Lexicographic Practices in Europe: Results of the ELEXIS Survey on User Needs

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

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

Aim

To get a good overview of

- lexicographic practices both for born-digital and retrodigitised resources
- different tools and methods used by lexicographers
- the needs that lexicographers have now or anticipate to have in the short-term and long-term future

- carried out in the context Horizon 2020 project European Lexicographic Infrastructure (ELEXIS)
- focus on Europe, but also disseminated outside Europe

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

Methodology

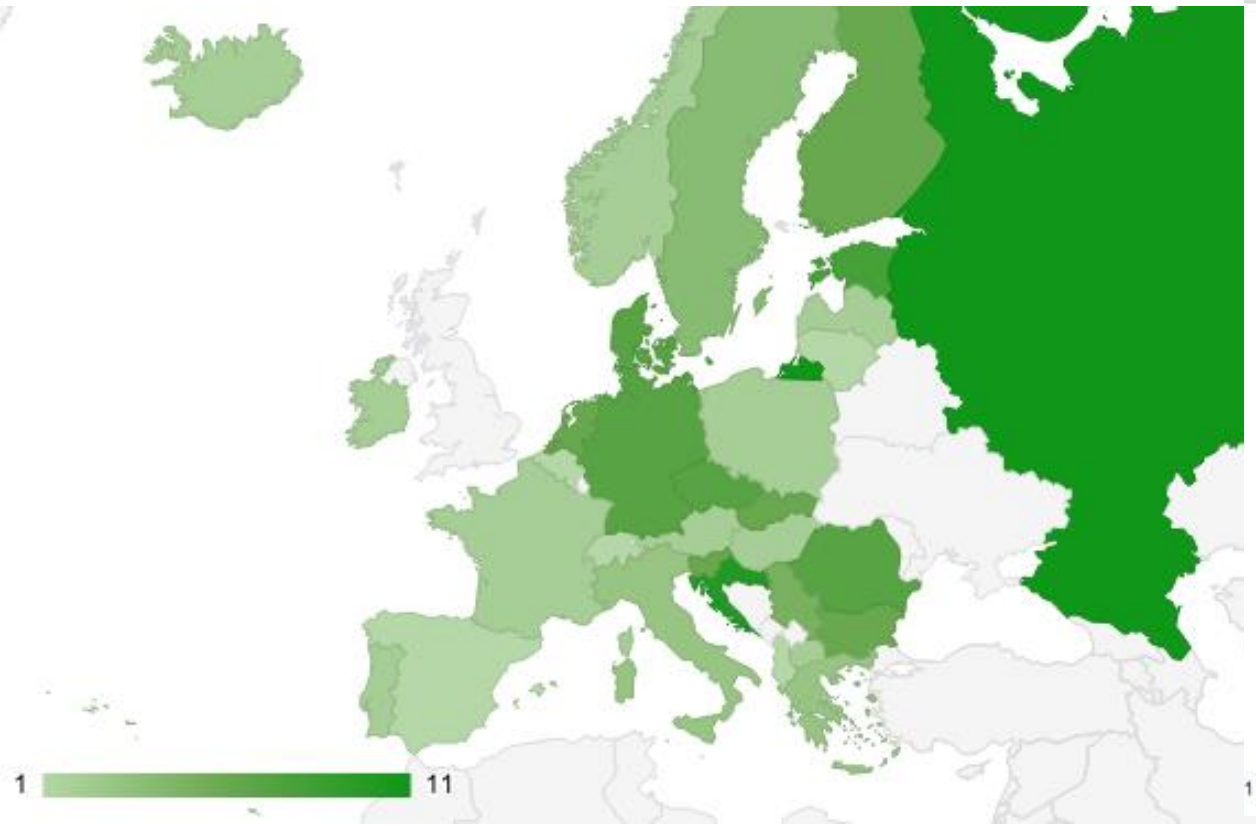
- online questionnaire (Google Forms)
- 13 July 2018 - 1 October 2018
- dissemination: mailing lists, social networks, conferences
- 44 questions
- 6 sections: (1) General information; (2) Ongoing work; (3) Software and tools; (4) Publication; (5) Retrodigitisation; (6) Past and future
- many open-ended questions
- ELEXIS Delivirable 1.1.

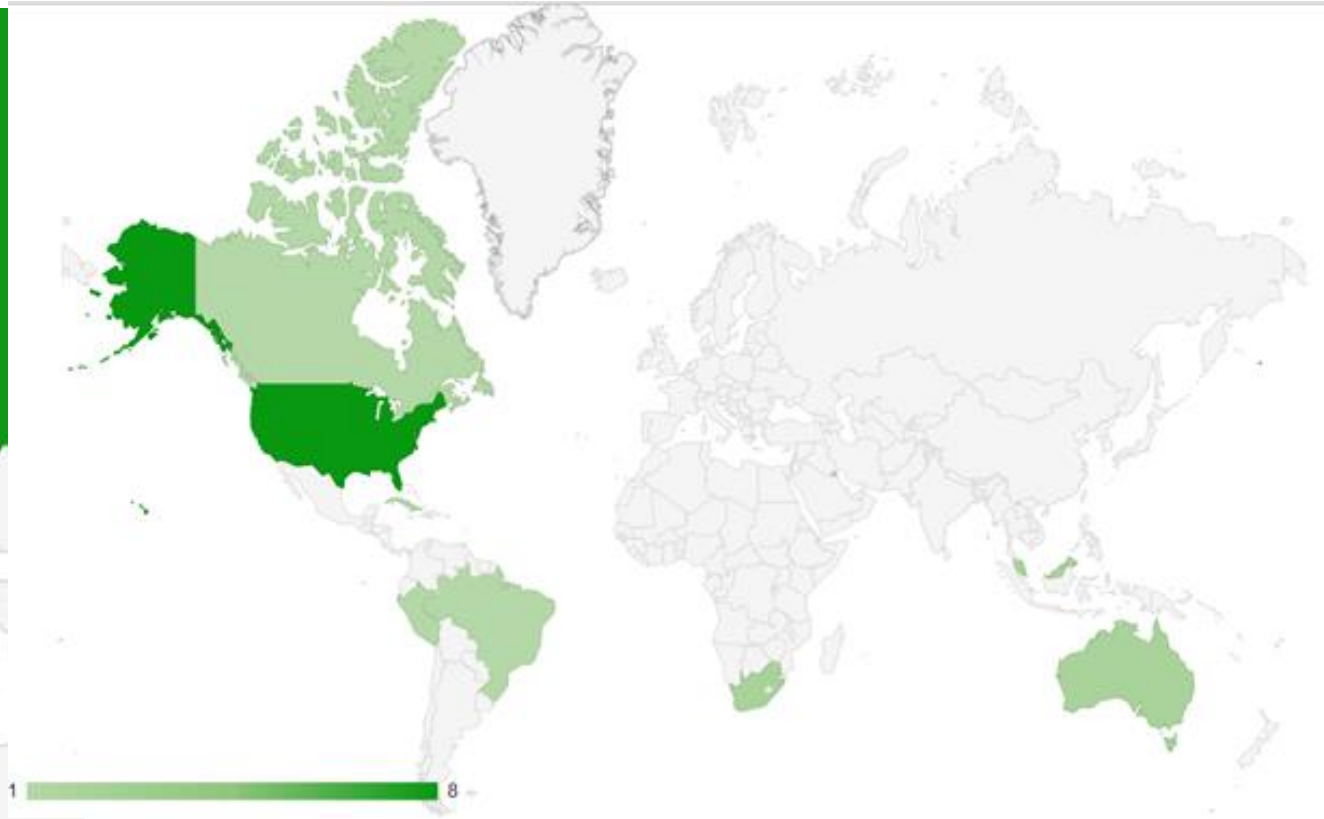
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.

Participation

159 respondents

45 countries (36 European and 9 outside Europe)





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

Methodological concerns

- survey design and formulation of questions affect the answers
- interpretation of terms may differ per person
  (e.g. *born-digital*, *IT-support*, *outsourcing*)
- incomplete participation lowers validity and creates a bias

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

Background surveys (2013-2017)

ISCH COST Action IS1305 European Network of e-Lexicography (ENeL) (2013-2017)

- 300 lexicographers
- 30 countries
- survey on the workflow of corpus-based lexicography (2014)
- survey on Dictionary Writing Systems and Corpus Query Systems (2014/2015)
- survey on automatic knowledge acquisition for lexicography (2015)

Horizon 2020 project European Lexicographic Infrastructure (ELEXIS) (2018-2022)

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

General background of respondent

- works at public institution on monolingual dictionary, which will be published online
- has a PhD and works for more than 20 years in lexicography
- has been trained within his/her own institution
- works in a team of 3-6 people that consists only of people from his/her institution

eLex 2017 group photo

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

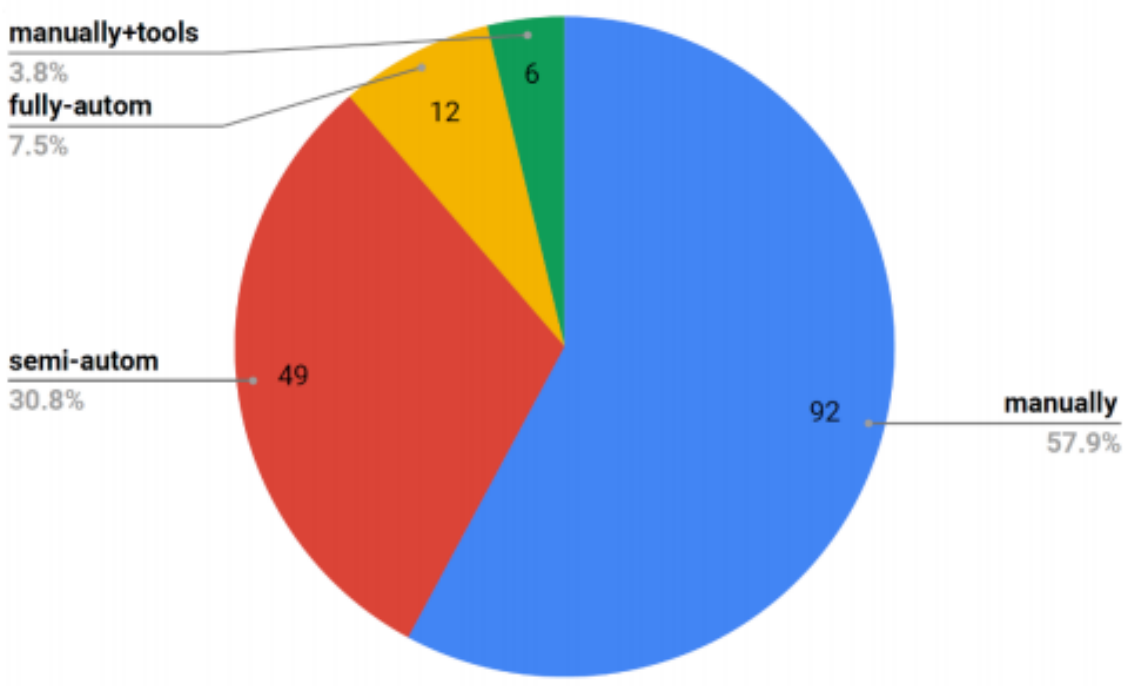
Software and tools (N=89)

55.7% use both, Dictionary Writing System (DWS) and Corpus Query System (CQS)

- 15 DWS
  - commercial, e.g. IDM, Tlex, iLex, Multiterm
  - open-source, e.g. Lexonomy, FLEx (Fieldworks Language Explorer)
  - in-house, e.g. INT-DWS (Holland), DGD (Germany), Ekilex (Estonia)
- 22 CQS
  - commercial, e.g. **Sketch Engine (54.8%)**, Tlex
  - open-source, e.g. Corpus Workbench (CWB), Korp, NoSketchEngine, AntConc, COSMAS II
  - in-house, e.g. CoRest (Denmark)
- Common model: in-house DWS + commercial CQS, one DWS and one CQS

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

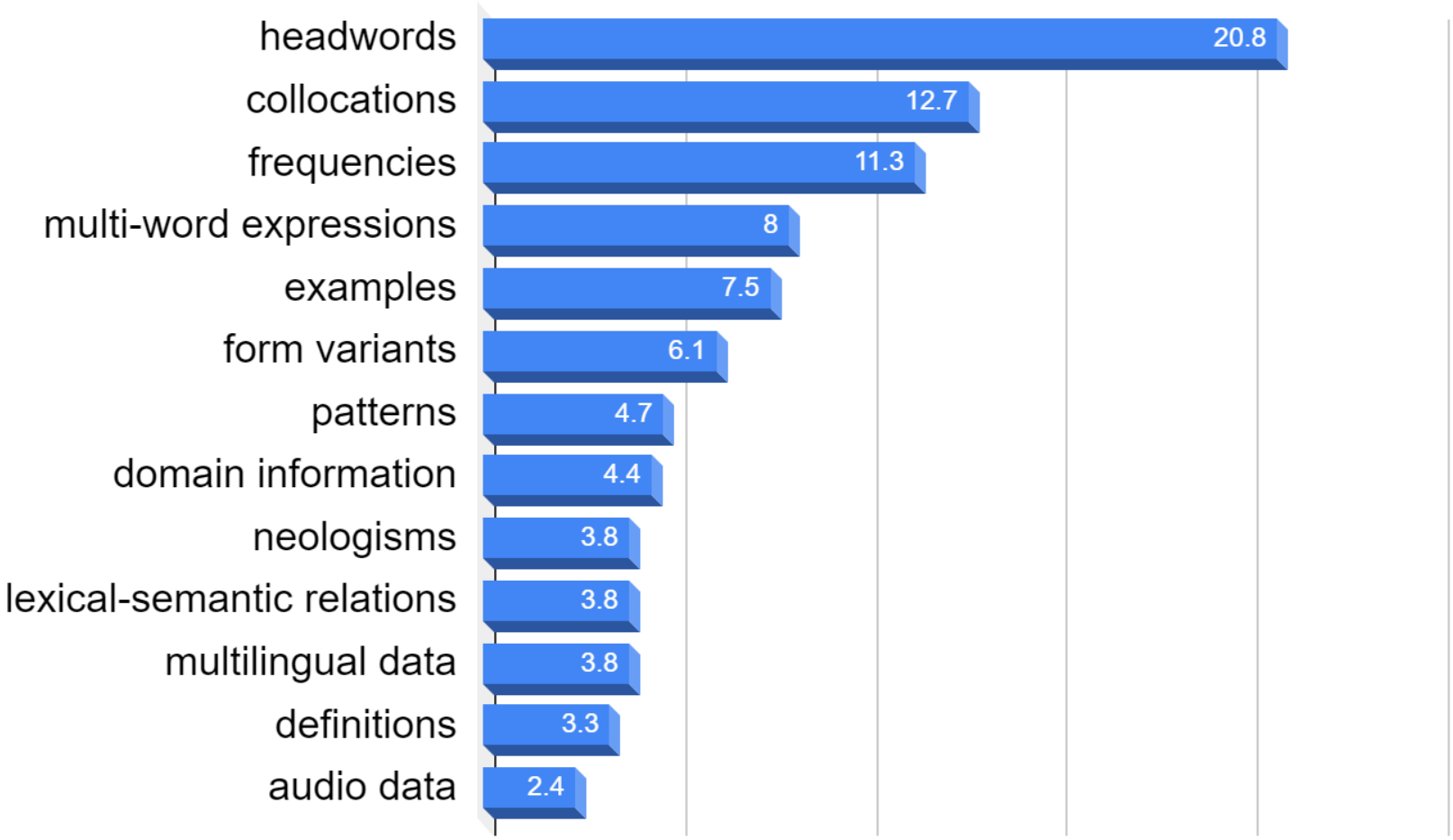
Compiling methods (N=159)



Compiling methods for all projects (N=159)

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

Automatic data extraction (N=150)



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

The changing job of the lexicographer

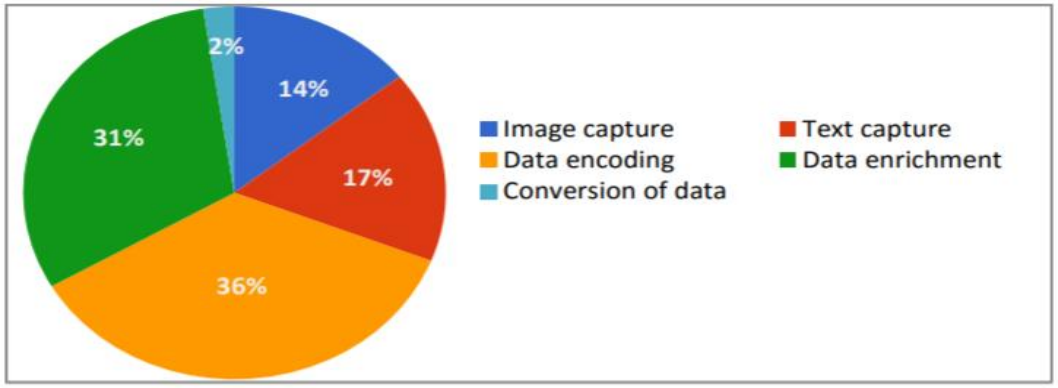
- not just editing dictionary entries
- online publication, user research and user involvement
  - analysis of user logs and user feedback
  - analysis of data from Google Analytics
  - creating add-on materials (blogs, word games)
  - promotional activities
  - project management

  - crowdsourcing
  - gamification

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

The changing job of the lexicographer

- evaluating the user interface
- communication with IT specialists, incl. UX/IX designers
  - IT tasks are the only tasks that are outsourced (from designing the online interface to offering support in the use of DWS and CQS)
  - problems: the cost and lack of regular funding
- retrodigitisation (10% of respondents)
  - data encoding
  - data enrichment



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

Wishes and Needs: software

- online, fast, open-source, browser independent, intuitive, easy to maintain, real-time saving
- **interoperability** with other resources, operating systems and tools (possibility for automatic pre-compilation of entries)
- **integration of CQS and DWS**
- better tools for extraction and automatic processing of data from corpora
- dictionary drafting
- **customisability** (in terms of functionalities and interface)

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

Wishes and Needs: software

- API access, proper documentation (not a black-box system)
- support for collaborative work
- online publishing of the results
- ideally free

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

Wishes and Needs: recent trends

- sense clustering
- syntactic and semantic annotation
- detection of neologisms
- automatic acquisition of translation equivalents and lexical-semantic relations
- diachronic analysis
- (cross-linguistic) interlinking

- data visualisation

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

Wishes and Needs: recent trends

- tools for harmonization of dictionary formats
- common **standard** for the development of lexicographic resources

- tools for crowdsourcing and gamification

- different presentation modes, including mobile applications

- combining a synchronic and a diachronic approach in one resource
- common infrastructure and better tools for retrodigitising

- central repository

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

Wishes and Needs: training

- degree programmes (e.g. EMLex, European Master in lexicography)
- training materials, more case studies
- **interdisciplinarity**

NB! Lexicographic community is very heterogeneous

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

How will ELEXIS help?

- training and education (e.g. DARIAH-Campus platform, grants for research visits)

- best practices and standardisation
  - Elexis deliverables:
    - Deliverable 3.1: Lexical-semantic analytics for NLP: sense clustering
    - Deliverable 6.1: Early ELEXIS interoperability report

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

elexis european lexicographic infrastructure

How will ELEXIS help?

- tools
  - Sketch Engine, Lexonomy, Elexifier, Naisc ...
  - crowdsourcing and gamification tools
- automatic knowledge extraction (Oneclick Dictionary)
- services
  - Elexifinder, News feed

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