

ELEXIS – European Lexicographic Infrastructure

Iztok Kosem

Reliable and accurate information on word meaning and usage is important in the information-driven society of the 21st century. In most European countries, elaborate efforts are put into the development of lexicographic resources describing the language(s) of the community. Although confronted with similar problems relating to technologies for producing and making these resources available, cooperation on a larger European scale has long been limited. Consequently, the lexicographic landscape in Europe is rather heterogeneous. Firstly, it is characterised by stand-alone resources, which are typically encoded in incompatible data structures due to the isolation of efforts, prohibiting reuse of this valuable content in other fields. Secondly, there is a significant variation in the level of expertise and resources available to lexicographers across Europe. This forms a major obstacle to innovative, transnational, data-driven approaches to dictionaries, both as tools and objects of research.

four-year project, which aims to establish a sustainable lexicographic infrastructure, started in February 2018. In order to achieve its objectives, ELEXIS will:

- (a) foster cooperation and knowledge exchange between different research communities in lexicography to bridge the gap between lesser-resourced languages and those with advanced e-lexicographic experience;
- (b) establish common standards and solutions for the development of lexicographic resources;
- (c) develop strategies, tools and standards for extracting, structuring and linking of lexicographic resources;
- (d) enable access to lexicographic standards, methods, data and tools for scientific communities, industries and other stakeholders;
- (e) promote an open access culture in lexicography, in line with the European Commission recommendation on access to and preservation of scientific information.



In 2013, the European lexicographic community was brought together in the European Network of e-Lexicography (ENeL) COST action (<http://www.elexicography.eu>). This initiative was set up to improve the access for the general public to scholarly dictionaries and make them more widely known to a larger audience. In the context of this network, a clear need has emerged for a broader and more systematic exchange of expertise, for the establishment of common standards and solutions for the development and integration of lexicographical resources, and for broadening the scope of application of these quality resources to wider communities, including the Semantic Web, artificial intelligence, NLP and digital humanities. At the end of the COST action, the initiative had been successfully transformed into an H2020 infrastructure project – European Lexicographic Infrastructure (ELEXIS). This

Join ELEXIS

While EU institutions can no longer take part in ELEXIS as partners, they are most welcome to join the project as observers. This provides priority access to certain infrastructure and data and will also include other benefits. Please contact the project leader Simon Krek, info@elex.is, for details.

Transnational visits

During the project lifecycle, ELEXIS will organise transnational calls enabling researchers to work with data with restricted access at host institutions, and to gain knowledge and expertise in close contact with lexicographers and experts in NLP and artificial intelligence. One of the reasons for limited accessibility of lexicographic data outside institutions which are the creators and owners of the data is the efforts needed for their compilation, which necessitate tighter control over access and availability of raw data.

Transnational activities represent one of the mechanisms of ELEXIS to enable access to such content to researchers from other countries. The results of research conducted in transnational activities will be available under open access licences according to the rules of the call, enabling the international community to familiarise itself with previously inaccessible resources. In total, ELEXIS will give access to eleven infrastructures/lexicographical milieus where researchers/lexicographers within the EU member states or associated countries are invited to apply for free-of-charge access via grant visits. During the visits (lasting up to 3 weeks), the hosting institutions will provide support in terms of both lexicographical and technical expertise.

The first deadline for applying is 30 July 2018, covering visits for the period from October 2018 to April 2019. New calls will be launched every six months. More information can be found on the ELEXIS website <https://elex.is/grants-for-research-visits/>.

The next four years thus promise to be very exciting for lexicography, and will likely have a major impact on its future, not only in Europe, but also worldwide.

As a new infrastructure, ELEXIS brings together research and consortium partners from different fields to support the community working in the emerging field of e-lexicography:

- “Jožef Stefan” Institute, Slovenia (leading partner)
<https://www.ijs.si/ijsw/V001/JSI>
- Lexical Computing, Czech Republic
<https://www.lexicalcomputing.com/>
- Dutch Language Institute, Netherlands
<http://www.inl.nl/>
- Sapienza University of Rome, Italy
<http://lcl.uniroma1.it/index.html>
- National University of Ireland, Galway, Ireland
<https://nuig.insight-centre.org/unlp/>
- Austrian Academy of Sciences, Austrian Centre for Digital Humanities, Austria
<https://www.oeaw.ac.at/acdh/acdh-home/>
- Belgrade Center for Digital Humanities, Serbia
<http://www.humanistika.org/>
- Hungarian Academy of Sciences, Research Institute for Linguistics, Hungary
<http://www.nytud.hu/eng/>
- Institute for Bulgarian Language "Prof Lyubomir Andreychin", Bulgaria
<http://ibl.bas.bg/en/>
- Universidade Nova de Lisboa, Faculty of Social Sciences and Humanities, Portugal
<http://www.fcsh.unl.pt/>
- K Dictionaries, Israel
<https://lexicala.com/>
- Institute for Computational Linguistics »A. Zampolli«, Italy
<http://www.ilc.cnr.it/en>
- The Society for Danish Language and Literature, Denmark
<http://dsl.dk/>
- University of Copenhagen, Centre for Language Technology, Denmark
<http://cst.ku.dk/english/>
- Trier University, Centre for Computational Linguistics and Digital Humanities, Germany
<http://kompetenzzentrum.uni-trier.de/de/>
- Institute of the Estonian Language, Estonia
<https://www.eki.ee/en/>
- Spanish Royal Academy, Spain
<http://www.rae.es/>

<https://elex.is/>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.

Free access to Sketch Engine and Lexonomy for academic institutions

First parts of the infrastructure in terms of virtual access provision have already been implemented by providing free access to Sketch Engine (<https://www.sketchengine.eu/>), a leading corpus query system, and Lexonomy (<https://www.lexonomy.eu/>), a simple web-based dictionary writing system. The access has been established as of April 1, 2018 through the single-sign-on facilities of the eduGAIN network of academic identity federations. Thereby anybody academic worldwide with an institution affiliated in eduGAIN can gain access to the services without the need of further registration.

During the first month of availability, more than 100 European universities have obtained such access and delegated an administrator to manage a local quota of 1 billion words that users have available for corpus building. Besides that, they gain access to over 500 preloaded corpora hosted at Sketch Engine as well as all its analytic functions, such as concordancing, building wordlists, compiling word sketches, thesauri and automatic dictionary drafting.

Events with ELEXIS presence

LREC 2018, Miyazaki (May 2018)

- GLOBALEX 2018 Workshop: Lexicography and WordNets (<https://globalex.link/globalex2018/>)
- 6th Workshop on Linked Data in Linguistics: Towards Linguistic Data Science (<http://ldl2018.linguistic-lod.org/>)

EURALEX XVIII, Ljubljana (July 2018, <http://euralex2018.cjvt.si/>)

- Presentation by Simon Krek
- wnlex2018 Workshop: WordNet as Lexicographical Resource (<https://wnlex2018.sciencesconf.org/>)
- TEI-Lex-0 Workshop: Developing a Baseline Encoding for Lexical Data (<http://euralex2018.cjvt.si/programme/co-located-events/>)

Lisbon Summer School: (Area 2) Terminology and Lexicography (July 2018, <http://clunl.fcsh.unl.pt/wp-content/uploads/sites/12/2018/05/PhD-Programme-2-2018-EN-1.pdf>)

CLARIN Annual Conference, Pisa (October 2018, <https://www.clarin.eu/event/2018/clarin-annual-conference-2018-pisa-italy>)

Lexical Data Masterclass, Berlin (December 2018, <https://lexmc.sciencesconf.org/>)

ELEXIS annual conference, Vienna, Austria (February 2019)