User needs

Lexicographic Practices: Carole Tiberius, INT

Skills analysis: Toma Tasovac, BCDH

Lexicographic Practices in Europe: A Survey of User Needs

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

T1.1

Aim of the survey

To get a good overview of

- different tools and methods used by lexicographers around Europe,
- the needs that they have now or anticipate to have in the short-term and long-term future

Method

- Two surveys
  - Survey for lexicographers (44 questions)
  - Survey for institutions (86 questions)
- Questions:
  (1) General information;
  (2) Types of lexicographic resources, software and tools supporting the workflow;
  (3) Publication and access. Crowdsourcing and gamification;
  (4) Retrodigitised dictionaries;
  (5) Data formats. Metadata. Availability; (only in survey for institutions)
  (6) Past and Future.

- Get good coverage
- Online questionnaire (Google Forms)

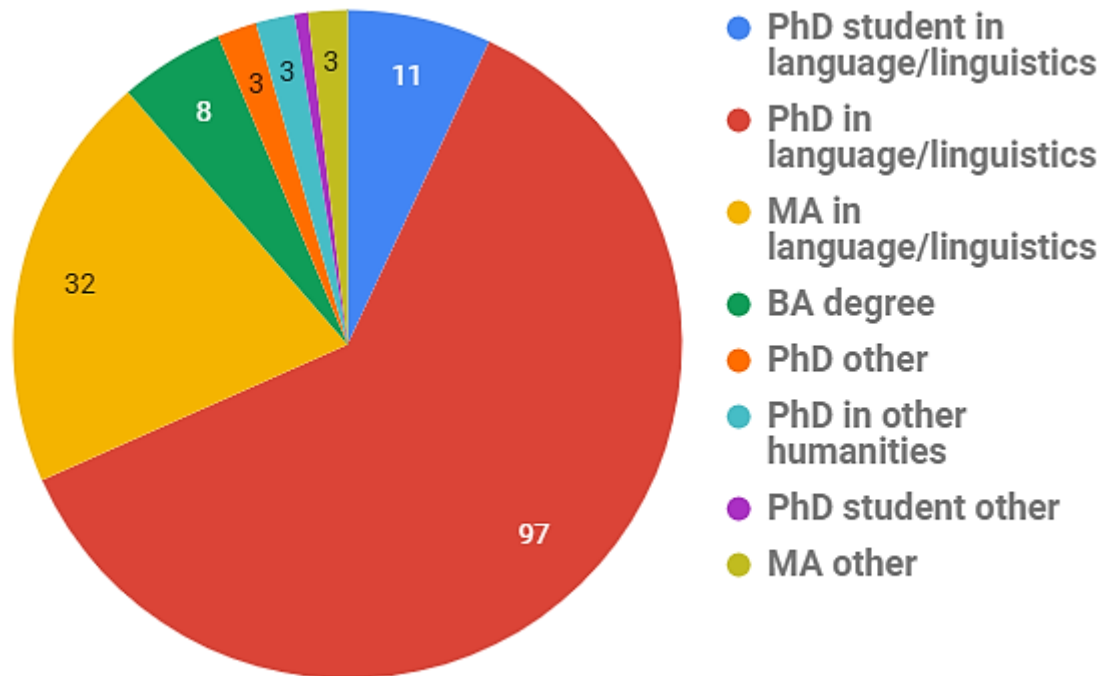# Results: survey for lexicographers (N=159)
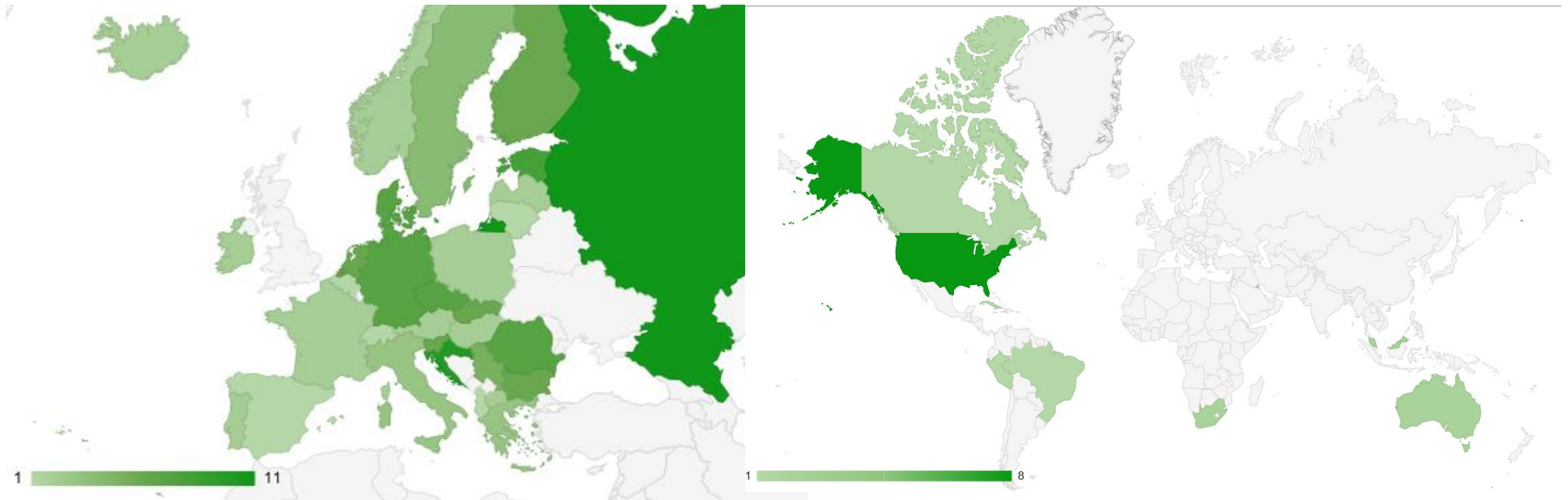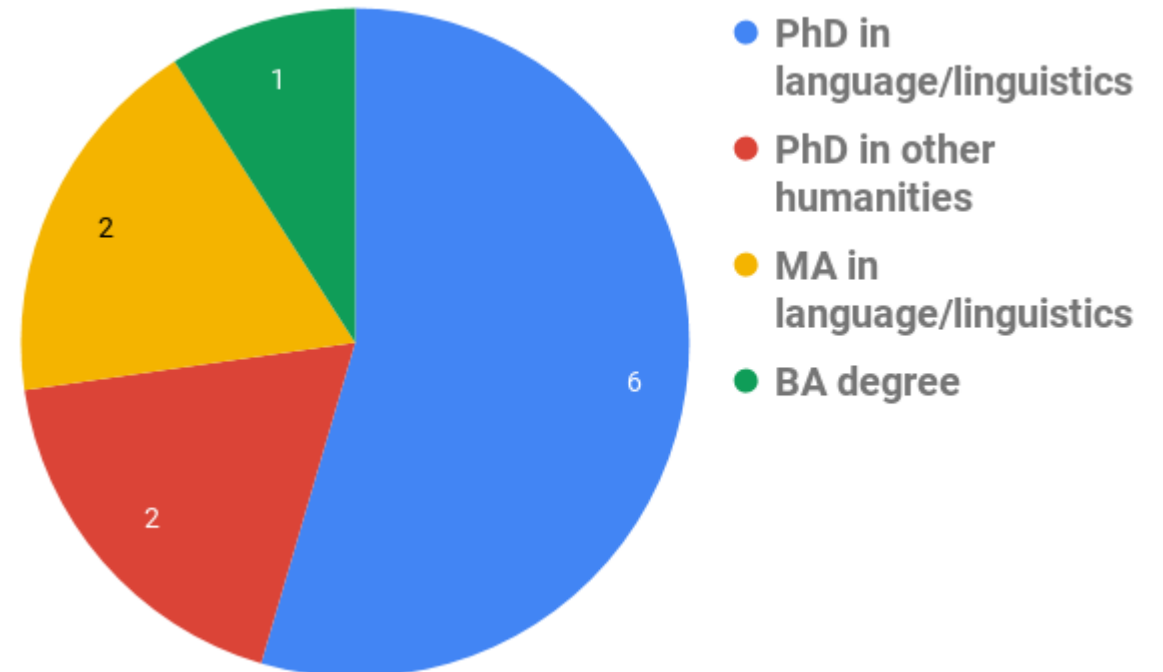
Results: survey for institutions

Name of Institution	Short name	Country
Austrian Academy of Sciences: Centre for Digital Humanities	OEAW	Austria
Institute for Bulgarian Language	IBL	Bulgaria
Society for Danish Language and Literature	DSL	Denmark
Institute of the Estonian Language	EKI	Estonia
Trier University, Trier Center for Digital Humanities	TCDH	Germany
Hungarian Academy of Sciences, Research Institute for Linguistics	RILMTA	Hungary
K Dictionaries	KD	Israel
Instituut voor de Nederlandse Taal	INT	Netherlands
Belgrade Center for Digital Humanities	BCDH	Serbia
"Jožef Stefan" Institute	JSI	Slovenia
Real Academia Española	RAE	Spain

Insights: Education

survey for lexicographers (N=159)

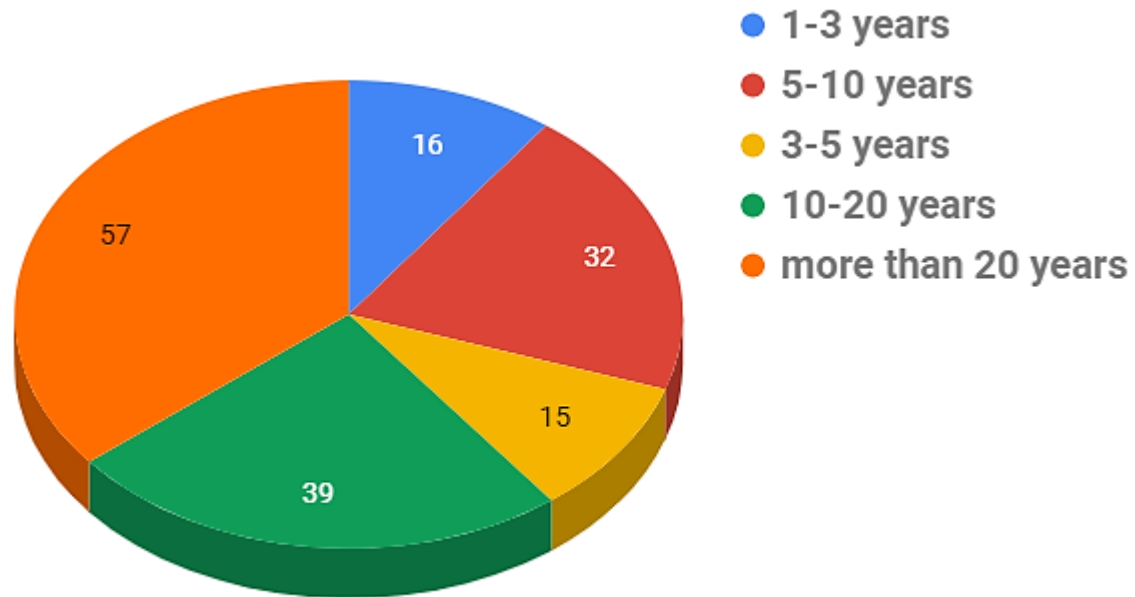**survey for institutions (N=11)**

Insights: Training

Specialised lexicographic training:

- In house training, courses

- External courses, workshops, summer schools

- University course

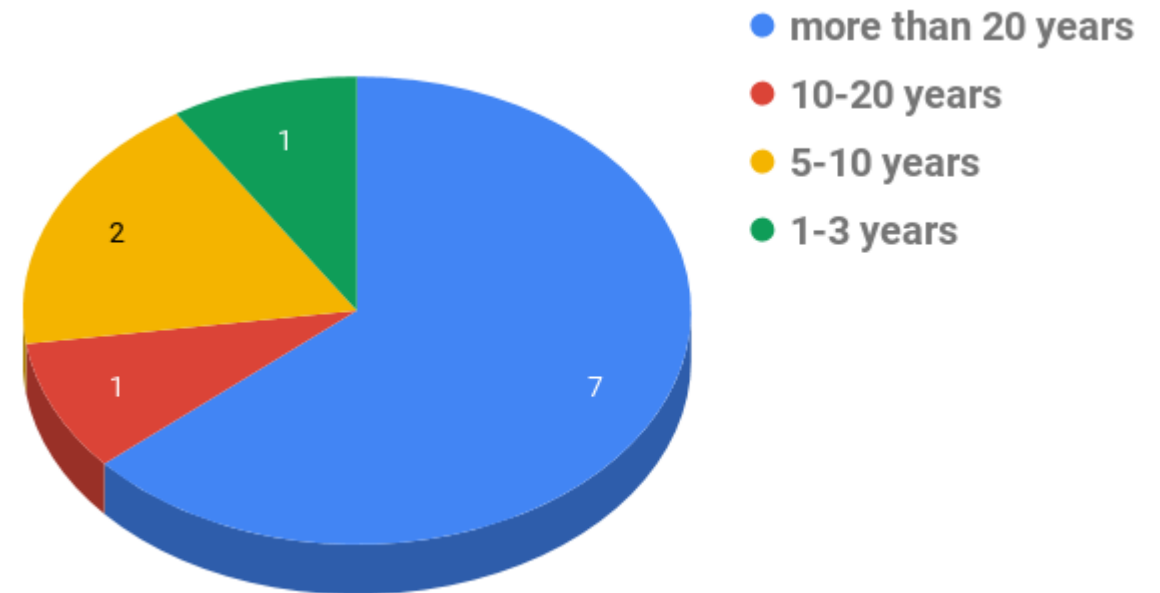# Insights: Working years as a lexicographer

survey for lexicographers (N=159)

- 1-3 years
- 5-10 years
- 3-5 years
- 10-20 years
- more than 20 years

survey for institutions (N=11)

- more than 20 years
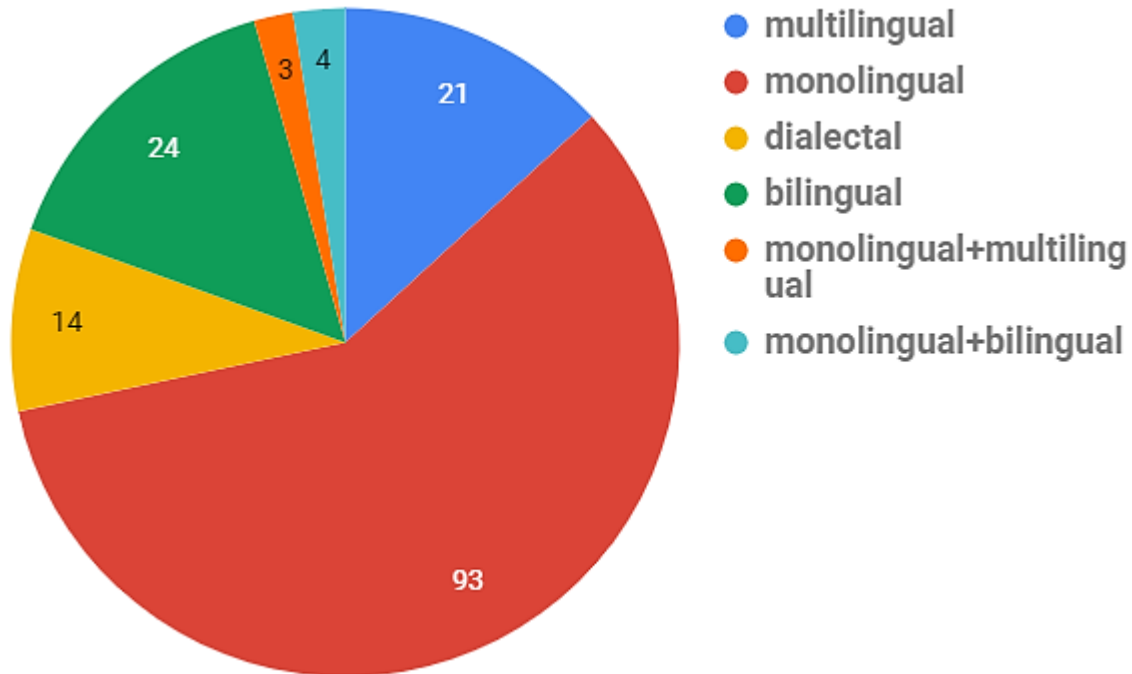- 10-20 years
- 5-10 years
- 1-3 years

Insights: lexicographic landscape

- Most respondents came from public institutions and non-profit organisations.

- Suggests that lexicographic work in Europe is mainly done in public institutions and non-profit organisations.

Insights: Type of projects and publication medium

survey for lexicographers (N=159)

- ● multilingual
- ● monolingual
- ● dialectal
- ● bilingual
- ● monolingual+multilingual
- ● monolingual+bilingual

21, 93, 14, 24, 3, 4

survey for lexicographers (N=150) (124 projects)
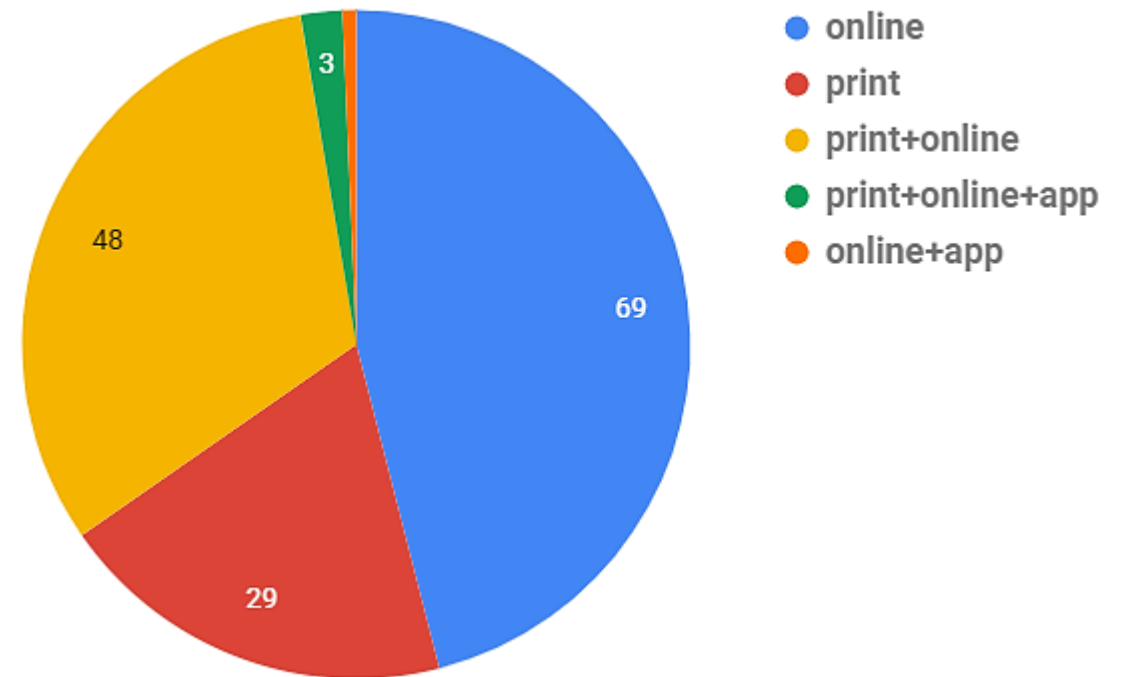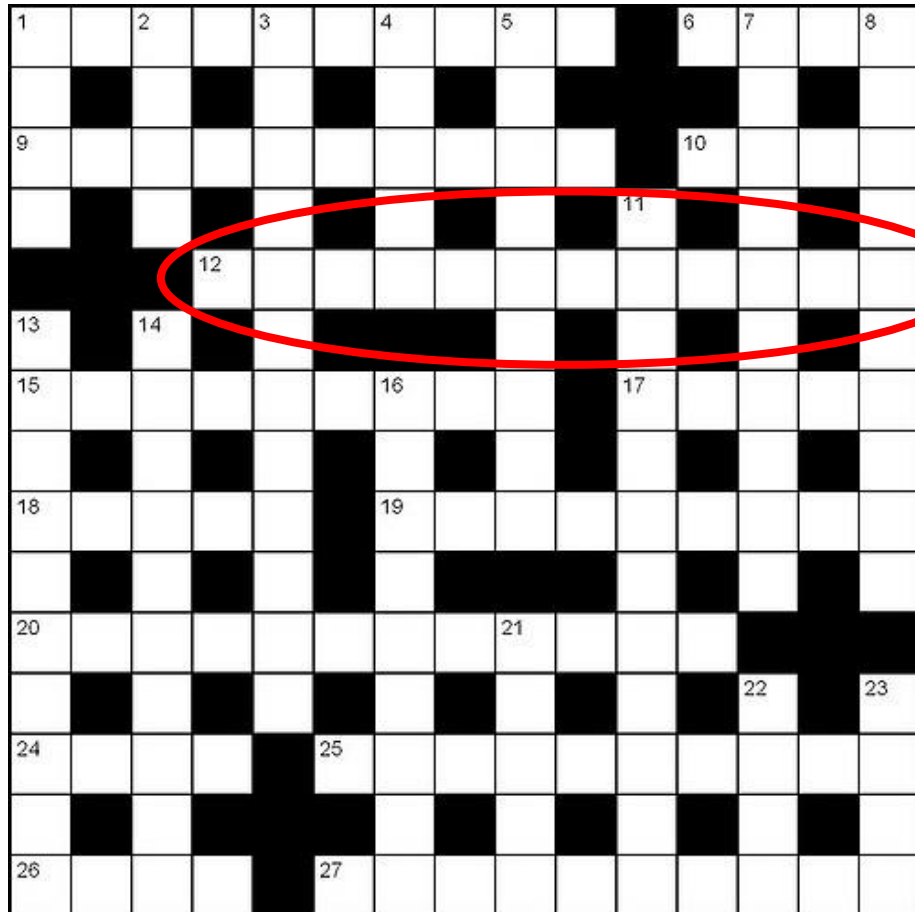
- ● online
- ● print
- ● print+online
- ● print+online+app
- ● online+app

69, 29, 48, 3

Is this definition clear?

ACROSS

10.   ...

12.   A dictionary conceptualized for the electronic medium, offering radically different options for organisation and presentation of lexical information

Is this definition clear?

ACROSS

10. ...

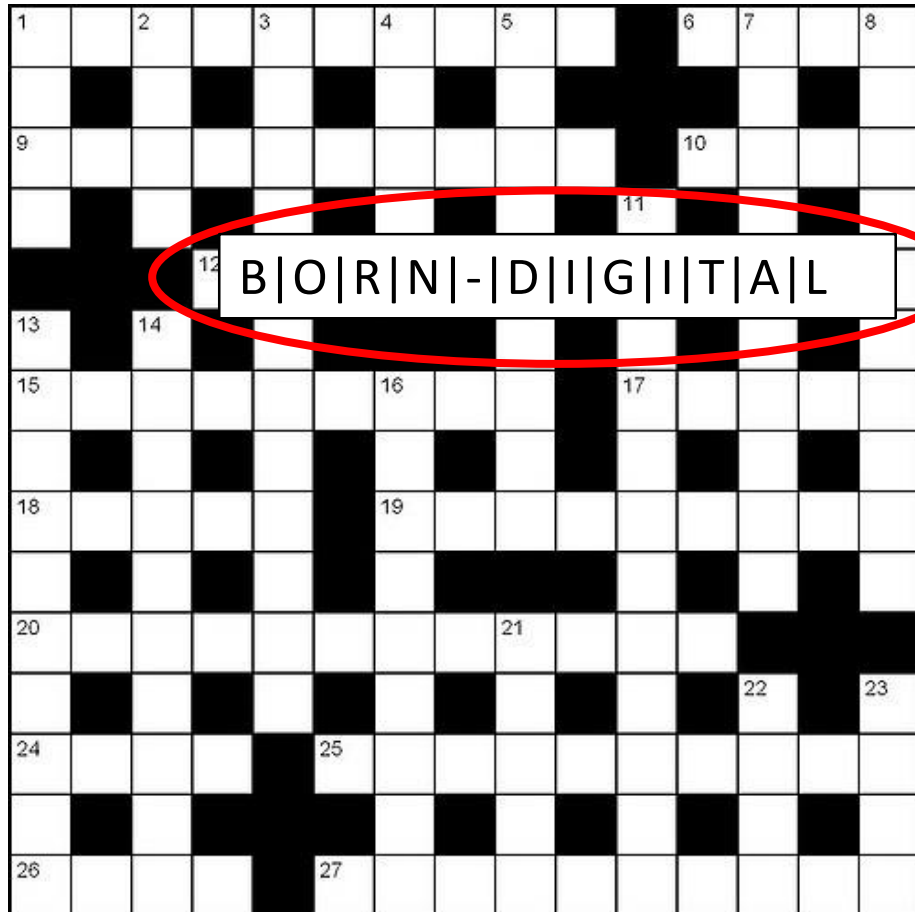12. A dictionary conceptualized for the electronic medium, offering radically different options for organisation and presentation of lexical information

Insights: Born-digital dictionaries and their compilation methods

survey for lexicographers (N=159)

among N=159: 65 born-digital
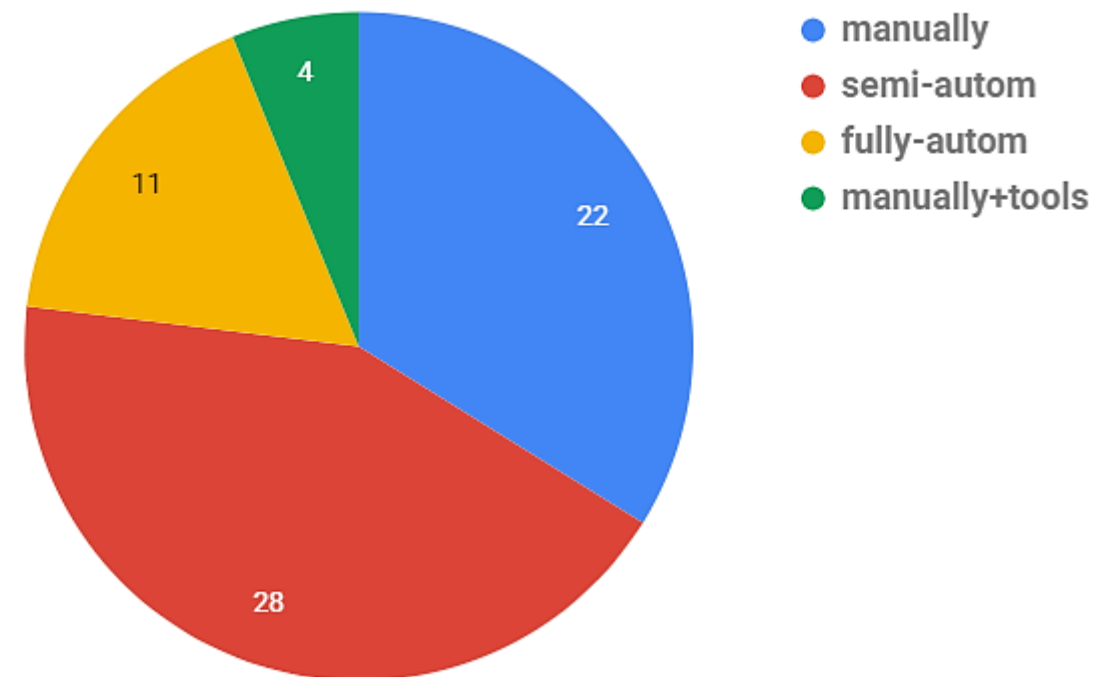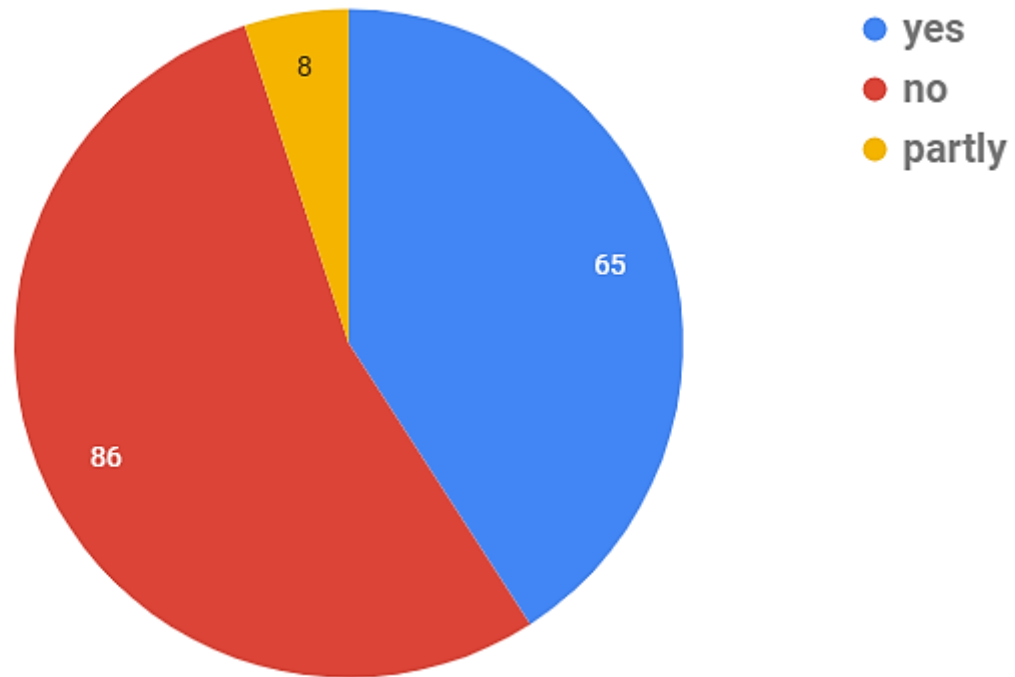
Implications for ELEXIS

- WP1 T1.3 "Best Practices for Lexicography",
- WP2 T2.1: "Common models and protocols for lexicon access",
- WP4: "NLP for lexicography"
- WP5: "Training and Education"

Implications for ELEXIS

ELEXIS Objective: establish **common standards** and **solutions** for the development of lexicographic resources (WP1,WP2,WP4,WP6)

Survey Results: a clear need for common standards and solutions

Implications for ELEXIS

ELEXIS Objective: establish **common standards** and **solutions** for the development of lexicographic resources

Two complementary sets of tools will be provided: lexicographic workflow tools and crowdsourcing and gamification tools (WP4)

Survey Results: wishes and needs for software and tools, e.g. integrated DWS/CQS, interoperability, user-friendly, online, open-source, support for collaborative work, tools for crowdsourcing, API access, fast

Implications for ELEXIS

ELEXIS Objective: promote an **open access culture** in lexicography, in line with the European Commission Recommendation on access to and preservation of scientific information.

Survey Results: free online availability is preferred by the academic partner institutions

Implications for ELEXIS

ELEXIS Objective: foster cooperation and **knowledge exchange** between different research communities in lexicography in order to bridge the gap between lesser-resourced languages and those with advanced e-lexicographic experience (WP5)

Survey Results: specific lexicographic training is often received on the job. In-house training is most common, usually given by a tutor or a senior lexicographer, but external courses, workshops or summer schools are also a popular means of training lexicographers.

The e-lexicographer?

- The role of lexicographers and the tasks they do are changing rapidly.

- These days, lexicographers have to be technically skilled.

- 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.

ELEXIS Skillset Report

Toma Tasovac

Belgrade Center for Digital Humanities (BCDH)

Why does ELEXIS need training and education?

- ELEXIS: knowledge network and service provider
- Diverse user base
- **Training and education is a core requirement for the social sustainability of research infrastructures**

ELEXIS WP5: Training and Education

- Terminology and Lexicography strand at the **Lisbon Summer School in Linguistics** 2018
  - *Semantics in Terminology: The Contribution of Concept to the Meaning of Term* (taught by Christophe Roche from the Université Savoie Mont-Blanc, France)
  - *Trends and Advanced Approaches in Lexicography* (taught by Teresa Lino, Raquel Amaro, and Rute Costa from the NOVA CLUNL, NOVA FCSH, Portugal); and
  - *From Print to Screen: The Theory and Practice of Digitizing Dictionaries* (taught by Toma Tasovac from the Belgrade Center for Digital Humanities, BCDH).
- **Lexical Data Masterclass** by DARIAH at the Berlin-Brandenburg Akademie der Wissenschaften, December 2018
  - TEI, XPath, XSLT, GROBID Dictionaries

ELEXIS Skillset Report

- In-depth video interviews with 11 lexicographers from across Europe chosen based on
  - their professional expertise in lexicography and related fields; and
  - their familiarity with the training and educational landscape in their own communities
- Qualitative assessment
- Does not pretend to be representative, yet our informants offer very different perspectives (digital-born vs. retro, lexicography vs. NLP, university professors vs PhD students, lesser-resourced vs. more advanced scholarly communities etc.)

Key findings

- Almost all of our informants received no formal lexicographic training

- Almost all of our informants describe themselves as lexicographic self-learners who honed their skills by:
    - consulting various dictionaries;
    - reading the scholarly literature;
    - attending workshops and seminars outside the regular university curricula; and/or
    - learning from their colleagues or on the job.

Challenges

- Lack of university curricula in lexicography or systemic, integrated training outside the university
  - *The knowledge gained in unconnected workshops is partial and segmented* (Hudeček)
- If lexicography is taught as a university subject, it is in some countries too theoretical and removed from contemporary trends and best practices
- If students have a background in linguistics, they don't necessarily have the necessary technical skills, and vice versa
- **Technology is not the ultimate challenge: the reluctance to adopt new methods is also a social and cultural challenge**
- **Not all lexicographers believe in a culture of open access, open source and sharing lexicographic data.**

Critical use of tools

- *The tools of today are not necessarily the tools of tomorrow. We should prepare students not only how to learn tools but also to be able to switch to different tools.* (Costa)

Face-to-face is the preferred training method

- presential settings facilitate communication between instructors and students
- intensive face-to-face workshops provide a dedicated space for learning, without the usual distractions of day-to-day obligations
  - *Face to face training is the best. In the five days I spent in Berlin, I managed to digitize a dictionary from scratch without any previous knowledge of GROBID dictionaries.* (Mickoski)
- learning is a social activity
- infrastructural services can benefit from being tested in face-to-face settings
  - *We see a lot of different real-life examples [when we conduct training measures] so we can make sure that what we put in TEI Lex-0 reflects the expectations and constraints of the projects we are dealing with in the classes we teach*(Romary)
- Face-to-face training is expensive and time-consuming. One model that can be used to overcome this obstacle is *blended learning*, an approach which combines online training materials with traditional face-to-face opportunities.

- Training materials should be on a wide range of topics and openly accessible to everybody
  - *ELEXIS should contribute towards setting a general training standard as well as a reference point for lexicographic practice in general.* (Grønvik)
  - *ELEXIS could play an important role in integrated curriculum development.* (Hudeček)

Formats?

- Many informants prefer multimodal training materials, although some still prefer written step-by-step instructions

- Videos in training materials:
  - should be short rather than long;
  - should be carefully prepared to satisfy concrete learning objectives;
  - should be accompanied by full transcripts and/or subtitles;
  - **should not be served on their own but embedded and textually contextualized.**

There is more in the report

- Including a list of topics that our informants think should be covered in our training materials

- Including features of the training materials that our informants think are important

- Lots of evidence (as if we needed it) that we are all total dictionary nerds

WP5: Next Steps

- Digesting the results

- Working toward our next deliverable: *D5.2 Guidelines for Producing ELEXIS Tutorials and Instruction Manuals* (M24) A set of practical recommendations for developing tutorials and instruction manuals for ELEXIS tools and services that lower the barrier for participation in the infrastructure.

- More training measures...

The end is just a beginning...

Thanks for bearing with me.