

Stimulating Knowledge Exchange via Trans-national Access – the ELEXIS Travel Grants as a Lexicographical 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 Academy of
Sciences, Austria

tanja.wissik@oeaw.ac.at

Anna Woldrich

Austrian Academy of
Sciences, Austria

Anna.Woldrich@oeaw.ac.at

Simon Krek

Jozef Stefan Institute,
Slovenia

simon.krek@ijs.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. 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 trans-national 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

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>

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

with 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 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 7 calls, i.e. 35-40 grants in total.

The following lexicographic institutions accept transnational visits during the ELEXIS project: Jozef Stefan Institute, Institute for Dutch Language, Austrian Academy of Sciences, Belgrade Center for Digital Humanities, Institute of Bulgarian Language Lyubomir Andreychin, Hungarian Academy of Sciences, K-Dictionaries, Det Danske Sprog- og Litteraturselskab, University of Copenhagen, Trier Center for Digital Humanities, Institute for Estonian Language, Real Academia Española.

Out of the 11 countries, where the hosting institutions are located, 8 countries participate in CLARIN and 5 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 ELEXIS website⁵, mailing lists, newsletters (e.g. CLARIN Newsflash), social media and via leaflets at conferences. Particular effort was invested in disseminating the ELEXIS travel grants via social media. For the first call, not only the call was advertised but also each hosting institution was presented in a separate post. Furthermore, we accompany the grant holders from the announcement through their travel visits with social media posts as well as website portraits. In total, we have published 129 posts related to the first three calls.

Besides the website portraits⁶ that are published before the research visits in form of a written interview, the final reports of the grant holders are published at the ELEXIS website and on social media.

3 Status after four Calls

After four calls, 13 visits have been completed, one of these ‘as a virtual visit’ and 10 visits have been postponed due to the uncertain situation with the current pandemic. The next call has been postponed till October for the same reason.

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

⁶ <https://elex.is/category/grants-for-research-visits/>

The winners of the first four calls come from institutions located in the countries shown in figure 1. The fact that the applications received originate from a wide range of countries proves that the transnational access (TNA) program has indeed reached out to the strong European lexicographic community as well as to communities that do not have an equivalent strong infrastructure.

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 visits, 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. In order to make sure that visits are somewhat evenly distributed among infrastructures, there have been cases where the TNA Committee, who selects the winners of each call, has given priority to applications addressing less popular infrastructures on condition that these applications were of sufficient quality.

In order to investigate their research experience, we divided the grant winners into expertise groups, as shown in figure 2. The grant holders represent a good mixture of different levels of experience. At their research visit most grant holders, experienced or not, prove to have limited or little previous experience in the specific fields of their proposed projects. Not surprisingly, most of the more experienced researchers applied with projects aiming to improve their individual skills in specific areas hitherto not part of their research practice.

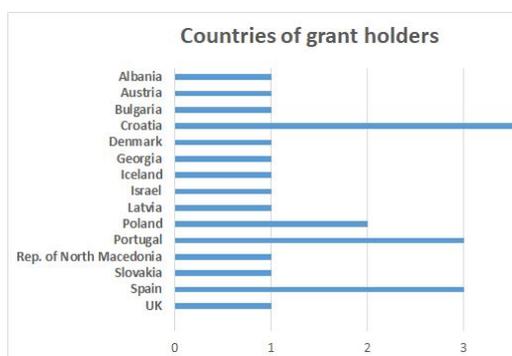


Fig. 1: Countries of grant holders

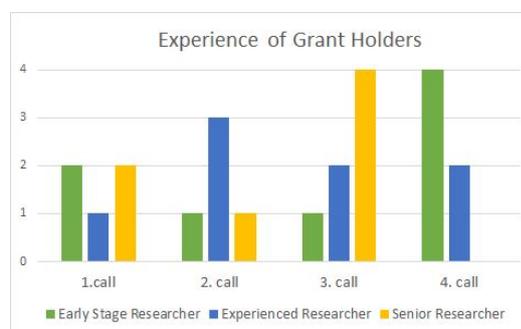


Fig. 2: Experience of grant holders

4 Scientific Outcome of the Visits

4.1 Research Visit Projects

The topics of the travel grant projects are quite diverse. Most of the projects focus on the compilation of dictionaries of different kinds and the primary objective of most visits are 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, see also Olsen & Pedersen 2020.

The projects span from retro-digitalization of older dictionaries, creation of dictionaries with terminological content, comparison of dictionary structures of different kinds, optimization of methods for automatic data extraction from corpora, the development of a data visualisation map and a study of business models of lexicography. Another topic concerns how to collect a well-balanced corpus for lemma selection. Many of the grant holders are very interested in gaining knowledge of the TEI guidelines for dictionaries, the ISO standards for lexicography and terminology as well as of dictionary encoding in XML, i.e. technical knowledge. In addition, a topic of current interest is the focus on the dictionary content in regard to ethical dilemmas in dictionary writing and potentially offensive content.

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

4.2 Further Dissemination Related to the Research Visit projects

Besides the research visit and their published travel grant reports, 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, publications or were part of a Master thesis (cf. Woldrich and Wissik, 2019, 2020).

4.3 Benefits for the CLARIN Network

Some of the ELEXIS travel grant recipients come from countries that are not yet part of CLARIN, and are introduced to CLARIN during their visit, shown by the examples below.

The travel grant holder from Spain, who was visiting the Austrian Academy of Sciences in December 2019, was introduced to CLARIN AT and to all the benefits 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. With the help of CLARIN-DK staff, NLP tools from the CLARIN-DK toolbox, i.e. lemmatiser and pos-tagger, were trained for Croatian for the grant holder's future benefit.

A grant holder from the Republic of North Macedonia will during his visit in Ljubljana 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.

Through the ELEXIS Travel Grant visits at various ELEXIS hosting institutions that are also part of the CLARIN network, CLARIN was and will continue to be introduced, to a community that without the ELEXIS travel grant opportunity 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 community 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 at establishing interoperability with CLARIN by forming an ELEXIS-CLARIN subgroup⁷ 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 - seeked to tailor their research visit

⁷ https://elex.is/wp-content/uploads/2019/04/ELEXIS_03_Observer_Session5_ELEXIS_CLARIN-DARIAH.pdf

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 50 observer institutions⁸ 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 (with 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.

Most presumably, the grant visits of the second half of the project 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.

Concludingly, 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 hopefully 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
- 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

⁸ <https://elex.is/observers/>