

Stimulating Lexicographical Knowledge Exchange via Trans-national Access – the ELEXIS Travel Grants as a Use Case

Sussi Olsen
University of Copenhagen,
Denmark
saolsen@hum.ku.dk

Bolette S. Pedersen
University of Copenhagen
Denmark
bspedersen@hum.ku.dk

Tanja Wissik
Austrian Centre for Digital
Humanities, Austria
Tanja.Wissik@oeaw.ac.at

Anna Woldrich
Austrian Centre for Digital
Humanities, Austria
Anna.Woldrich@oeaw.ac.at

Simon Krek
Jozef Stefan Institute,
Ljubljana, Slovenia
simon.krek@guest.arnes
.si

Abstract

This paper describes the intermediate outcome of one of the initiatives of the ELEXIS project: Transnational Access. The initiative aims at facilitating interaction between lexicographers/researchers from the EU and associated countries and lexicographical communities throughout Europe by giving out travel grants. Several of the grant holders have visited CLARIN centres, have been acquainted with the CLARIN infrastructure and have used CLARIN tools. The paper reports on the scientific outcome of the visits that have taken place so far: the origin of the grant holders, their level of experience, the kind of research projects the grant holders work with and the outcomes of their visits. Every six months ELEXIS releases a call for grants, the fourth call closed January 2020. Since then calls and visits have been suspended due to the COVID-19 situation. So far 23 visits have been granted in total; 13 of these visits have been concluded and the reports of the grant holders are publicly available at the ELEXIS website.

1 Background and Motivation

Even though lexicography has a long history of international research conferences, it has traditionally been a research area with limited knowledge exchange outside of each lexicographical institution, and in many cases lexicographic data has only been accessible to researchers from the institution who created the data and held the copyright. This tradition is partly related to the fact that practical lexicography has a strong commercial basis; lexicographical data used to be good business. But it also relates to the fact that enabling easy access to restricted data requires significant effort into facilitating and controlling this access - which again requires time and money not easily found in the budgets of lexicographic projects.

To this end, an important objective of the ELEXIS project is to stimulate knowledge exchange between lexicographical research facilities, infrastructures and resources throughout Europe, which can consequently mutually benefit from the vast experience and expertise that exist in the community. Inspired by other EU projects such as EHRI¹, RISIS², InGRID³, and sobigdata⁴, ELEXIS offers transnational access activities in the form of visiting grants that enable researchers, research groups and lexicographers to work with lexicographical data, which are not fully accessible online. Furthermore, grants offer access to professional on the spot expertise in order to ensure and optimise mutual knowledge exchange. Finally, grant recipients can gain knowledge and expertise by working with

This work is licenced under a Creative Commons Attribution 4.0 International Licence.
Licence details: <http://creativecommons.org/licenses/by/4.0/>

¹ <https://ehri-project.eu/ehri-fellowship-call-2016-2018>

² <http://datasets.risis.eu/>

³ <http://www.inclusivegrowth.eu/visiting-grants>

⁴ <http://www.sobigdata.eu/access/transnational>

lexicographers and experts in NLP and artificial intelligence. The CLARIN infrastructure is one of the important infrastructures for these travel grant visits.

The trans-national access activities are expected to have a long-term impact specifically but not only for lesser-resourced languages, boost the network and infrastructure of the European lexicographic community, and facilitate future collaboration and knowledge exchange.

The objectives of the ELEXIS trans-national activities can be summarised as follows:

- to offer opportunities to researchers or research teams to access research facilities with an excellent combination of advanced technology and expertise
- to support training of new specialists in the field of e-lexicography in order to conduct high-quality research and ensure sustainability of the infrastructure
- to ensure support for excellent scholarly research projects and innovative enterprises and also support the complex multi-disciplinary research
- to encourage the integrative use of technology and methodologies as developed in ELEXIS and in the lexicographical institutions.
- to improve the overall services (lexicographic and technical) available to the research community
- to exchange knowledge and experience and to work towards future common projects and objectives
- to create an interdisciplinary community, collaborating on activities that are fully or partially of relevance to the proposed work of the grant holder
- to create knowledge at the interaction between academia and society

The trans-national activities represent a way of ELEXIS to enable access to restricted data, which has so far not been available outside of the hosting institutions, to researchers from other institutions and countries. As the results of research conducted in trans-national activities become available under open-access licenses, the international lexicographic community will become acquainted with previously inaccessible resources.

2 The Grants

The transnational activities consist of visiting grants of 1 to 3 weeks for researchers to experiment with and work on lexicographical data in a context of mutual knowledge exchange with the hosting institutions. Around five visiting grants are made available twice a year during the entire project period, amounting to 35-40 grants in total. However, since February 2020 calls and visits have been temporarily suspended due to the COVID-19 situation.

The following lexicographic institutions accept transnational visits during the ELEXIS project:

1. ELEXIS-SL: Institut Jozef Stefan (JSI, Slovenia)
2. ELEXIS-NL: Institute for Dutch Language (INT, The Netherlands)
3. ELEXIS-AT: Austrian Academy of Sciences (OEAW, Austria)
4. ELEXIS-RS: Belgrade Center for Digital Humanities (BCDH, Serbia)
5. ELEXIS-BG: Institute of Bulgarian Language Lyubomir Andreychin (IBL, Bulgaria)
6. ELEXIS-HU: Hungarian Academy of Sciences (RILMTA, Hungary)
7. ELEXIS-IL: K-Dictionaries (KD, Israel)
8. ELEXIS-DK: Det Danske Sprog- og Litteraturselskab, University of Copenhagen (DSL/UCPH, Denmark)
9. LEXIS-DE: Trier Center for Digital Humanities (TCDH, Germany)
10. ELEXIS-EE: Institute for Estonian Language (EKI, Estonia)
11. ELEXIS-ES: Real Academia Española (RAE, Spain)

Out of the 11 countries, where the hosting institutions are located, 8 countries participate in CLARIN and five out of the 11 hosting institutions are operating CLARIN B Centres.

2.1 The Calls

Researchers and lexicographers within the EU member states and associated countries are invited to apply for a visit of free access to and support from one of the lexicographical institutions.

The calls for applications include descriptions of the institutions and the lexicographical resources, tools, and expertise that are made available for the visitors. Researchers and lexicographers interested in visiting a particular host institution are encouraged to make motivated applications describing their background, the purpose of the visit etc.

2.2 Dissemination and Reporting

The calls are disseminated through the ELEXIS website⁵, mailing lists, newsletters, as well as through Facebook and Twitter. Additionally, attention to the open call(s) is drawn at presentations and/or booths at various conferences, and flyers are distributed to partners and the audience at relevant events. For the dissemination activities we used CLARIN channels such as mailing lists of national CLARIN consortia, the CLARIN Newsflash and the CLARIN Twitter channel.

Particular effort was invested in disseminating the ELEXIS travel grants via social media campaigns on Facebook and Twitter. For the first call, not only the call was advertised but also each hosting institution was presented in a separate post⁶, as seen in Figure 1. For the second, third and fourth call, only the call itself was advertised.

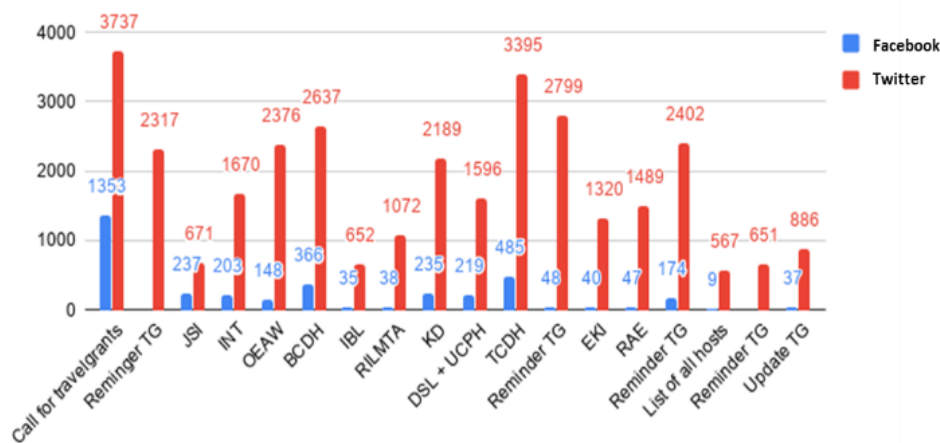


Figure 1. Reach (Facebook)/ Impressions (Twitter) of the dissemination campaign of the 1st call

Furthermore, we accompany the grant holders from the announcement through their travel visits with social media posts as well as website portraits as shown in Figure 2.

⁵ <https://elex.is/grants-for-research-visits/>

⁶ For instance, see: <https://twitter.com/i/events/1025308219126763520>

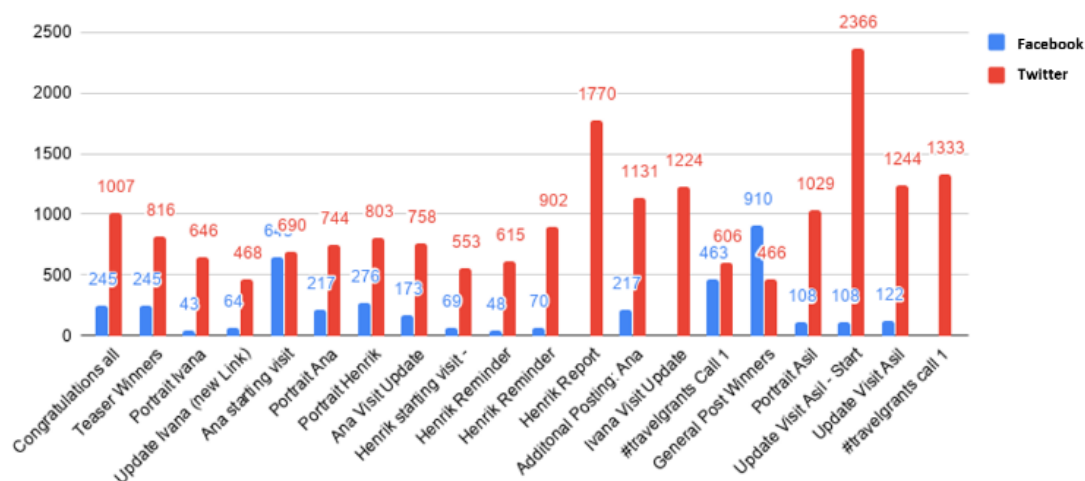


Figure 2. Reach of the dissemination campaign of grant holders of the 1st call.

Besides the website portraits that are published before the research visits in form of a written interview⁷, the grant holders report on their visits afterwards, and the reports are published at the ELEXIS website⁸ and on social media.

The promotion of the open calls via Facebook and Twitter resulted in 88 postings that inform about the open call for research grants (1, 2, 3 and 4) and 106 postings about the grant winners, their projects and journey.

Summing up, in the course of quarter 2 2018 until quarter 1 2020 ELEXIS EU published a grand total of 194 social media postings about the calls for research visit grants on both platforms.

3 Status after four Calls

Four calls for applications were launched and 23 grants were granted, of which 13 were completed, while the other 10 visits have been postponed due the COVID-19 situation. The fifth call has been postponed several times for the same reason and currently it is not to be foreseen when the next call can be launched.

The four calls received applications from Albania, Austria, Bulgaria, Croatia, Denmark, Georgia, Germany, Iceland, Ireland, Israel, Latvia, Poland, Portugal, Republic of North Macedonia, Russia, Slovakia, South Africa⁹, Spain, Turkey, and the UK. The final winners of the first four calls come from institutions located in the countries shown in Figure 3. The fact that the applications received originate from a wide range of countries proves that the transnational access program has indeed reached out to both the strong European lexicographic community as well as to communities that do not have an equivalent strong infrastructure.

⁷ <https://elex.is/category/grants-for-research-visits/>

⁸ <https://elex.is/travel-grant-reports/>

⁹ Since neither Russia nor South Africa are associated countries in the Horizon 2020 program, an application from these countries unfortunately cannot be granted.

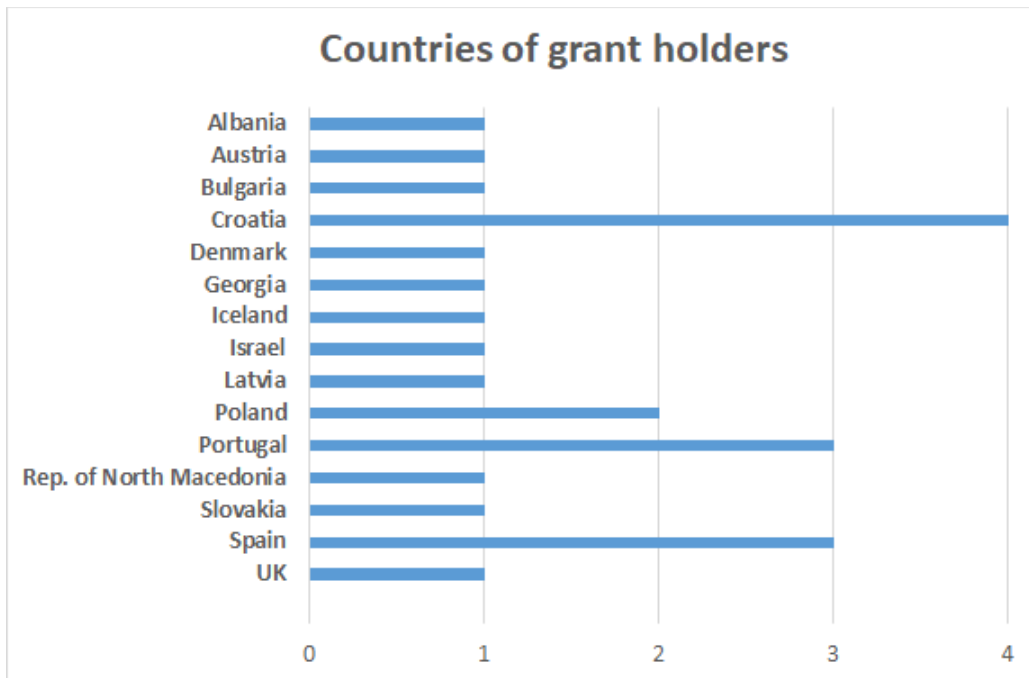


Figure 3. List of countries of the grant winners of the first four calls and number of winners from each country.

After four calls, most of the infrastructures have had one or more visits. Some infrastructures are much more applied to than others. However, since each host has a fixed budget for approximately three visitors, infrastructures that have reached their maximum number of visits and spent their budget are left out of the list of hosting infrastructures for upcoming calls. Hence, in order to make sure that visits are somewhat evenly distributed among infrastructures, there were cases where the Transnational Access Committee, which selects the winners of each call, has given priority to applications addressing less popular infrastructures on the condition that these applications were of sufficient quality.

In order to investigate the research experience of the grant winners, we divided the grant holders into three groups as can be seen in Figure 4.

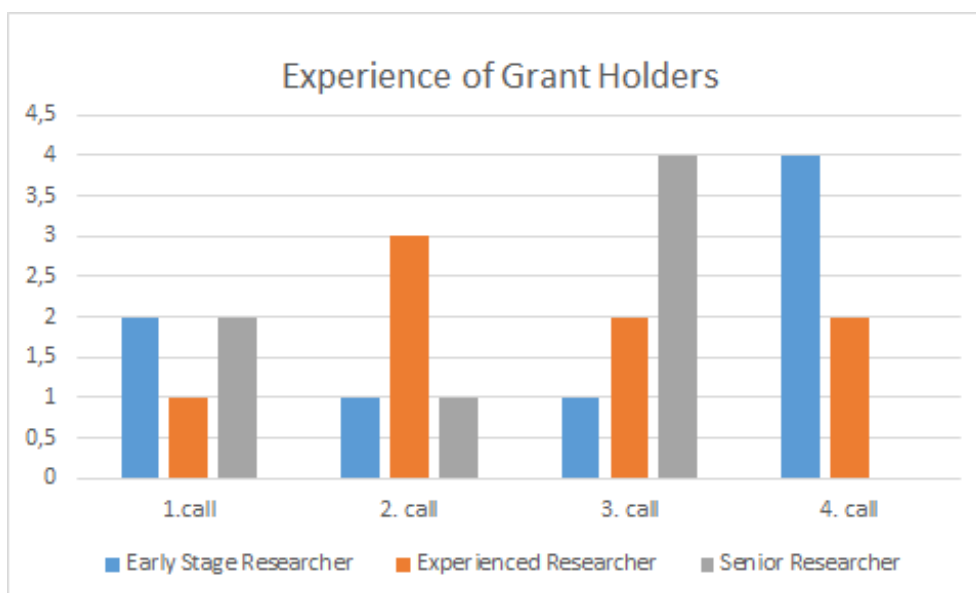


Figure 4. Experience of grant holders divided into three groups.

The early stage researchers span from students doing their master thesis to PhD scholars. In these cases, the visits are reported as having a career boosting effect. Researchers with a couple of years of research experience (after their PhD) belong to the second group, and the last group is defined by senior researchers or lexicographers with years of experience.

As shown in Figure 4, the grant holders represent a good mixture of different levels of experience: Eight early stage researchers, eight experienced researchers and seven senior researchers. At their research visit, most grant holders prove to have limited previous experience in the fields that they work in. Thus, not surprisingly, most of the experienced researchers apply for projects that gratify them with an upgrade of skills in areas that are not previously within their range of experience.

We also did a study of the gender distribution, see Figure 5.

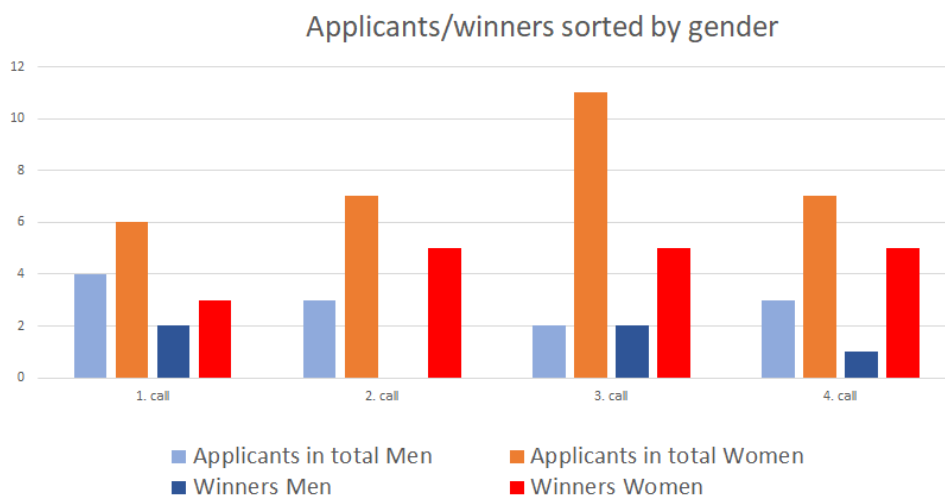


Figure 5. Gender distribution of applicants and grant winners.

We received more than twice as many female applications as male, 12/31. Of the 23 grants given in total, 18 winners were female, 5 male. The success rate has been 53 % on average; for female applications, the success rate is 58 % while only 41 % for the male applicants. It is a fact that lexicography is a field with a strong female representation but it is still a bit unclear what the reason can be for such a large discrepancy.

4 Scientific outcome of the visits

4.1 Research visit projects

Just as the grant holders come from a great variety of countries and communities, the topics of the grant holder projects are quite diverse. Most of the projects focus on the compilation of dictionaries of different kinds and the primary objective of most visits is to be acquainted with the hosts' dictionary writing systems, corpus tools, the methodology behind the dictionary and corpus compilation, standards such as TEI and ISO, and to discuss their own project with experienced lexicographers and terminologists.

From the reports of the visits carried out, the following summary can be made: two projects are about retro-digitalization of older dictionaries, a Latin-Bulgarian dictionary¹⁰ and a Croatian dictionary of literary language¹¹. The grant holders visited different hosting institutions but both gained knowledge

¹⁰ https://elex.is/wp-content/uploads/2019/10/Elexis_report_Elina_Boeva.pdf

¹¹ https://elex.is/wp-content/uploads/2019/03/Elexis-report-on-travel-grant_Ivana-Filipovic-Petrovic.pdf

of corpus digitalization, tools and methods of similar lexical projects as well as hands on experience with dictionary encoding in XML and introduction to the TEI guidelines for dictionaries.

Two projects deal with the creation of dictionaries in combination with terminological content: one proposes a multilingual and multimodal online dictionary combining resources of various kinds (Ramos, 2019), the other an English-Georgian bilingual dictionary of maritime terminology (Tenieshvili, forthcoming). The two grant holders were introduced to relevant terminological as well as lexical projects and to tools and systems that can facilitate their future work, including introduction to international standards such as the ISO standard for term definition.

Two other projects are of a more technical kind: One focuses on comparing the macro/microstructure of a new Portuguese dictionary with the structure of the Diccionario de la Lengua Española (RAE), to which inside access as well as communication with the lexicographers was necessary (Salgado, 2019). In this project, the focus was on the treatment of terminological content in general language dictionaries. The other project dealt with the optimization of methods for automatic extraction of data from a corpus and import into a dictionary writing system¹². Both researchers gained invaluable knowledge for their project thanks to the hosts' introductions to the tools and the various technical solutions.

Another project investigated Nordic e-dictionaries compared to the Croatian Web Dictionary, with a focus on the Danish Dictionary DDO during the visit¹³. The grant holder gained knowledge of the structure of the dictionary as well as the methodology behind, e.g. the lemma selection. The grant holder also participated in workshops about ethical dilemmas in dictionary writing e.g. potentially offensive content, a topic of current interest.

A rather different lexicographic project was a map based data visualization application that analyses the variants of the Spanish language¹⁴. This application could not have been carried out without access to the Dictionary of the Real Academia Española and assistance from its staff. Furthermore this project was also part of a Master's thesis, submitted in 2019¹⁵.

Yet another project dealt with business models of lexicography¹⁶: Based on earlier studies, the grant holder developed a general and more broadly founded business model by arranging several workshops and conducting research interviews at the hosting institution K-Dictionaries.

Common to all the visits are the introduction to dictionaries at the host institution and to the tools, systems and methodology behind. Several visitors received valuable introduction into dictionary writing and structuring including both the linguistic and the technical aspects. Another interesting topic was how to collect a well-balanced corpus for lemma selection.

In their reports all grant holders emphasize the importance of the communication and discussions with the experts of the hosting institution. Several participated in workshops and other events and everybody reports of an instructive and rewarding visit.

4.2 Further scientific dissemination and output related to the research visit projects

Besides the research visit projects and the travel grant reports that are available on the ELEXIS website¹⁷, the travel grant holders of the first call had the opportunity to present their projects during the poster session at the ELEXIS Observer Event in Vienna 2019. Furthermore, several of the research visit projects have led to scientific presentations or publications or were part of a Master thesis. Examples of this kind of output for the first two calls can be found in Woldrich and Wissik (2019).

4.3 Collaboration with and benefits for the CLARIN Network

Some of the ELEXIS travel grant recipients come from countries that are not yet part of CLARIN, hence, they were introduced to the infrastructure during their visit:

¹² https://elex.is/wp-content/uploads/2019/11/final-report_ELEXIS_TanaraZKuhn.pdf

¹³ <https://elex.is/wp-content/uploads/2019/08/Elexis-report-Daria-Lazic.pdf>

¹⁴ https://elex.is/wp-content/uploads/2019/05/Final_Report_ELEXIS_Grant_CETIN_Asil-1.pdf

¹⁵ Asil Çetin (2019). Multi-Faceted Visual Data Analysis for Corpus Research. Master's Thesis. University of Vienna. <http://othes.univie.ac.at/60378/1/64562.pdf>

¹⁶ https://elex.is/wp-content/uploads/2019/01/Report_Transnational_Research_Grant_Henrik_K%C3%B8hler_Simonsen_Research_Visit_Report_final_final.pdf

¹⁷ <https://elex.is/travel-grant-reports/>

The travel grant holder from Spain, who was visiting the Austrian Academy of Sciences in December 2019, was introduced to CLARIN AT and learned about all the facilities that CLARIN offers to researchers. Consequently a visit to the CLARIN K-Centre for Terminology Resources and Translation Corpora at the University of Vienna was organized. This was a perfect opportunity to exchange knowledge in the field of terminology with researchers involved and to learn more about the CLARIN K-Centre infrastructure.

A Croatian grant holder that visited the Society for Danish Language and Literature and University of Copenhagen was introduced to the CLARIN-DK infrastructure and CLARIN EU. Especially the CLARIN-DK NLP tools were of interest to her and with the help of CLARIN-DK staff, some tools from the CLARIN-DK toolbox, i.e. lemmatizer and pos-tagger, were trained for Croatian for the grant holder's future benefit.

During his visit in Ljubljana, a grant holder from the Republic of North Macedonia will be working with CLASSLA, the CLARIN knowledge centre for South Slavic languages. The objective of the visit is to obtain knowledge about corpora management software in order to be able to learn how to train POS taggers for Macedonian, and to create a corpus, ultimately to be used for corpus-based lexicographic work at the Macedonian Academy of Sciences and Arts. Thus it represents an ideal match between CLASSLA's offer of expertise on language resources and technologies for South Slavic languages, and the ELEXIS objectives of bridging the gap between more advanced and lesser-resourced lexicographic communities.

Thanks to the call for ELEXIS Travel Grants, various ELEXIS hosting institutions introduced and will continue to introduce CLARIN to a community that would not have approached a CLARIN centre due to a lack of knowledge. Hence we expect to observe a snowball effect in the future, where ELEXIS grant winners introduce CLARIN within their (national) research communities and approach a CLARIN centre thanks to ELEXIS acting as an intermediate. Through the ELEXIS travel grants, the usage of certain CLARIN tools and services, introduced during the ELEXIS research visits, might increase due to additional users and user scenarios.

In addition, ELEXIS aims to provide interoperability with CLARIN by forming an ELEXIS-CLARIN subgroup¹⁸ to prepare a strategy of integration of ELEXIS services into CLARIN at the end of the project (Summer 2022). We foresee that the subgroup will be formed by members of national CLARIN infrastructures, especially those with CLARIN B centres. Thus, ELEXIS sustainability will be enabled via national consortia to guarantee an afterlife for the efforts made.

5 Conclusion

We have presented the mid-term outcomes of the transnational access initiative of the ELEXIS project where 13 visits to 9 different lexicographical infrastructures in Europe have been completed to date.

The grant holders - be they early stage researchers or senior staff - sought to tailor their research visit in a manner that enabled them to gain new knowledge by physically visiting lexicographical milieus with specific expertise in certain topics and technologies that are highly relevant to their research.

The reports of the grant holders clearly show that the travel grants serve several purposes: the above-mentioned gaining of new knowledge and the network building and knowledge exchange are obvious results. For the individual visiting researchers, however, the visits also serve as a career boost either by helping the early stage researchers establish themselves in the field or by leading the more experienced ones towards new fields. The fact that many experienced researchers apply for a grant to deepen their knowledge and gain new expertise shows that the travel grants meet an existing need not covered by other initiatives.

ELEXIS has a network of 52 observer institutions¹⁹ that benefit from early access to newly developed tools and services, as well as to activities aimed at improving and enriching their own lexicographic data. In the upcoming calls, we expect to receive more applications from these observer institutions. Furthermore, the observing institutions that have lexicographic data, will be invited to join in the network of ELEXIS infrastructures who host the travel grants. They will not receive compensation for work at the institution but visitors will be compensated in the same manner as when visiting existing infrastructures.

¹⁸https://elex.is/wp-content/uploads/2019/04/ELEXIS_03_Observer_Session5_ELEXIS_CLARIN-DARIAH.pdf

¹⁹ <https://elex.is/observers/>

At the time of writing, the second half of the ELEXIS grant activities has been put on hold due to the covid-19 pandemic. Consequently, new calls are being postponed, and already granted and planned visits have not yet been completed. Hopefully, in autumn 2021 it will again be possible to travel around Europe to complete fruitful exchange of lexicographical knowledge. 9 out of 10 of the grant holders who had their visits postponed, claim that they are still very interested in visiting the host institutions when it will once more be possible.

Most presumably, the future grant visits will follow the same line as the previous ones. However, we expect to see an increased interest in the integrative use of the lexicographical tools, methodologies and resources that are just currently being developed and made available through ELEXIS, e.g. the automated linking tool NAISC (McCrae, 2018) and Lexonomy, an open-source platform for writing and publishing dictionaries (Měchura, 2017), which several grant holders already report having worked with.

In conclusion, the ELEXIS travel grants are an opportunity for the CLARIN infrastructure to get known in a community that without ELEXIS would not approach a CLARIN centre. When the project ends we aim at establishing full interoperability with existing CLARIN centres, enabled via national consortia.

References

- McCrae, J. & Buitelaar, P. (2018). Linking Datasets Using Semantic Textual Similarity. *Cybernetics and Information Technologies*. 18. 109-123. 10.2478/cait-2018-0010.
- Měchura, M. B. (2017). 'Introducing Lexonomy: an open-source dictionary writing and publishing system '. In *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 conference*, 19-21 September 2017, Leiden, The Netherlands.
- Olsen, S. & Pedersen, B.S. (2020). D.9.2 Report on trans-national access – year 2. Available at: https://elex.is/wp-content/uploads/2020/07/ELEXIS_D9_2_Report-on-TNA-2.pdf
- Ramos, M., Costa, R. & Roche, C. 2019. Dealing with specialized co-text in text mining: The terminological verbal collocations. In C. Roche, ed. Terminology and Text Mining, TOTh 2019 Proceedings. (https://www.academia.edu/39887099/Dealing_with_specialized_co-text_in_text_mining_The_terminological_verbal_collocations).
- Salgado, A. and Costa, R. (2019). Marcas temáticas en los diccionarios académicos ibéricos: estudio comparativo. In *III Jornadas internacionales sobre investigaciones lexicográficas y lexicológicas (inLÉXICO2019)*, 4–5 abril, 2019, Universidad de Jaén, Spain.
- Tenieshvili, A. (forthcoming). Why It Is Necessary to Create and Adopt Georgian Maritime Terminology?. In *Journal of International Terminology: Translation and Standardization*. Batumi State University, Georgia.
- Woldrich, A. and Wissik, T. (2019). D7.5 First Year Communication and Dissemination Report and Updated Communication Plan. Available at https://elex.is/wp-content/uploads/2019/03/ELEXIS_D7_5_First_year_dissemination_and_communication_report_and_updated_communication_plan.pdf
- Woldrich, A. and Wissik, T. (2020). D7.6 Second Year Communication and Dissemination Report. Available at https://elex.is/wp-content/uploads/2020/05/ELEXIS_D7_6_Second_Year_Dissemination_and_Communication_Report.pdf