Petra (eds.) *Dictionaries and Society. Proceedings of the XX EURALEX International Congress*. Mannheim: IDS-Verlag, 129-141. https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202022/EURALEX2022_Pr_p129-141_Mueller-Spitzer-Ruediger.pdf

Villa-Vigoni-Theses. 2018. https://www.emlex.phil.fau.eu/about-us/publications/other/

Wiegand, Herbert Ernst. 1983. Überlegungen zu einer Theorie der lexikographischen Sprachbeschreibung. In Hyldgaard-Jensen, Karl & Zettersten Arne (eds.) *Symposium zur Lexikographie. Symposium on Lexicography. Proceedings of the Symposium on Lexicography*. September 1-2, 1982, at the University of Copenhagen (GL 5-6/82). Hildesheim, 35-72.

Wiegand, Herbert Ernst; Gouws, Rufus; Kammerer, Matthias; Mann, Michael, & Wolski, Werner (eds). 2020. *Wörterbuch zur Lexikographie und Wörterbuchforschung*. Berlin / Boston: De Gruyter.

Williams, Geoffrey. In Praise of Lexicography, and Lexicographers. 17th EURALEX International Congress, Tinatin Margalitadze, Sep. 2016, Tbilisi, Georgia, 77-88. https://shs.hal.science/halshs-01502498/document