International Journal of Lexicography, 2024, **37**, 1–28
https://doi.org/10.1093/ijl/ecad023
Advance access publication 6 October 2023
Article

OXFORD

A Lexicographic Practice Map of Europe

Carole Tiberius, Jelena Kallas, Svetla Koeva, Margit Langemets, Iztok Kosem

Instituut voor de Nederlandse Taal, The Netherlands (carole.tiberius@ivdnt.org)
Institute of the Estonian Language, Estonia (jelena.kallas@eki.ee)
Institute for Bulgarian Language, Bulgarian Academy of Sciences, Bulgaria (svetla@dcl.bas.bg)
Institute of the Estonian Language, Estonia (margit.langemets@eki.ee)
Faculty of Arts, University of Ljubljana & Jožef Stefan Institute, Slovenia (iztok.kosem@ijs.si)

Abstract

This article presents the combined results of three international surveys that were carried out in the context of the Horizon 2020 European Lexicographic Infrastructure project (ELEXIS). The aim of these surveys was to gain more insight into lexicographic practices and the needs of lexicographers in Europe. The surveys were delivered via online platforms. Based on the combined results, we sketch a map of lexicographic practices in Europe, both for born-digital and retrodigitized resources, analyze current needs in terms of tools, functionalities and training, and identify emerging trends that will affect lexicography in the short and long term.

1 Introduction

Although elaborate efforts are put into the development of lexicographic resources in most European countries, cooperation on a larger European scale has long been limited. As a result, the European lexicographic landscape is quite diverse, with different languages having different lexicographic traditions and different levels of expertise and resources available. However, to be able to tackle the challenges of modern-day dictionary-making brought about by technological progress and the move from print to online, the need for cooperation has become even greater.

To enable more collaboration within the field of lexicography as well as with other fields and to bridge the gap between more advanced and less-resourced communities working on lexicographic resources in Europe, the Horizon 2020 ELEXIS project[1] (2018-2022) was set up to create a sustainable infrastructure for lexicography (Krek et al. 2018, 2019, Pedersen et al. 2018, Woldrich et al. 2020).

Within the project, three international surveys were conducted to get an overview of existing lexicographic practices across Europe, tools and methods used for compiling both born-digital and retrodigitized lexicographic resources, and the needs that lexicographers have now or anticipate to have in the short-term and long-term future. In this, the project built on previous work carried out within the COST action European Network of e-Lexicography (ENeL)[2] that helped gather a great deal of information on lexicographic practices and workflows across Europe (cf. Tiberius and Krek 2014, Krek et al. 2015, Tiberius et al. 2015). However, due to rapid changes in the field, an update and extension to this work was very much needed. Therefore, three surveys were carried out within the ELEXIS project to collect data from lexicographic institutions and lexicographers. In chronological order:

- Survey on Lexicographic practices: A Survey of Lexicographers' Needs (Kallas et al. 2019a; Kallas et al. 2019b) targeted at individual lexicographers. In total, 159 lexicographers from 45 countries (36 European and 9 outside Europe) completed the survey.
- Survey on Lexicographic practices: A Survey of Lexicographers' Needs for Lexicographic Partner Institutions (Kallas et al. 2019a) targeted at the 11 lexicographic partner institutions in ELEXIS.
- Survey on Lexicographic practices: A Survey of Lexicographers' Needs for Observer Institutions (Tiberius et al. 2022) targeted at the observer institutions[3] in ELEXIS. In total, the survey was completed by 54 institutions.

The first two surveys were conducted in 2018 (the first year of the ELEXIS project). The survey for observer institutions was carried out in the second half of the project. In this paper we report on the combined results of the three surveys to get a more general picture of the lexicographic community in Europe and to sketch a map of lexicographic practices in Europe. We will mainly focus on the data provided by institutions, but we will complement this data with findings from the survey for individual lexicographers where relevant.

The remainder of this paper is structured as follows. We begin with a brief introduction of the ELEXIS project and the project's organizational structure in Section 2. In Section 3, we describe the methodology of the surveys, setting out their general principles and aims, as well as the implementation. Section 4 presents a detailed analysis of the survey results, followed by a discussion in Section 5. We will end with a conclusion and plans for further research in Section 6.

2 The ELEXIS project

The main objective of ELEXIS was to create a sustainable infrastructure for lexicography to (1) enable efficient access to high-quality lexicographic data so that it can also be used by other fields, including Natural Language Processing (NLP), artificial intelligence (AI) and digital humanities, and (2) bridge the gap between more advanced and less-resourced scholarly communities working on lexicographic resources. To realize these goals, ELEXIS used an inclusive multi-layered organizational structure aimed at engaging different user groups with various levels of intensity during the project (see Figure 1).

The core of the project consisted of a consortium of 17 partners, including content-holding institutions and researchers with complementary backgrounds: lexicography, digital humanities, standardization, language technology, the Semantic Web, and AI. Out of the 17 consortium partners, 11 were defined as lexicographic partners in the ELEXIS grant agreement bringing quality lexicographic data and lexicographic expertise in the consortium. These were the Austrian Academy of Sciences, the Institute for Bulgarian Language "Prof Lyubomir Andreychin", the Society for Danish Language and Literature, the Institute of the Estonian Language, Trier University - Trier Center for Digital Humanities, the Hungarian Academy of Sciences - Research Institute for Linguistics, K Dictionaries Ltd, the Dutch Language Institute, the Belgrade Center for Digital Humanities, "Jožef Stefan" Institute, and the Real Academia Española. It is important to mention that the consortium included also two industrial partners, who were responsible for the industrial/commercial involvement in the project.

Another organizational layer was formed by observer institutions that were directly included in outreach and dissemination activities through various channels. The central group of institutions falling under the observer category were typically, but not exclusively, those producing high-quality lexicographic data and resources. At the end of the project, ELEXIS had 56 observers. Appendix 1 contains an overview of the number of observers and partners per country.

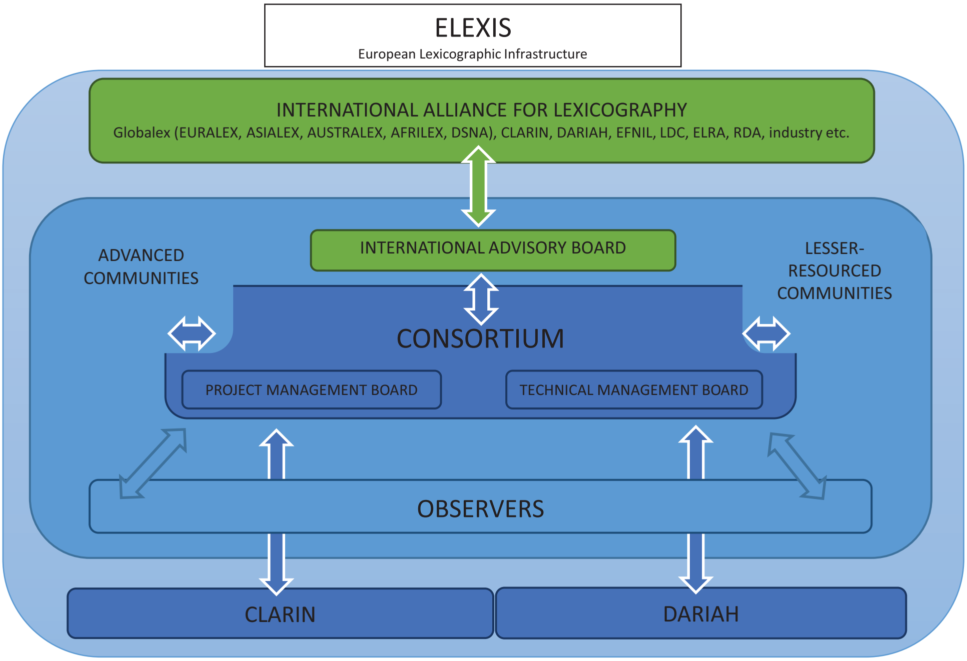

ELEXIS
European Lexicographic Infrastructure
INTERNATIONAL ALLIANCE FOR LEXICOGRAPHY
Globalex (EURALEX, ASIALEX, AUSTRALEX, AFRILEX, DSNA), CLARIN, DARIAH, EFNIL, LDC, ELRA, RDA, industry etc.
ADVANCED COMMUNITIES
INTERNATIONAL ADVISORY BOARD
LESSER-RESOURCED COMMUNITIES
CONSORTIUM
PROJECT MANAGEMENT BOARD
TECHNICAL MANAGEMENT BOARD
OBSERVERS
CLARIN
DARIAH

Figure 1: ELEXIS organizational structure.

The majority of countries represented in ELEXIS, either by a partner or an observer institution, were European countries including countries with close cultural ties to Europe and inclusive status in EU-funded initiatives. In addition, a few institutions from countries outside Europe were involved in the project as observers.

3 Methodology and practical aspects of survey design and implementation

At the start of the ELEXIS project, the idea was to carry out one European-wide survey focusing on workflows, metadata and data formats used in lexicographic projects within Europe. However, while preparing the first survey in 2018, it became clear that one survey could not cover all the aspects we were interested in. One of the problems was that certain types of questions, e.g. of a more technical nature, could not necessarily be answered by a lexicographer without the help of a computational linguist or an IT specialist. Another issue was the potential length of the survey; with all the questions included, the survey would be very long, which would likely put off potential respondents, or we would get many partially completed surveys. Therefore, it was decided to conduct two separate surveys, one targeted at institutions and one targeted at individual lexicographers. To get as many responses as possible from individual lexicographers (and not just the opinion of their institutions), the survey targeted at individual lexicographers was limited in length.

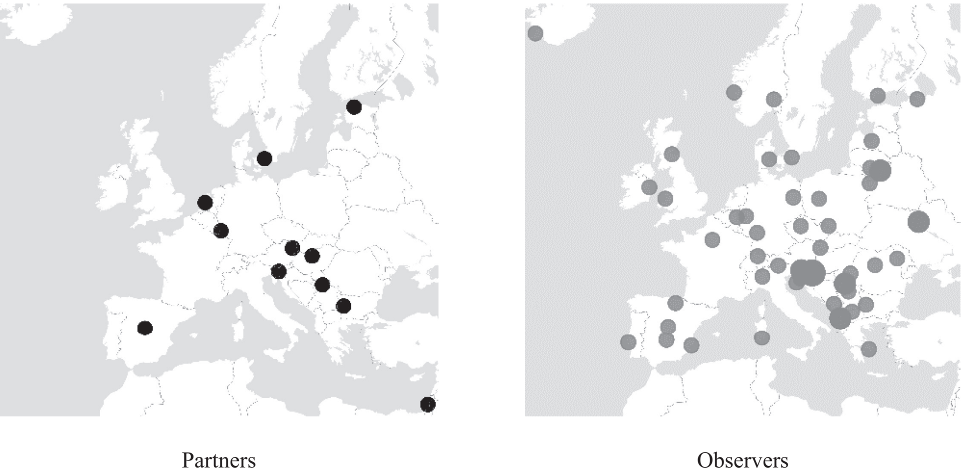
The survey targeted at institutions was initially only sent to the ELEXIS lexicographic partner institutions. However, as we were aiming to enable a comparison of lexicographic practices across Europe, it was important to get good coverage to ensure that the data would be representative of the lexicographic community in Europe as a whole and would not be biased. Therefore, a revised and upgraded version of the survey for institutions was conducted later in the project among the observer institutions. The intention for the surveys targeted at institutions was that one survey would be completed per institution and that it would be completed by a representative on behalf of the institution. We cannot, however, exclude that some personal opinions are reflected in some of the answers given. Moreover, in some of the open-ended questions, we did actually ask the respondents about their views. If this was the case, this was clearly marked in the question.

The method chosen for the surveys was an online questionnaire. Questionnaires had already proven to be a very effective and useful method for approaching the lexicographic community in the COST ENeL network (e.g. Tiberius and Krek 2014, Krek et al. 2015, Tiberius et al. 2015, Kosem et al. 2019). The first two surveys, conducted in 2018, were implemented in Google Forms, as it is simple to use and manage and seemed to cover the majority of our needs. However, the fact that Google Forms does not support the nesting of questions turned out to be a problem as it led to unexpected results in the analysis. Therefore, we decided to switch to more advanced survey software, 1ka[4], for the third survey.

All three surveys were divided into sections: (1) General information, (2) Ongoing work/projects, (3) Software and tools, (4) Publication, (5) Retrodigitization, (6) Past and future. The survey targeted at individual lexicographers contained 44 questions. The surveys for institutions were more elaborate and contained a separate section on data formats, metadata and availability, as well as some questions on crowdsourcing and gamification which were included in the section on Publication. Overall, the survey for the lexicographic partner institutions contained 86 questions. The survey for the observer institutions contained even more questions, i.e. 121, as in this survey some questions were split[5] and further refined (see, for instance, Section 4.3 on lexicographic expertise). A few additional questions were also included related to online lexicographic resources that the respondents use themselves, and Lexonomy, the Dictionary Writing System that was further developed within the project.[6]

All the surveys contained three different types of questions: (1) “yes/no" questions, (2) multiple choice questions, and (3) open-ended questions. Open-ended questions were included as we were not just interested in quantitative data, but also in qualitative data. It should also be noted that not all questions were mandatory.

Overall, we got a fairly good response to all three surveys. We received 159 responses from individual lexicographers from a total of 45 countries, comprising 36 European countries (140 respondents) and 9 countries outside Europe (19 respondents). For the institutional surveys, we achieved a 100% response rate from the 11 lexicographic partner institutions and a 96% response rate from the observers (54 out of 56), totalling 65 institutions from 35 different countries. Figure 2 shows the location of the institutions that took part in the surveys. The larger dots indicate that there is more than one institution in the same location. There were, for instance, three institutions from Zagreb among the observers.



Partners

Observers

Figure 2. Location of institutions participating in the surveys (N=65).

Based on their involvement in ELEXIS and in the surveys, we can assume that these 65 institutions are fairly representative of where in Europe lexicographic work is carried out.

4 Analysis of the surveys' results

In this section, we present the combined results of the institutional surveys. We sketch a map of lexicographic practices in Europe on the basis of the quantitative data gained from the surveys. This data includes information on the type of institution, type of funding, type of lexicographic expertise, tools and software used, publication medium and training of lexicographers. We will end with an analysis of the open questions where respondents were asked about their views on the past and the future of lexicography, more specifically which were the major changes they observed in the field in the past 10-15 years, and which wishes and needs they anticipated for the next 10-15 years.

4.1 Institutions doing lexicographic work in Europe

Looking at the type of institution (Figure 3), we establish that lexicographic work in Europe is mostly carried out at public institutions and universities. There were slightly more universities (26) than public institutions (23) among the responding observer institutions, whereas the lexicographic partner institutions were mostly public institutions. Two institutions, one observer and one partner, indicated that they were a mix of public and private (public-private partnership, PPP). Among the partner institutions, there was also one private non-commercial and one private/commercial organization. There were no private/commercial organizations among the observers.

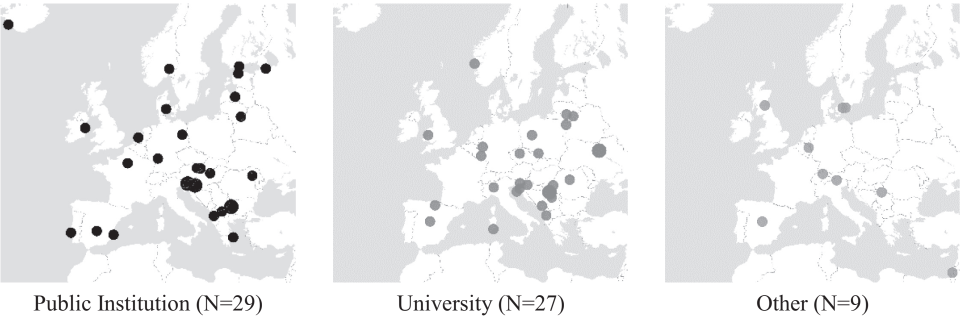


Figure 3. Overview of types of institutions that took part in the surveys (N=65).

Although Kosem et al. (2019: 96) note that commercial publishers dominate over public institutions in Greece, Germany, France, Israel, Italy, Portugal, and the UK, and that they still play an important role in Denmark, Ireland, and the Netherlands in the publication of monolingual dictionaries, commercial companies are clearly underrepresented in the ELEXIS lexicographic landscape.

One possible explanation is that research institutions in some countries do not receive adequate public funding, which drives their desire to participate in infrastructures for sharing lexicographic tools and resources rather than creating their own or purchasing off-the-shelf solutions, whereas commercial publishing houses have their own methods and tools and are less or not inclined to share them.

When we take a closer look at public institutions, we observe that there is only a small number of countries where this type of institution was not represented in our results. Data from Bosnia and Herzegovina, Malta and Luxembourg are missing as we did not obtain any results from these countries. Furthermore, public institutions from Belarus, the Czech Republic, Italy, Montenegro, Poland and Ukraine are missing, but they are represented in ELEXIS by universities. Sweden, Switzerland, and the UK are not represented by a public institution in ELEXIS either, but there are non-profit organizations conducting lexicographic work in these countries in the project. Some countries do not seem to have a public institution for language and are not a member of the European Federation of National Institutions for Language (EFNIL)[7] either (e.g. Bosnia and Herzegovina, Montenegro).

4.2 Funding used for lexicographic work

Most institutions taking part in the surveys rely on public funding for carrying out lexicographic work, sometimes in combination with private or other kinds of funding (see Figure 4). Only institutions from Belarus, the Czech Republic, Montenegro, Poland, and Sweden indicated that they rely solely on private funding or other kinds of funding for their lexicographic work. Other kinds of funding that were mentioned included fundraising as a charity organization and institutional calls for funding. Some respondents also indicated that no funding is provided.

The majority of the institutions that receive public funding for their lexicographic work receive it at the national level, i.e. 47 institutions (of which 29 are public institutions). Most of these are directly funded by the government, others rely on grants from national research agencies. When we consider public funding at the international level (reported by 23 institutions), we see that the number of universities and public

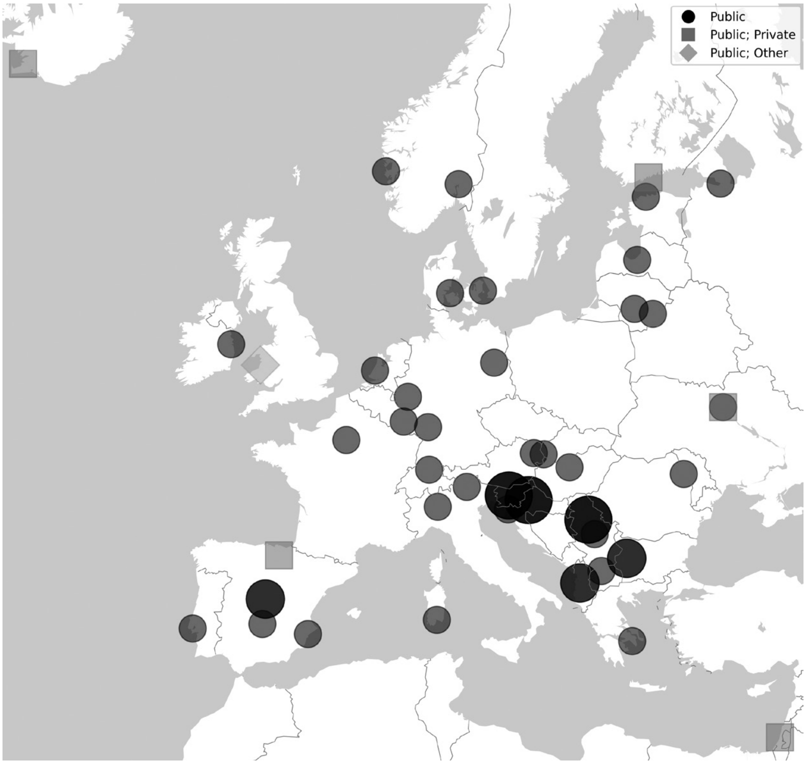


Figure 4. Institutions receiving public funding for their lexicographic work (N=53).

institutions that receive this type of funding is more or less equal. Furthermore, we note that universities rely more on other types of funding in addition to public funding than public institutions.

These observations suggest that lexicographic work in Europe is heavily dependent on national funding by the government. This corresponds with the findings of the European survey on dictionary use and culture (Kosem et al. 2019: 96), where it was reported that in the majority of the countries participating in the survey, monolingual dictionaries are published solely or mainly by public institutions funded by the government, especially in the case of countries/languages with a small number of native speakers.

4.3 Lexicographic expertise

In both institutional surveys, respondents were asked about the lexicographic expertise of their institution. However, the options offered were not completely the same in the two surveys as following new insights, these options were refined in the later survey. The partner institutions could choose from a) monolingual general dictionaries (modern, synchronic), b) monolingual specialized dictionaries (e.g. dictionary of collocations, phrasal verbs, synonyms, rhyming), c) historical dictionaries (e.g. diachronic, etymological, old literary languages), d) dialect dictionaries, e) bilingual or multilingual general dictionaries, f) multilingual terminological or specialized dictionaries (e.g. dictionary of legal terms, accounting), and g) other. Figure 5 shows the lexicographic expertise of the partner institutions. More than one answer could be selected.

We see that in this survey, the information on the type of dictionary and the number of languages involved was combined in the answers offered. For instance, expertise on terminological dictionaries was implicitly multilingual, whereas for specialized dictionaries the options monolingual and multilingual were offered separately. Whilst preparing the survey for the observers, we realized that this was not ideal and in order to get a clearer picture of lexicographic expertise, this information was split. Learner's dictionaries were also added as a separate category as it was not listed among the examples for specialized dictionaries in the survey for the lexicographic partner institutions.

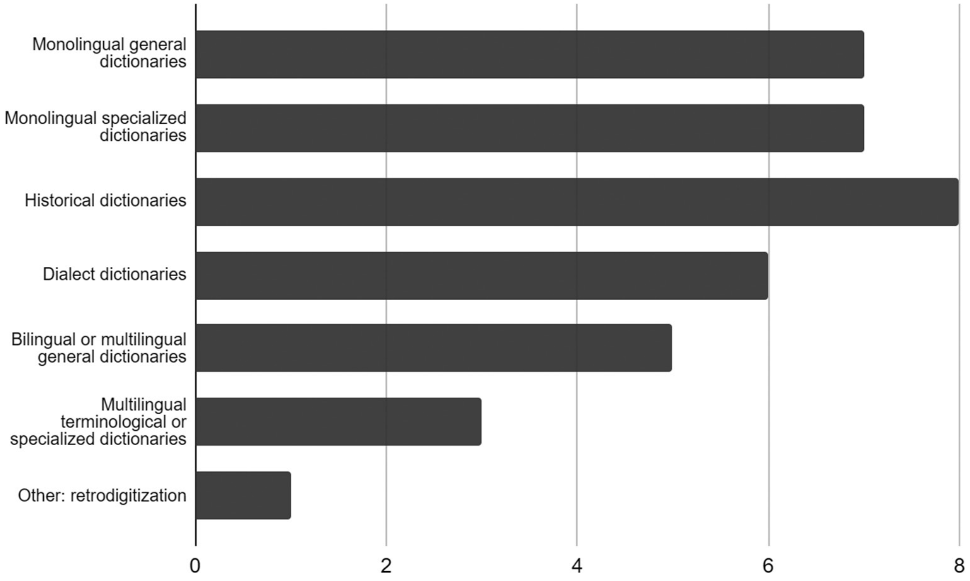


Figure 5. Lexicographic expertise at the partner institutions (N=11).

Figure 6 shows the results for the observer institutions. Again, multiple answers could be selected.

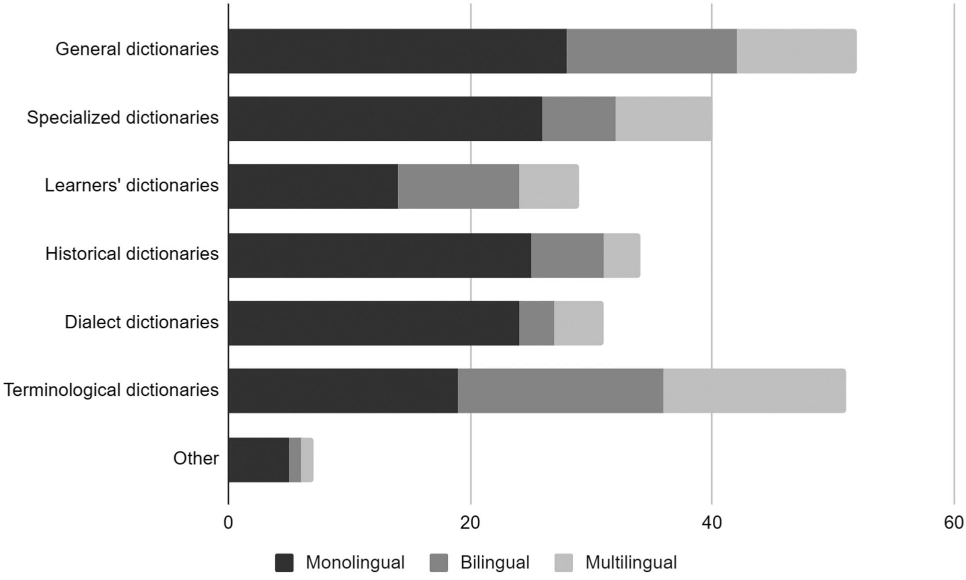


Figure 6. Lexicographic expertise at the observer institutions (N=54).

Figures 5 and 6 show that both the partner institutions and the observer institutions have varied lexicographic expertise, most reporting expertise in more than one type of dictionary.

The maps in Figure 7 show the distribution of expertise in general (monolingual, bilingual and multilingual) dictionaries per country.

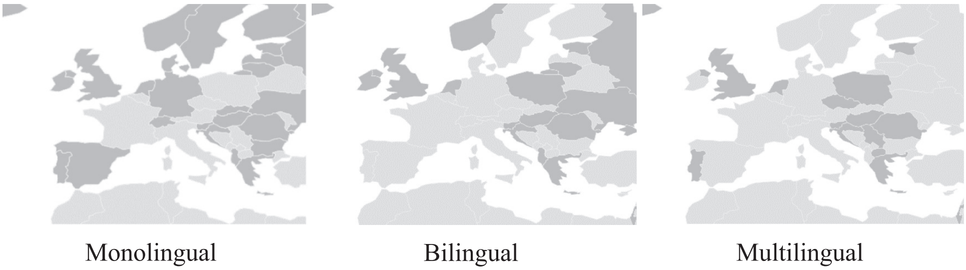
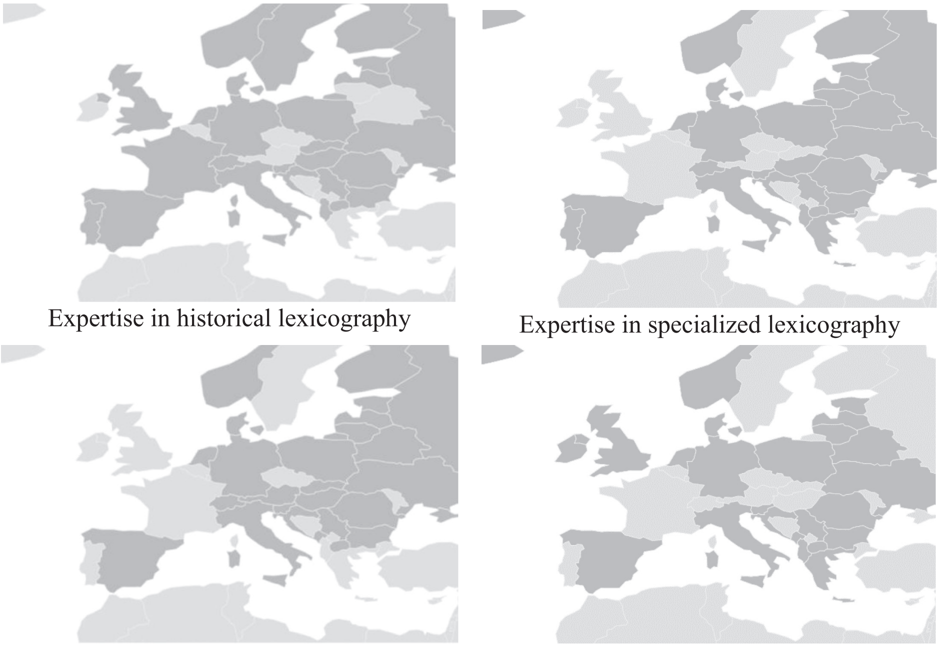


Figure 7. Countries in ELEXIS with expertise in general dictionaries.

From the countries represented within ELEXIS, expertise in general dictionaries was lacking in France, Belarus, Austria, and Italy, according to our data. This of course does not mean that such dictionaries do not exist in these countries. We already noted the dominance of commercial publishers in the publication of monolingual dictionaries in France and Italy (cf. Kosem et al. 2019), and that commercial publishing houses are underrepresented in ELEXIS. Austria is a special case, as the variant of German spoken in Austria is generally considered to be mutually intelligible with standard German. However, there are dictionaries that focus on the differences in vocabulary spoken in Austria and Germany. We do not have enough information about general monolingual, bilingual, or multilingual dictionaries in Belarus.



Expertise in historical lexicography

Expertise in specialized lexicography

Expertise in dialect lexicography

Expertise in terminology.

Figure 8. Countries in ELEXIS and type of lexicographic expertise..

Looking at the other types of lexicographic expertise present in the different countries in Europe taking part in the surveys, we get the following spread as shown in Figure 8.

Note that these maps should be seen as an approximation of the distribution of the different kinds of lexicographic expertise in Europe. For instance, we see that according to our data, there is no expertise in historical lexicography in Austria, but we know that the Austrian partner institute is involved in historical dialect dictionaries. Note also that we know that terminological expertise is present at the Portuguese partner institution, but they were listed as a standardization partner (not as a lexicographic partner) in the ELEXIS grant agreement and only lexicographic partners took part in the survey.

When we consider the distribution of monolingual versus bilingual/multilingual expertise and the type of institution, we observe that for all types of lexicographic expertise, monolingual expertise is more frequently mentioned by public institutions, whereas bilingual/multilingual expertise is more often reported by universities. The difference is largest for general monolingual dictionaries, with 24 public institutions reporting having this kind of expertise compared to 9 universities. Only for terminological dictionaries, we note that bilingual/multilingual expertise is mentioned almost as frequently by public institutions as by universities. This type of expertise is more prevalent among the observer institutions where expertise on terminological dictionaries is reported almost as frequently as expertise on general dictionaries.

4.4 Retrodigitization

As we were not only interested in born-digital lexicography, the surveys also included a separate section on retrodigitization in order a) to reveal the institutions' involvement in the various stages of the retrodigitization process (i.e. the process of converting a paper-based dictionary into a digital, computer-readable format, including scanning, OCRing, data encoding and enrichment); b) to provide an overview of the software used in this process, and c) to provide an insight into opinions on which dictionaries should be retrodigitized. In general, the lexicographic community – including ELEXIS partners and observers – is interested in retrodigitizing printed dictionaries. The respondents value the data described in printed dictionaries as they pointed out many dictionaries (mainly historical, dialect and specialized) as a possible target for retrodigitization. Looking at the percentages, partners are more often involved in retrodigitization (64%: 7 out of 11) than observers (46%: 25 out of 54 or 25 out of 50 (50%) as 4 did not answer). The dots in the map of Figure 9 show the locations of institutions that are or have been involved in retrodigitization.

Overall, 21 public institutions and 9 universities reported working on retrodigitization, while 12 public institutions and 15 universities did not (see Table 1). Of the 4 private and mixed organizations in the project, 2 also reported retrodigitizing activities. This suggests that retrodigitization occurs more frequently in specialized lexicographic centers than in universities. There were 4 institutions that did not answer the question on whether they are or have been involved in retrodigitization, i.e. one institution from Latvia, one from Serbia, and two from Croatia.

For retrodigitization, dialectal, historical, and onomastic dictionaries are particularly appealing, and there is a correlation between lexicographic expertise mentioned by the institutions and their involvement in retrodigitization: 30 out of the 32 institutions that are involved in retrodigitization indicated having expertise in dialect, historical and/or specialized dictionaries. Furthermore, in retrodigitization, emphasis is placed on

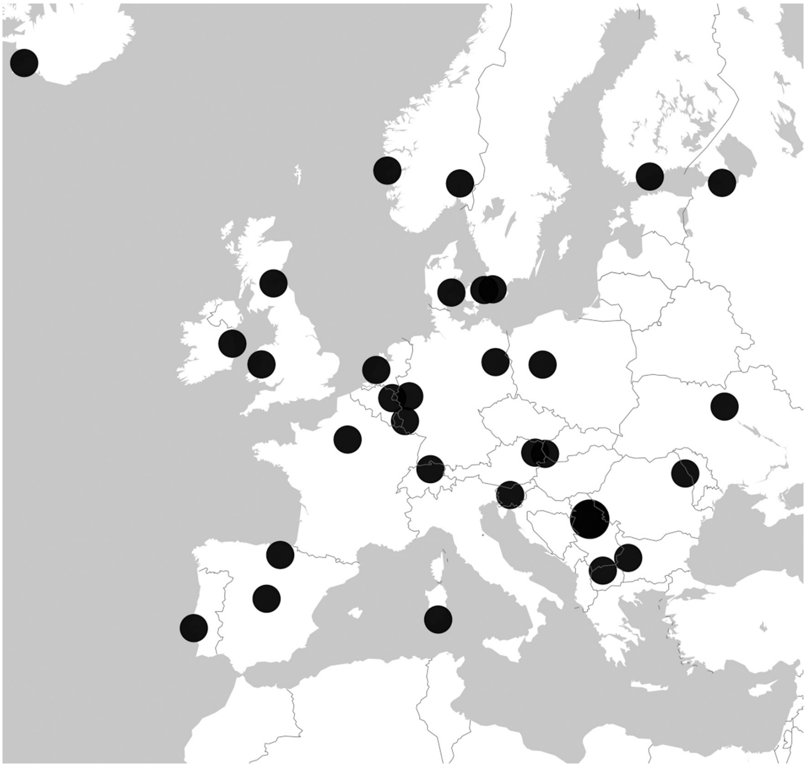


Figure 9. Institutions that are or have been involved in retrodigitization (N=32).

Table 1. Retrodigitization and type of institutions (N=61)

	Yes	No
Public/Non-Profit	21	12
University	9	15
Private commercial	0	1
Private non-commercial	1	0
Mixture	1	1

multi-volume dictionaries with broad vocabulary that were published in the second half of the twentieth century.

Similar procedures and software tools were mentioned in both surveys for the various stages of retrodigitization (image capture, text capture, data encoding and data enrichment). This is reassuring, as it implies that some best practices for the retrodigitization workflow are already in place.

The content of the retrodigitized dictionaries has to be structured in order to be suitable for online access, to be easily upgraded and expanded, or to be linked to the (usually complex) structure of born-digital dictionaries. Thus, a conversion from plain text to structured text (for example, XML) should be performed to achieve an explicit structure comparable to the structure of born-digital dictionaries. XML is the preferred format, and XML-based tools and editors are the preferred technologies according to the respondents.

The added value of retrodigitized dictionaries may come in two forms: first, as a source for online references, and second, as the foundation for creating new dictionaries. Institutions provide access to their retrodigitized materials in various ways: 20 institutions provide access via an institutional portal or website, 4 via an API, 6 via the download of image files, and one via the download of full text; 10 institutions stated that they make the full text of their retrodigitized dictionaries available to users. The most common justifications for not sharing retrodigitized materials are copyright restrictions and ongoing work.

4.5 Tools supporting lexicographic work: DWS and CQS

Nowadays, two tools are more or less indispensable in a good lexicographic workflow, i.e. a dictionary writing system (DWS) and a corpus query system (CQS).

In the surveys, a DWS was defined as a piece of software for writing and producing a dictionary. It might include an editor, a database, a Web interface and various management tools – for allocating work, etc. Specialized dictionary editing software includes customizations of existing/standard (XML) editors.

A CQS was defined as a piece of software that lets the lexicographer/researcher see the concordances for any word, phrase or grammatical construction in a text corpus, an extensive collection of texts, usually annotated with additional information about words such as their base form or lemma, part-of-speech category and similar.

The survey results show that the use of a DWS in combination with a CQS (dots) is most common for the 44 institutions currently working on lexicographic projects (Figure 10) and can be found at institutions located all over Europe.

The use of CQS alone (squares) in the traditional work of creating dictionaries can be viewed as a transition to the simultaneous use of DWS and CQS, and thus it is relatively widely spread compared to the independent use of DWS (diamonds), which is not very common. There are still six institutions that reported not using a DWS or a CQS at all to support their lexicographic work. Most of those indicated though that they feel the need at least for a DWS to support their work.

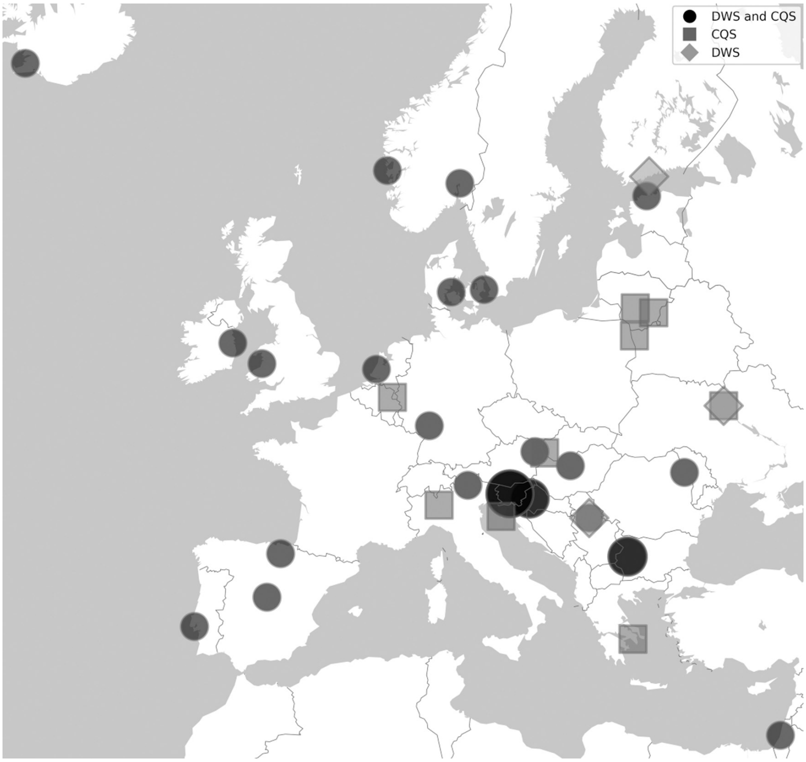


Figure 10. Use of DWS and CQS to support lexicographic work in institutions with at least one ongoing lexicographic project[8] (N=44).

If we take a closer look at the type of institution, we observe that at public/non-profit institutions, the use of a DWS in combination with a CQS is much more common than at universities, where the use of a CQS alone is as common as using both tools.[9] Organizations which are a mixture of public and private as well as private institutions follow the pattern of the public/non-profit institutions here. They all use a combination of a DWS and a CQS for their lexicographic work.

These findings are similar to the results from the survey for individual lexicographers. Of the 89 lexicographers who answered the question on whether they use tools to support their lexicographic work, just over half reported that they use both a DWS and a CQS. However, among the individual lexicographers, the use of a DWS (15.9%) without a CQS was more common than the use of a CQS alone (10.2%), which is the exact opposite of the institutional result. These results show that the community is technologically still quite heterogeneous.

The reasons mentioned mostly by observers and individual lexicographers for not using a DWS were financial difficulties in purchasing lexicographic software or tools, as well as the absence of suitable knowledge and technical skills. Indeed, choosing the most appropriate tool to use may not be straightforward as many different DWSs and CQSs were reported. See Kallas et al. (2019a) for an overview of the different systems that were mentioned by the individual lexicographers and the partner institutions. The respondents in the survey targeted at individual lexicographers mentioned 15 DWSs and 22 CQSs. The ELEXIS lexicographic partners mentioned 11 DWSs and 8 CQSs, and the observers named 26 DWSs and 31 CQSs. Of the various systems mentioned, the Sketch Engine[10] is the most mentioned CQS and Lexonomy[11] the most mentioned DWS.[12] For DWS, in-house solutions are still very common (which is in line with earlier results, cf. Tiberius and Krek 2014), whereas for CQS, commercial systems tend to be used most. It thus seems that at the time the surveys were conducted, the landscape did not change significantly compared to 2014 and existing off-the-shelf DWSs still do not meet (all) the needs of lexicographic projects. Furthermore, the integration of a DWS and a CQS into one tool has not yet become common practice in modern lexicography, although institutions feel that this would be beneficial, especially for the linking, selection and retrieval of examples and collocations.

From the results of the surveys, we can conclude that overall the partner institutions and observers are satisfied with the CQS they use. The most frequently mentioned wishes for further development of CQSs included advanced corpus creation and annotation tools (including spoken and monitor corpora); better metadata management; additional functionalities (e.g. diachronic analysis; detection of translation equivalents; (bilingual) term extraction); improved user ergonomy and customization of the user interface according to user profile, for instance, CQS for learners. Individual lexicographers expressed more specific wishes depending on the type of their projects, such as support for corpus annotation, including tagging mistakes on the fly; better support for data evaluation; better access to certain types of texts (e.g. transcriptions); possibility to present legally sensitive data; better support for data acquisition (e.g. multi-level extraction, (syntactic) pattern detection). All three surveys revealed the need for more advanced tools for semantic analysis, including enhanced sense annotation and disambiguation, sense clustering and embeddings.

As for DWSs, most partner institutions and observer institutions are quite satisfied with the DWS they use at the moment. How satisfied they are with a DWS seems to depend on factors such as the availability of support; available functionalities; the possibility to adapt and add functionalities; the ability to work with multiple users and real-time updating of the database. Observer institutions also expressed concerns about the long-term sustainability of the system and about keeping up with technical improvements. In relation to DWS, the following important features were mentioned: better support for (automatic) data collection (simple import and export of files, mapping transcripts, the inclusion of media files (e.g. audio files with a linked transcript); better support for data management and data processing (e.g. version history, assignment tools, change tracking, statistics, complex searching, advanced visualization options, automatic validation tools, internal and external reference facilities); support for data publishing (e.g. print and export functions); tools for processing user-generated content. Customization needs of DWSs concern mostly schemas, DTDs and menus, search options, and export options (including export for saving and transformation (e.g. XML, CSV, JSON, TEI), for printing (e.g. pdf, Indesign), and for publishing online).

Respondents in all three surveys also raised the issue of data formats and expressed the need for (more) stable and established formats for data encoding in lexicography. Although a shift can be observed from non-structured data to structured data, quite a few institutions (44% of the observer institutions and 36% of the lexicographic partner institutions) reported using a non-structured data format (e.g. Microsoft Word) for at least some of their projects. Furthermore, it should be noted that using a structured format does not automatically enforce that each dictionary is encoded in an exactly specified manner. Within TEI, this situation inspired the development of TEI Lex-0, a more constrained version of TEI which aims to establish a baseline encoding and a target format to facilitate the interoperability of heterogeneously encoded lexical resources (cf. Romary and Tasovac 2018). In addition, the use of a special metadata schema (e.g. CMDI) and the use of a standard licensing schema (e.g. Creative Commons) are not yet widespread among lexicographic institutions. All this makes it harder to share data across different projects and applications.

It also hinders linking individual lexicographic resources to other (lexicographic and NLP) resources, which forms a significant obstacle for reusing the data in other fields. An additional requirement which came to the fore in the survey for observers is the need for API access for lexicographic tools and resources.

4.6 Publication medium of lexicographic data

Figure 11 shows the medium that institutions use to publish their (ongoing/new) lexicographic resources. Note that we restricted the results here to those institutions that have at least one ongoing lexicographic project and that we focus on the current situation, that is, the publication of ongoing and new lexicographic projects.

The dots show that at the moment the online medium is by far the most popular publication medium for lexicographic projects. Sometimes both options – online and print – are offered (squares), and a minority of institutions (Belarus, North Macedonia, and Russia) indicated publication in print only (diamonds). Note though that the print-only option was selected for 24 out of the 124 projects that were mentioned in the survey for the individual lexicographers (primarily by respondents from eastern and southeastern Europe). The main reason for publishing in print is tradition. The dictionary is part of a larger, long-term project and previous volumes have also appeared in print. Lack of technical support or software and user demand are also mentioned as motivations for publishing in print. Overall, these results are in line with what was reported by Kosem et al. (2019: 109-111) on the status of lexicography (types of dictionaries being compiled and their format) in the 26 countries involved in their study.

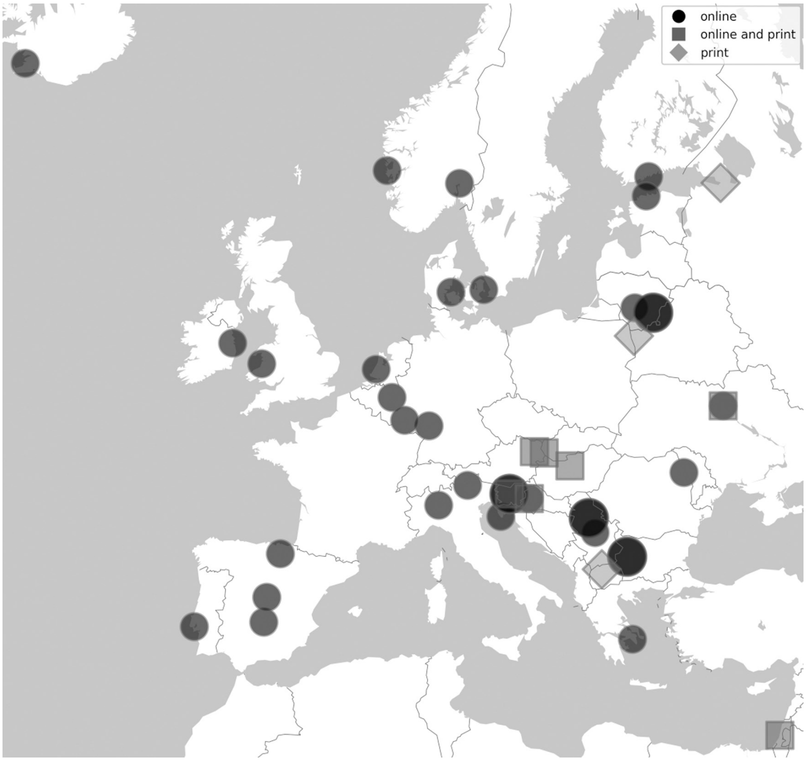


Figure 11. Publication medium at institutions with at least one ongoing lexicographic project (N=44).

Most dictionaries that are published online are made available for free to users through an interface. However, access to the actual underlying data for reuse by others is often still restricted. From the partner institutions only two mentioned that their data is accessible via an API for free, whereas all others required a license sometimes in combination with a fee. Among the observer institutions, the option 'Free download and use under certain license' was selected most often, followed by 'Customized preparation of datasets' and 'Free API access'. Common licensing schemes that are used, are Creative Commons and the CLARIN licensing framework with the Creative Commons' licenses being slightly more used by the observers. More details on licensing issues can be found in Kosem et al. (2021) which reports on the results from a separate survey conducted within ELEXIS on licensing of lexicographic data.

4.7 Automatic Knowledge Extraction

The automatic extraction of data and knowledge is finding its way into lexicography and we observe that automatic data extraction takes place at institutions located all over Europe. Figure 12 shows the percentage of partner and observer institutions that reported using some kind of automatic data extraction.

Interestingly, the results are very similar to the results from the survey which was conducted in 2015 in the context of the COST ENeL network (Tiberius et al. 2015). At the top of the list, we still find the extraction of frequency information and headword lists, whereas definitions and knowledge-rich contexts[13] are still at the bottom. Looking in detail at the results from our surveys, we see that the extraction of headword lists is most common for the partner institutions, whereas the extraction of frequency information is most common for the observer institutions. Observer institutions are also more involved in the extraction of form variation, sense information and neologisms, and they reported extracting more different types of data. For instance, the extraction of audio data, knowledge-rich contexts, register information and diachronic sense distributions is currently only mentioned by the observer institutions. We should note here that the survey for observers was conducted two years after the survey for the partner institutions and

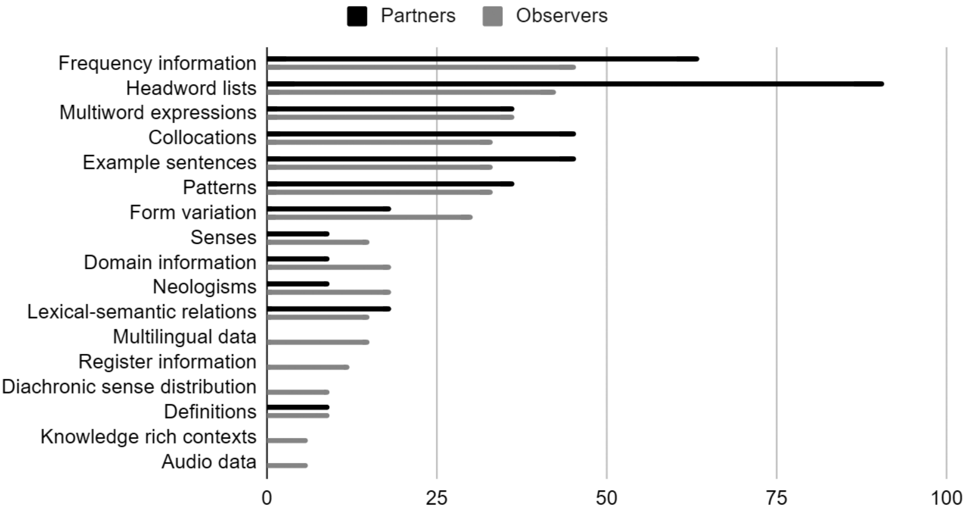


Figure 12. Types of data extracted by lexicographic partners and observers with an ongoing lexicographic project (N=44).

that the differences may be partly explained by technological advances in automatic data extraction in general.

For instance, the Sketch Engine has added new functionalities between 2018 and 2020, such as filtering on text type in Word Sketch and trends[14], a feature for detecting words which undergo changes in the frequency of use over time (diachronic analysis).

4.8 Crowdsourcing and gamification in modern lexicography

Another emerging trend is crowdsourcing and gamification (cf. Čibej et al. 2015). The results show though that crowdsourcing and gamification are not yet common practice in the dictionary compilation process. Only 11 institutions (of which 4 partners and 7 observers) reported involvement in crowdsourcing. Projects that were mentioned include the Sprachatlas der deutschen Schweiz (SDS) – Steno-Labor[15], the Thesaurus of Modern Slovene (Arhar Holdt et al. 2018), the Collocations Dictionary of Modern Slovene (Kosem et al. 2018), and the Taalradar project[16] which hosts various lexicographic crowdsourcing experiments for Dutch.

Even fewer institutions (only 3) are or have been involved in gamification. The projects that were mentioned were online educational language games of the Institute of the Lithuanian Language[17] and the Slovene Game of Words "Igra besed"[18], a mobile application purposed for a gamified improvement of two automatically compiled dictionaries for Slovene: the Collocations Dictionary of Modern Slovene and the Thesaurus of Modern Slovene. This latter game inspired the development of the word games app in ELEXIS, which focuses on word combinations and has been made available for five languages (i.e. Dutch, English, Estonian, Portuguese, and Slovene)[19]. The game can be used to clean (semi)-automatically extracted collocational data from corpora. Within ELEXIS a second app was developed, i.e. CrossTheWord[20], an innovative crossword game in which puzzle clues are generated starting from definitions retrieved from WordNet[21], and players have to find the corresponding word. The CrossTheWord game leverages crowdsourcing for the purpose of removing noise within semi-automatically created lexico-semantic resources, in this case WordNet.

A possible explanation for the low figures for crowdsourcing and gamification may be that institutions are still searching for the best ways of including crowdsourcing methodologies in the lexicographic workflow. Potential issues could be the lack of suitable case studies, the lack of relevant features in existing DWSs, or the lack of tools supporting these methods. The need for such kinds of tools was expressed explicitly in the surveys. Another issue could be the scalability of the crowd (cf. Jakubíček 2022). How can we motivate large enough crowds to take part in crowdsourcing and gamification for lexicography?

4.9 Training of lexicographers

Over the past few decades, the field of lexicography has undergone some radical changes and has become far more interdisciplinary. As noted, for instance, by Leroyer and Kohler Simonsen (2020: 184) "the digital revolution … is leading to metamorphoses not only in the dictionary making processes and dictionary forms, but also in dictionary use and the general status of lexicography".

Indeed, the findings of our surveys indicate that nowadays a lexicographer's job is far from monotonous. Most lexicographers do not just edit dictionary entries anymore, they are also involved in various other aspects of dictionary-making, such as project management, promotional activities, responding to user questions and feedback, as well as communication with computational staff, which requires research and analytical skills, including the understanding of corpus and computational linguistics as well as good command of foreign languages. Lexicographers at the observer institutions tend to spend more than 50% of their time on other tasks. The partner institutions reported that their lexicographers mainly work on lexicographic projects (especially in the case of third-party funded projects), but not exclusively. However, we do not know how much of their time lexicographers at the partner institutions spend on other tasks as this additional question was added later only to the survey for the observer institutions.

These changes in the role of a lexicographer and the tasks that they do have implications for their training and education. However, our results show that training lexicographers does not seem self-evident. More institutions answered that they do not offer any form of training (e.g. in-house tutoring, external courses, workshops or summer schools) to their lexicographers than that they do (35 versus 29). The dots in Figure 13 indicate that training is offered and the squares indicate that no training is offered. One observer institution did not answer this question.

If we confine ourselves to the institutions with at least one ongoing lexicographic project (N=44), the situation becomes more positive and more institutions (26) offer training than not (18). We further note that it is more common for the partner institutions to provide training than for the observers, which may be related to the fact that lexicographic training is more often provided by public institutions than by universities.

Out of 11 partner institutions, only one indicated not providing training, whereas 17 of the 33 observer institutions with lexicographic projects do not offer training (10 public/

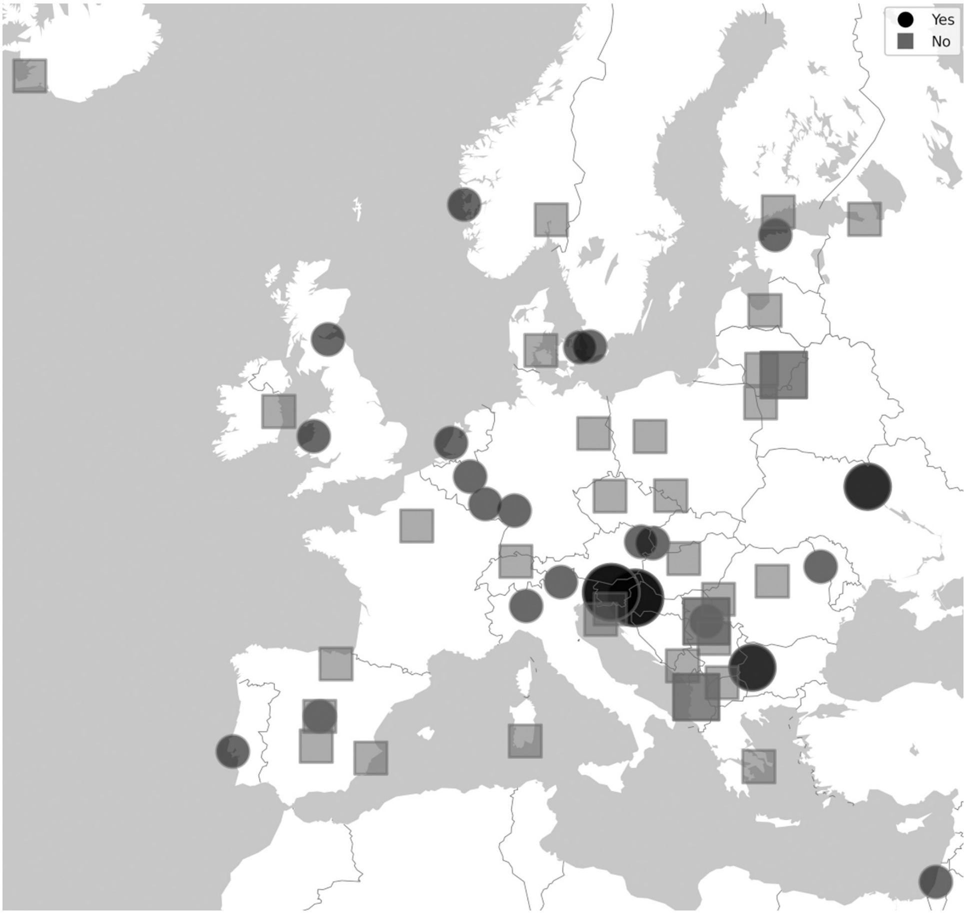


Figure 13. Institutions offering training to their lexicographers (N=64).

non-profit institutions and 7 universities) and 16 do (6 universities, 9 public/non-profit institutions and one mixture of public and private).

When training is offered, in-house training is most common, especially for training/mentoring new recruits, followed by attending special workshops and training schools. Formal education programmes are hardly mentioned, and although no such information was gathered, it is possible to conclude that universities in Europe usually do not offer BA or MA degree programs in lexicography. This emphasizes the importance of international degree programmes such as EMLex (European Master in Lexicography)[22] for training lexicographers.

The results from the institutions are confirmed by the data from the survey for the individual lexicographers, where more than one third of the respondents reported having been trained within their own institute, usually by a tutor or a senior lexicographer.

These findings on training and the changing job of a lexicographer are also in line with the profile of the respondents who completed the surveys on behalf of their institution. The respondents were primarily corpus linguists, computational lexicographers, or computational linguists holding a PhD in language or linguistics.

In sum, the needs of a modern lexicographer extend beyond linguistic knowledge, meaning that continuous training and development in various areas should become a regular part of a lexicographer's job, and institutions have to be prepared for extra costs for training and need to plan projects accordingly.

4.10 Changes observed over recent decades and looking ahead

In this section, we present a joint analysis of the answers to the open-ended questions, where respondents were asked to foresee major changes in lexicographic work during the next 10-15 years and to indicate the major changes that took place during the last 10-15 years.

The respondents from the institutions envisage a 'new lexicography' using not-yet-known technologies but still representing the art and craft of explaining the world around us.

For the future, a majority of respondents of all three surveys would like to see increased linking, sharing and reuse of resources, more open-source programs and platforms, as well as training on how to use them. Among future trends, the use of APIs, aggregated search, responsive design and crowdsourcing were mentioned repeatedly. From the thoughts about the near-past expressed by the respondents from the institutions, two crucial shifts can be concluded: (a) the radical move from paper to online publishing, as well as free online access to the dictionaries; (b) working with a lot of (large) corpora, as well as using new methods and tools for corpus analysis. As a consequence, corpus-driven lexicographic treatment of data should result in a better representation of linguistic phenomena (e.g. semantic change, specialized use in different domains, sense and sentiment annotation).

While respondents from institutions were more focused on general trends, the individual lexicographers focused more on the needs of their current projects. For individual lexicographers, the most relevant topics were a) the need for better tools for the extraction and automatic processing of data from corpora (with a focus on more semantically-based analysis, e.g. senses, definitions, and lexical relations); b) new methods for corpus creation; c) better integration of dictionaries and corpora, along with better integration of CQS and DWS; d) (semi-)automatic compilation and the advent of post-editing lexicography. Also, individual lexicographers were more concerned about the presentation modes of the lexicographic content, including online publishing and mobile applications. Several respondents emphasized that the overestimated value of dictionaries presented on smartphones may result in a neglect of the quality and reliability of lexicographic data. But generally, it was noted that the impact of mobile phones is immense as a distribution method and a mobile-first approach has to be adopted. As a positive change, individual lexicographers have also pointed out better interaction with end users, since users can now directly contact lexicographers online about words they are looking for, technical issues, etc. In turn, these changes influence the nature of the lexicographer's job and result in a shift in skills: the task is changing from creating a dictionary to maintaining and expanding a dictionary; besides lexicographers will be more and more expected to participate in project management, data management, fundraising, and public relations.

Considering the obstacles that were mentioned, one of the biggest concerns for all groups of respondents seems to be funding. The need for funding is expressed in all the ELEXIS surveys and in all parts of Europe, although it seems more urgent in Eastern Europe, where the respondents speak of a 'lack' of funding, whereas in Western Europe 'difficulties' in obtaining funding is used. In addition, concerns are expressed about the low status of lexicographic work, which is a constant worry for individual lexicographers and institutions.[23] Serious concerns were often expressed in connection with the quality and reliability of (semi-)automatically built resources and presenting (semi-)automatic results without proper linguistic and lexicographic expertise, while high-quality lexicographic data is still kept closed under restrictive licenses (both, public institutions and private publishing houses). Other concerns mentioned were the multiplicity of encoding schemas/grammars available, information overload, rapid technology development, the potentially reduced value of lexicographic skills in digitally oriented projects, and last but not least, the old-fashioned authoritarian tradition of the 'wrong' and the 'right' language use still dominating in some parts of the lexicographic community. One of the challenges is also how to constantly update all kinds of data.

Overall, we can conclude that these two open-ended questions revealed clearly that the lexicographic community is very heterogeneous; some issues that are favourably mentioned by some lexicographers were considered as negative by others, for example, moving from paper to online would not be good as "paper is more durable than web", also there were individual lexicographers who reported just moving from typewriting or handwriting to using the computer as the major change during the past 10-15 years.

4.11 Lexicographers' perception of a good lexicographic resource

Although our research was not a study into dictionary use, we were interested to find out what people working in lexicography value most when consulting online lexicographic resources themselves. Therefore, an additional open-ended question was added to the survey addressed at the observer institutions. We asked the respondents whether they use online lexicographic resources themselves (e.g. to look up the meaning, translation, spelling, or use of a word when reading or writing a paper) and what they like about those resources. The word cloud below gives an impression of the words that were most frequently mentioned in the answers given. Note that this was an open-ended question and that we manually coded the answers removing function words, summarizing very detailed answers and harmonizing the spelling.

If we look at Figure 14, we see that words such as *access, easy, fast, quick, free, online, updated, reliable, accurate, comprehensiveness* and *coverage* are frequently mentioned. Although we did not offer a list of criteria to choose from (cf. Figure 15), there are striking similarities between the words in the answers given to this open-ended question and the criteria included in the studies on dictionary use from Müller-Spitzer and Koplening (2014) and Kosem et al. (2019).

According to both these studies, dictionary users expect a good dictionary to be reliable, up-to-date, easy-to-use, and freely accessible. As can be seen in the word cloud, these qualities are not just appreciated by dictionary users, but also by lexicographers using online dictionaries. Similarly, media-specific features such as adaptive ways of presenting dictionary content or integrating multimedia features like audio files were ranked as less important in the user studies, and words relating to these features are absent from the word cloud.

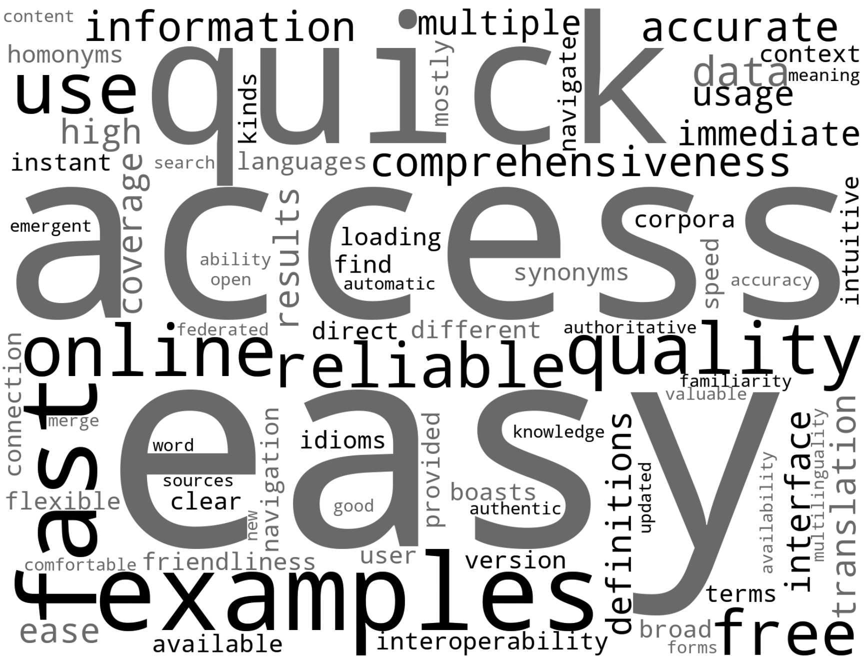


Figure 14. Word cloud illustrating what lexicographers value about lexicographic resources.

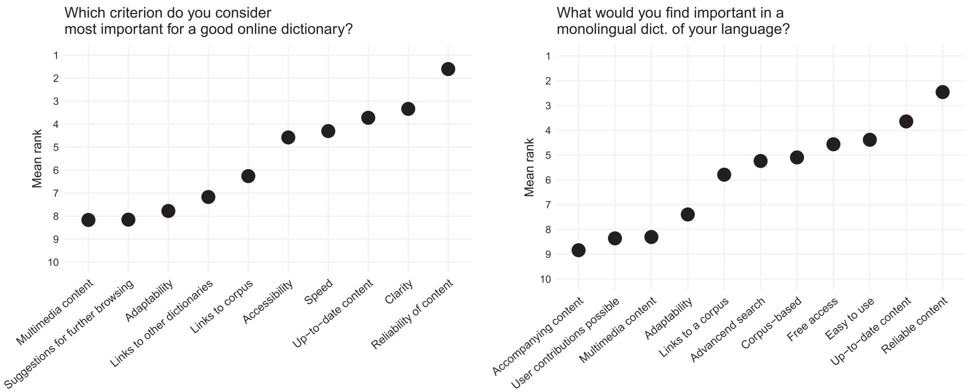


Figure 15. Mean ranks from the Müller-Spitzer and Koplenig (2014) study (left) and mean ranks from Kosem et al. (2019) (right). The data presented in the left panel was collected in 2010, while the data in the right panel comes from 2017.

The word cloud also gives a hint of the type of lexicographic information that our respondents tend to use as it contains words such as definitions, examples, idioms, translations and synonyms/homonyms. This information was not explicitly present in the studies from Müller-Spitzer and Koplenig (2014) and Kosem et al. (2019) but can possibly be linked to the criteria ‘corpus-based’ and ‘links to a corpus’. These features seemed to be of moderate importance to the non-lexicographic user, but their presence in the word cloud suggests that these are important features in a dictionary for people working in lexicography.

5 Discussion and implications for lexicography in relation to the main outcomes of ELEXIS

In the foregoing, we have presented an analysis of the results of the ELEXIS surveys on lexicographic practices and lexicographers' needs. In this section, we will focus on some key findings and consider these in relation to the relevant outcomes of ELEXIS. We will also discuss the implications of these in the light of future directions in lexicography.

One of the main objectives of ELEXIS was to promote an open access culture in lexicography enabling efficient access to high-quality lexicographic data so that it can also be used in other fields. The surveys show that the process of making lexicographic resources more openly available has started in the lexicographic community, but that there is still much to do. Although most partners and observers make their dictionaries available in an online application for free, access to the underlying data for reuse by others is often still limited and subject to restrictive licenses. Only a few institutions indicated that their data is available for free without any restrictions. Within ELEXIS serious efforts have been dedicated to address Intellectual Property Rights (IPR) issues and a number of flexible and diverse licensing options have been identified (Boelhouwer et al. 2020) to encourage the sharing of lexicographic data. In total, 106 lexicographic resources have been contributed to the ELEXIS infrastructure during the project duration. 60 resources have been contributed by the ELEXIS partner institutions, and 46 by the observer institutions. The majority of the data contributed by observer institutions are made available under an open license (76%), whereas the data from the partner institutions mostly comes with a restrictive license (65%). The data that is available under an open-access license has been integrated in the ELEXIS Dictionary Matrix[24], a universal repository of linked lexicographic data.

Another important outcome is that the surveys clearly show that the lexicographer's job has changed and that continuous training and development are required to keep up with new opportunities offered by technological advances and the digital medium. To counter the lack of university curricula for lexicography and of systemic training opportunities outside the university, ELEXIS has developed a special curriculum that provides a broad coverage of topics relevant to lexicographic practice in the computer age (cf. Tasovac et al. 2020, 2022). This newly developed ELEXIS curriculum[25] provides a good starting point for learning about lexicography. It is freely available and offers a wide range of courses covering the basic skills needed to educate a new generation of lexicographers so that they will understand the full potential of digital research infrastructures; will be able to optimally exploit existing state-of-the-art tools to create open, standards-compliant lexical datasets that can be fed back into the infrastructure, shared and reused. The curriculum is available on the online teaching platform DARIAH-Campus, which is maintained by DARIAH[26] and offers a stable framework for sustaining the ELEXIS learning resources. While this guarantees access to the training materials beyond the end of ELEXIS as a funded Horizon 2020 project, separate efforts will have to be made to keep the materials up to date when new technologies emerge. It remains to be investigated how this can best be realized, but there seems to be a trend towards developing more modular online course material (cf. e.g. UPSKILLS[27], which aims to upgrade the digital skills of linguistics and language students). Also services such as Elexifinder[28], offering access to lexicographic scientific output, are important in facilitating knowledge exchange in the lexicographic community.

The results from the sections on tools supporting lexicographic work and automatic knowledge acquisition show that there are still gains to be made. Nowadays, lexicographic projects throughout Europe typically incorporate a DWS and/or a CQS into their workflow, but there are still some institutions that do not use a DWS or a CQS at all for a variety of reasons including finances. The availability of open-source tools is therefore particularly important for institutions that have limited funding for their lexicographic projects. Within ELEXIS, a set of open-source services and tools have been developed (i.e. Elexifier[29] for the conversion of lexicographic resources to a uniform data format, Lexonomy for editing, linking, enriching and publishing lexicographic data, and Publex[30] for the online publishing of (retrodigitized) XML dictionary data).

These tools support the whole lexicographic workflow and exhibit a high level of interoperability, which is also a requirement expressed by many respondents. Another aspect crucial to ensuring interoperability is a standard for encoding dictionaries. To this end, an open-standards-based framework for internationally interoperable lexicographic work is being developed in OASIS.[31] In addition, further developments of TEI Lex-0 (Tasovac et al. 2018), a baseline encoding for TEI with a special focus on retrodigitized dictionaries, and Ontolex-Lemon (Cimiano et al. 2016), the de facto standard for representing lexical information as RDF, have been supported within ELEXIS. All these developments will lead to an increase in sharing, reusing and linking of lexicographic resources.

The results on automatic knowledge extraction suggest that dictionary projects are starting to incorporate automatically extracted data in their resources, but that the rate at which this is happening is not keeping up with technological developments. In fact, the results from the institutional surveys in ELEXIS (2018-2021) show a very similar picture to the results from the ENeL survey conducted in 2015, which is worrying and suggests that lexicographers need to become more pro-active as developments in AI may affect lexicographic work more profoundly than anyone imagines now. This is also why the innovative research that has been carried out within ELEXIS into lexical-semantic analytics for NLP is so important. As a result of this work, new algorithms have been proposed for sense clustering (Martelli et al. 2019, 2022a), domain labelling (Campagnano et al. 2021), and the diachronic distribution of senses (Martelli et al. 2022b). These techniques again enable the development of improved tools for the production of structured proto-lexicographic data in an automated process, meeting a wish expressed by the respondents, i.e. new possibilities for (semi-)automatic dictionary compilation (see e.g. Jakubíček et al. 2020, 2022).

Related to this, we see that crowdsourcing and gamification have not really taken off in lexicography. This is probably not that surprising, as crowdsourcing and gamification have only become well-received topics in lexicography in the past 5 to 7 years, so it is understandable that many projects are still a little cautious. Nevertheless, there are a few success stories such as the semi-automatically created Thesaurus of Modern Slovene (Arhar Holdt et al. 2018) that warrant further research into how crowdsourcing and gamification can best be integrated into the dictionary compilation process. The popularity of the Slovene resource suggests that crowdsourcing can be used to create a 'good' lexicographic resource, counterbalancing the concerns about the low-quality of automatically generated data that were raised frequently by the respondents. With more and more data that needs to be processed, analyzed and (post-)edited, and lexicographers getting more and more tasks and consequently less and less time for editing dictionaries entries, relying at least in part on the wisdom of the crowd seems unavoidable.

From the other side, we clearly see that due to instant online feedback lexicographers interact much more with end users and feel a greater need for deeper user behaviour analytics. Tools and training to help facilitate research into dictionary use form another important future direction in lexicography.

Considering the obstacles mentioned, one of the biggest concerns remains funding. The need for funding is voiced in all parts of Europe, although it seems more urgent in Eastern Europe. A second major source of concern is the low quality and reliability of (semi-) automatically generated resources, while high-quality lexicographic data is still kept under restrictive licenses.

All in all, the surveys allowed us to collect a considerable amount of information and online questionnaires continue to be an efficient means for approaching the lexicographic community. As with all methods of data collection, we are aware that survey research also has its drawbacks. It is a quantitative (rather than qualitative) methodology and even though we did include many open-ended questions to get additional information (which proved to be a good decision), we realize that in certain cases, other methods such as an interview might have been better to allow further clarifications. In this sense, the ELEXIS Skillset report (Tasovac et al. 2019) which was based on in-depth interviews and formed the basis for the ELEXIS curriculum, can be seen as complementary to the quantitative data on training and education collected in our surveys (see Section 4.9).

Furthermore, it is important to note that the lexicographic practice map that we sketched is based on the responses from institutions that were represented in ELEXIS as partners or observers. If only one institution from a country is represented in the surveys, this does not necessarily mean that there are no other institutions doing lexicography in this country. It only means that they were not represented in ELEXIS. We note a few gaps in our data. For instance, there is no separate institute from Belgium, but as the Dutch Language Institute is a binational institute, this partner institute can be considered to represent both the Netherlands and Belgium in the project. Information from Bosnia and Herzegovina, Luxembourg and Malta is also missing as there were no institutions from these countries that joined ELEXIS. These countries were not members in COST ENeL either and were also absent in the survey on the monolingual dictionary (Kosem et al. 2019). We also noted that commercial publishing houses are underrepresented in the ELEXIS landscape.

These are all points that need to be taken into account for future research if we want to get the complete picture of lexicographic practices in all European countries without any exception. A separate study on freelancers would also be useful in order to understand the differences in working conditions and funding schemes they encounter.

6 Conclusion and further research

In this paper, we have presented the combined results of the three surveys that were carried out in the context of the ELEXIS project and we sketched a map of the lexicographic practices in Europe revealing similarities and differences between them at different institutions in different countries.

The ELEXIS lexicographic landscape consists mainly of public institutions and universities that rely heavily on public funding (national and international) for their lexicographic work. They generally possess expertise on different types of dictionaries (general, specialized, historical, dialect, and/or terminological dictionaries). Expertise on terminological resources is, however, more prevalent among the observer institutions than among the ELEXIS lexicographic partners. The results reveal that in Europe, general monolingual dictionaries are commonly compiled at public institutions, whereas bilingual/multilingual dictionaries of all types are more in the domain of universities. The results also suggest that retrodigitization occurs more frequently in public institutions specialized in lexicography than in universities.

In the near future, we can state that the era of stand-alone (paper) dictionaries and closed data will be irrevocably over. Lexicography will encounter the shift towards open access structured data enabling re-use and linking of dictionary data along with aggregating stand-alone lexicographic (and terminological) resources into numerous dictionary portals. Also, the turn towards unified data is expected, with the vision that dictionary publishers will produce a single resource containing all the linguistic data that the publisher has about the language and become more of a data provider and less of a dictionary publisher. This tendency is specially visible at commercial publishing houses. The ones that are not closing are moving away from traditional dictionaries to dictionary portals and API services. The lexicographic community also anticipates intensive integration of lexicographic data into the Semantic Web, AI, NLP, and CALL applications.

Thanks to the COST ENeL Action and the ELEXIS project the lexicographic community in Europe has been brought together at a large scale and the resulting network forms a solid ground for future-oriented lexicographic projects. As a way of securing the sustainability of the infrastructure, the ELEXIS Association has been proposed (cf. Krek et al. 2022). The objective of this new association is the organization and coordination of activities related to lexicography, and activities related to NLP tasks on the topic of semantics, insofar as they are of interest to lexicography. Furthermore, an application has been submitted to CLARIN to establish an ELEXIS Knowledge Centre for Lexicography. On the basis of the ELEXIS Association and the CLARIN Knowledge Centre for Lexicography we plan to continue the research into lexicographic practices and the needs of lexicographers in Europe and beyond in order to monitor the changes in the field and provide input for tasks for the sake of enhancing lexicographic work in the future.

Acknowledgments

The research received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015 and it was supported by the Estonian Research Council grant (PRG 1978). The authors also acknowledge the financial support from the Slovenian Research Agency (research core funding No. P6-0411, *Language Resources and Technologies for Slovene*).

Notes

[1] https://elex.is
[2] https://www.elexicography.eu/
[3] https://elex.is/observers/
[4] https://www.1ka.si/d/en
[5] This applied particularly to those questions that benefited from nesting which was not possible in Google Forms. For instance, in the survey for observers, we first introduced a "yes/no" question such as 'Do you use automatic data extraction from corpora in lexicographic projects at your institution?' and the follow-up question on the types of data that are automatically extracted was only shown in the case of a positive answer to the first question. In the survey for partners this was merged into one question.
[6] The questions for which we do not have results from both surveys or where the results are not entirely comparable are marked explicitly in the analysis.
[7] http://www.efnil.org/
[8] If we consider all the institutions that answered this question and not just those actually working on a lexicographic project, the use of a DWS combined with a CQS is still the most common. However, there are then more institutes using a CQS only (17) as well as institutes not using any tool at all (17).
[9] This explains the differences between the results from the observers and the partners. Among the observer institutions (including more universities), the use of a DWS was less common than among the partners.
[10] https://www.sketchengine.eu/
[11] https://lexonomy.elex.is/ and https://www.lexonomy.eu/
[12] Note that access to the Sketch Engine was funded by the EU through the ELEXIS project between 1 April 2018 and 1 April 2022. The access was provided at no cost to academic institutions and ELEXIS observers, and applied to non-commercial use. Lexonomy is an open-source tool which was further developed and promoted in the context of ELEXIS.
[13] In terminography, a sort of hybrid of a good example and a definition, illustrating the meaning characteristics of a term, but not being a formal definition.
[14] https://www.sketchengine.eu/guide/trends/
[15] https://digital.sprachatlas.ch/stenolabor
[16] https://github.com/INL/taalradar
[17] http://lki.lt/socialine-ir-kulturine-lituanistikos-pletra/zaidimai/
[18] https://play.google.com/store/apps/details?id=si.cjvt.igrabesed&hl=sl&gl=US
[19] https://github.com/elexis-eu/word-games

[20] https://github.com/elexis-eu/CrossTheWord
[21] https://wordnet.princeton.edu/
[22] https://www.emlex.phil.fau.eu/
[23] These concerns are also voiced by associations for lexicography such as EURALEX which led to the adoption of the Resolution at its XVII congress in 2016: https://euralex.org/resolution2016/.
[24] https://matrix.elex.is/
[25] https://campus.dariah.eu/curriculum/the-elexis-curriculum
[26] https://www.dariah.eu/
[27] https://www.clarin.eu/content/factsheet-clarin-upskills
[28] https://elex.is/tools-and-services/elexifinder/
[29] https://elexifier.elex.is/
[30] http://publex.uni-trier.de/
[31] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=lexidma

References

Arhar Holdt, Š., J. Čibej, K. Dobrovoljc, P. Gantar, V. Gorjanc, B. Klemenc, I. Kosem, S. Krek, C. Laskowski and M. Robnik Šikonja. 2018. 'Thesaurus of Modern Slovene: By the Community for the Community' In Čibej, J., Gorjanc, V., Kosem, I. and S. Krek (eds), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, Ljubljana, Slovenia, 17–21 July 2018*, 401–410. Accessed on 20 July 2023. https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202018/118-4-2991-1-10-20180820.pdf.

Boelhouwer, B., I. Kosem, S. Nimb, M. Jakubíček, C. Tiberius, S. Krek and M. Rosenmeier. 2020. ELEXIS Deliverable 6.2 Recommendations on Legal and IPR Issues for Lexicography. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/2020/02/ELEXIS_D6_2_Reccommendations_on_Legal_and_IPR_Issues_for_Lexicography.pdf.

Campagnano, C., F. Martelli, R. Navigli and P. Velardi. 2021. ELEXIS Deliverable 3.3 Lexical-semantic Analytics for NLP: Domain Labeling (Software). Accessed on 20 July 2023. https://elex.is/wp-content/uploads/2021/02/ELEXIS_D3_3_Lexical-Semantic_Analytics_for_NLP_Domain_Labeling.pdf.

Čibej, J., D. Fišer and I. Kosem. 2015. 'The Role of Crowdsourcing in Lexicography' In Kosem, I., Jakubíček, M., Kallas, J. and S. Krek (eds), *Electronic lexicography in the 21st century: linking lexical data in the digital age*. Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd, 73–80. Accessed on 20 July 2023. https://elex.link/elex2015/proceedings/eLex_2015_05_Cibej+Fiser+Kosem.pdf.

Cimiano, P., J.P. McCrae and P. Buitelaar. 2016. Lexicon Model for Ontologies: Community Report, 10 May 2016 Specification. Accessed on 20 July 2023. https://www.w3.org/2016/05/ontolex/.

Jakubíček, M., O. Matuška, M. Cukr and M. Měchura. 2020. ELEXIS Deliverable 4.2. Dictionary Drafting Module. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/2020/02/ELEXIS_D4_2_Dictionary_Drafting_Module.pdf.

Jakubíček. M. 2022. ELEXIS Deliverable 4.9 Evaluation and Assessment of Methods for Crowdsourcing in Lexicography. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/ELEXIS_D4_9_Evaluation_and_assessment_of_methods_for_crowdsourcing_in_lexicography.pdf.

Jakubíček, M., V. Kovář and A. Rambousek. 2022. ELEXIS Deliverable 4.7. Evaluation and Assessment of Methods for Automatic Drafting. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/ELEXIS_D4_7_Evaluation_and_assessment_of_methods_for_automatic_drafting_of_lexicographic_resources.pdf.

Kallas, J., S. Koeva, I. Kosem, M. Langemets and C. Tiberius. 2019a. ELEXIS Deliverable 1.1 Lexicographic Practices in Europe: A Survey of User Needs. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/2020/06/Revised-ELEXIS_D1.1_Lexicographic_Practices_in_Europe_A_Survey_of_User_Needs.pdf.

Kallas, J., S. Koeva, M. Langemets, C. Tiberius and I. Kosem. 2019b. 'Lexicographic Practices in Europe: Results of the ELEXIS Survey on User Needs' In Kosem, I., Zingano Kuhn, T., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubíček, M., Krek, S. and C. Tiberius (eds), Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o., 519–536. Accessed on 20 July 2023. https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_30.pdf.

Kosem, I., S. Krek, P. Gantar, Š. Arhar Holdt, J. Čibej and C. Laskowski. 2018. 'Collocations Dictionary of Modern Slovene' In Čibej, J., Gorjanc, V., Kosem, I. and S. Krek (eds), Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, Ljubljana, Slovenia, 17–21 July 2018, 989–997. Accessed on 20 July 2023. https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202018/118-4-2939-1-10-20180820.pdf.

Kosem, I., R. Lew, C. Müller-Spitzer, M. Ribeiro Silveira and S. Wolfer. 2019. 'The Image of the Monolingual Dictionary across Europe. Results of the European Survey of Dictionary Use and Culture.' *International Journal of Lexicography* 32.1: 92–114. Accessed on 20 July 2023. https://doi.org/10.1093/ijl/ecy022.

Kosem, I., S. Nimb, C. Tiberius, B. Boelhouwer and S. Krek. 2021. 'License to Use: ELEXIS Survey on Licensing Lexicographic Data and Software' In Mitits, L. and S. Kiosses (eds), Lexicography for Inclusion Proceedings of the 19th EURALEX International Congress, 7-9 September 2021, Alexandroupolis, *Vol.* 2, 705–712. Accessed on 20 July 2023. https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202020-2021/EURALEX2020-2021_Vol2-p705-712.pdf.

Krek, S., A. Abel and C. Tiberius. 2015. Dictionary Writing Systems & Corpus Query Systems. Survey – WG3 ENeL. Accessed on 20 July 2023. https://www.elexicography.eu/wp-content/uploads/2015/04/ENeL_WG3_Vienna_DWS_CQS_final_web.pdf.

Krek, S., J.P. McCrae, I. Kosem, T. Wissik, C. Tiberius, R. Navigli and B. S. Pedersen. 2018. 'European Lexicographic Infrastructure (ELEXIS)' In Čibej, J., Gorjanc, V., Kosem, I. and S. Krek (eds), *Proceedings of the XVIII EURALEX International Congress on Lexicography in Global Contexts, Ljubljana, Slovenia, 17–21 July 2018*, 881–892. Accessed on 20 July 2023. http://doi.org/10.5281/zenodo.2599902.

Krek, S., T. Declerck, J.P. McCrae and T. Wissik. 2019. 'Towards a Global Lexicographic Infrastructure' In Adda, G., Choukri, K., Kasinskaite-Buddeberg, I., Mariani, J., Mazo, H., and S. Sakriani (eds), *Collection of research papers of the 1st International Conference on Language Technologies for All,* Paris, 4–6 December 2019, 120–122. Accessed on 20 July 2023. http://doi.org/10.5281/zenodo.3607274.

Krek, S., G. Leban, I. Kosem, A. Repar and A. Sršen. 2022. ELEXIS Deliverable 6.5 Final ELEXIS Interoperability Report including Interaction with CLARIN/DARIAH Services. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/ELEXIS_D6_5_Final_interoperability_report.pdf.

Leroyer, P. and H. Køhler Simonsen. 2020. 'Reconceptualizing Lexicography: The Broad Understanding' In Gavriilidou, Z., Mitsiaki, M. and A. Fliatouras (eds), Proceedings of XIX EURALEX Congress: Lexicography for Inclusion, *Vol.* I. Democritus University of Thrace, 183–192. Accessed on 20 July 2023. https://euralex.org/publications/reconceptualizing-lexicography-the-broad-understanding/.

Martelli, F., R. Navigli, P. Spadoni, G. Stil and P. Velardi. 2019. ELEXIS Deliverable 3.1. Lexical-semantic Analytics for NLP: Sense Clustering. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/2019/08/ELEXIS_D3_1_Lexical_semantic_analytics_for_NLP_sense_clustering_Final.pdf.

Martelli, F., M. Maru, C. Campagnano, R. Navigli, P. Velardi, R.J. Ureña-Ruiz, F. Frontini, V. Quochi, J. Kallas, K. Koppel, M. Langemets, J. de Does, R. Tempelaars, C. Tiberius, R. Costa, A. Salgado, S. Krek, J. Čibej, K. Dobrovoljc, P. Gantar and T. Munda. 2022a. ELEXIS Deliverable 3.8 Lexical-semantic Analytics for NLP - Final Report. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/ELEXIS_D3_8_Lexical-Semantic_Analytics_for_NLP_final_report.pdf.

Martelli, F., R. Navigli and P. Velardi. 2022b. ELEXIS Deliverable 3.7 Lexical-semantic Analytics for NLP: Diachronic Distribution of Senses - Software. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/ELEXIS_D3_7_Lexical-Semantic_Analytics_for_NLP_Diachronic_Distribution_of_Sense_software.pdf.

Romary, L. and T. Tasovac. 2018.'TEI Lex-0: A Target Format for TEI-Encoded Dictionaries and Lexical Resources' In *Book of Abstracts of TEI 2018: TEI as a Global Language*. The 18th Annual TEI Conference and Members' Meeting. September 9-13, 2018 Hitotsubashi Hall, Tokyo, 274–275. Accessed on 20 July 2023. https://zenodo.org/record/2613594#.Y8R4aHbP2Uk.

Müller-Spitzer, C. and A. Koplening. 2014. 'Online Dictionaries: Expectations and Demands' In Müller-Spitzer, C. (ed.), *Using online dictionaries*, Berlin, Boston: De Gruyter, 143–188.

Pedersen, B.S., J.P. McCrae, C. Tiberius and S. Krek. 2018. 'ELEXIS – a European Infrastructure Fostering Cooperation and Information Exchange among Lexicographical Research Communities' In Bond, F., Kuribayashi, T., Fellbaum, C. and P. Vossen (eds), Proceedings of the 9th Global WordNet Conference (GWC 2018), Global Wordnet Association, Singapore, 339–344. Accessed on 20 July 2023. https://doi.org/10.5281/zenodo.2599954.

Tasovac, T., L. Romary, P. Banski, J. Bowers, J. de Does, K. Depuydt, T. Erjavec, A. Geyken, A. Herold, V. Hildenbrandt, M. Khemakhem, S. Petrović, A. Salgado and A. Witt. 2018. TEI Lex-0: A Baseline Encoding for Lexicographic Data. Version 0.9.2. DARIAH Working Group on Lexical Resources. Accessed on 20 July 2023. https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html.

Tasovac, T., M. Monachini and F. Khan. 2019. ELEXIS Deliverable 5.1 ELEXIS Skillset Report.' Accessed on 20 July 2023. https://elex.is/wp-content/uploads/2019/02/ELEXIS_D5_1_ELEXIS_Skillset_Report-1.pdf.

Tasovac, T., R. Costa, F. Khan, I. Kosem, J.P. McCrae, M. Monachini, O. Matuška, S. Petrović, C. Roche, C. Tiberius and T. Wissik. 2020. ELEXIS Deliverable 5.2 Guidelines for Producing ELEXIS Tutorials and Instruction Manuals. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/2020/02/ELEXIS_D5_2_Guidelines_for_Producing_ELEXIS_Tutorials_and_Instruction_Manuals.pdf.

Tasovac, T., C. Tiberius, C. Bamberg, A. Bellandi, T. Burch, R. Costa, M. Ďurčo, F. Frontini, J. Hennemann, K. Heylen, M. Jakubíček, F. Khan, A. Klee, I. Kosem, V. Kovář, O. Matuška, J.P. McCrae, M. Monachini, K. Mörth, T. Munda, V. Quochi, A. Repar, C. Roche, A. Salgado, H. Sievers, T. Váradi, S. Weyand, A. Woldrich, and S. Zhanial. 2022. ELEXIS Deliverable 5.3 Overview of Online Tutorials and Instruction Manuals. Accessed on 20 July 2023. https://elex.is/wp-content/uploads/ELEXIS_D5_3_Overview-of-Online-Tutorials-and-Instruction-Manuals.pdf.

Tiberius, C. and S. Krek. 2014. *Workflow of Corpus-based Lexicography. Deliverable COST-ENeL-WG3 Meeting, July 2014, Bolzano/Bozen*. Accessed on 20 July 2023. https://www.elexicography.eu/wp-content/uploads/2015/04/LexicographicalWorkflow_DeliverableWG3BolzanoMeeting2014.pdf.

Tiberius, C., K. Heylen and S. Krek. 2015. Automatic Knowledge Acquisition for Lexicography. Survey – COST-ENeL-WG3 Meeting, August 2015, Herstmonceux. Accessed on 20 July 2023. https://www.elexicography.eu/wp-content/uploads/2015/10/ENeL_WG3_Survey-AKA4Lexicography-TiberiusHeylenKrek.pptx.

Tiberius, C., J. Kallas, S. Koeva, M. Langemets and I. Kosem. 2022. 'An insight into Lexicographic Practices in Europe. Results of the Extended ELEXIS Survey on User Needs' In Klosa-Kückelhaus, A., Engelberg, S., Möhrs, C and P. Storjohann (eds), Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag, 509–521. Accessed on 20 July 2023. https://euralex.org/publications/an-insight-into-lexicographic-practices-in-europe-results-of-the-extended-elexis-survey-on-user-needs/.

Woldrich, A., T. Goli, I. Kosem, O. Matuška and T. Wissik. 2020. 'ELEXIS: Technical and Social Infrastructure for Lexicography.' K Lexical News. Accessed on 20 July 2023. https://lexicala.com/review/2020/elexis/.

Appendix 1: Partners and Observers in ELEXIS per country

Country	Partner	Observer	Total
Albania		2	2
Austria	1		1
Belarus		1	1
Bulgaria	1	1	2
Croatia		5	5
Czech Rep.	1	1	2
Denmark	2	1	3
Estonia	1		1
Finland		1	1
France		1	1
Germany	1	3	4
Greece		1	1
Hungary	1		1
Iceland		1	1
Indonesia		1	1
Ireland	1	1	2
Israel	1		1
Italy	2	3	5
Latvia		1	1
Lithuania		3	3
Montenegro		1	1
Netherlands	1	1	2
North Macedonia		1	1
Norway		2	2
Poland		2	2
Portugal	1	1	2
Romania		3	3
Russia		1	1
Serbia	1	3	4
Slovakia		1	1
Slovenia	1	2	3
Spain	1	4	5
Sweden		1	1
Switzerland		1	1
Ukraine		2	2
United Kingdom		2	2
USA		1	1
Total	**17**	**56**	**73**