The UPSKILLS project

Iulianna van der Lek
iulianna@clarin.eu

CLARIN ERIC

L-Università ta' Malta

Co-funded by the Erasmus+ Programme of the European Union

13 July 2022
TaLC Conference 2022, Limerick, Ireland
Overview of the UPSKILLS project by Iulianna van der Lek

Hands-on activity by Francesca Frontini
**UP**grading the **SKI**lls of **Linguistics** and **Language Students**

An Erasmus+ strategic partnership  
September 2020 - August 2023

<table>
<thead>
<tr>
<th>1 University of Malta (coordinator)</th>
<th>6 University of Rijeka</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 University of Belgrade</td>
<td></td>
</tr>
<tr>
<td>3 University of Bologna</td>
<td>and with funding from Movetia:</td>
</tr>
<tr>
<td>4 CLARIN ERIC</td>
<td>7 University of Geneva</td>
</tr>
<tr>
<td>5 University of Graz</td>
<td>8 University of Zurich</td>
</tr>
</tbody>
</table>
Aims

Why?
A gap between the knowledge and skills (typically) acquired at universities and those (commonly) required by the labour market

How?
Four intellectual outputs:
IO1 – Needs analysis
IO2 – Best practices for research-based teaching
IO3 – Learning content creation
IO4 – Incorporation of content into educational games

What?
(Further) identify the gap
Tackle the gap

When?
Completed
Ongoing
UPSKILLS Needs Analysis

**Overall objective:** Exploring the gap between the knowledge and skills (typically) acquired at universities and those (commonly) required by the labour market

**Six components:**
1. A survey of curricula
2. A literature review
3. A corpus-based analysis of job advertisements
4. A questionnaire for the language industry
5. Interviews with industry representatives
6. A new profile for language and linguistics students
Language Data and Project Specialist
- knowledge, skills and competences -

**DISCIPLINARY**
- Knowledge of specific languages
- Ability to conduct linguistic analysis
- Translation, interpreting, post-editing, localisation

**(INTER)CULTURAL**
- Awareness of specific cultural contexts and cultural differences
- Cultural agility

**TRANSVERSAL**
- Creative and innovative thinking
- Problem-solving skills
- Presentation skills
- Writing for different audiences

**DATA-ORIENTED**
- Ability to collect, manage, curate, analyse language data
- Knowledge of statistics
- Familiarity with data standards

**RESEARCH-ORIENTED**
- Knowledge of research design
- Analytical, logical and hypothetical thinking
- Accessing and processing information critically

**TECHNICAL**
- Understanding and use of language technologies and resources
- Understanding of computational linguistics / NLP
- Knowledge of a programming language

**ORGANISATIONAL**
- Entrepreneurship
- Project management
- Quality control
- Planning
- Teamwork

Illustration by Freepik Storyset, https://storyset.com/data
Language data and project specialist
- typical tasks and responsibilities -

DATA-ORIENTED

- Linguistic data collection
- Transcription of audio files
- Linguistic annotation
- Language data exploration and analysis

TECHNICAL

- Translation, interpreting, localisation, post-editing
- Development, analysis and testing of software and technological tools
- Development, testing and improvement of machine learning models

RESEARCH-ORIENTED

- Language data research
- Research on business processes and market needs

ORGANISATIONAL

- Communication with teams, clients and/or vendors
- Project management
- Process evaluation
Needs Analysis Reports

1.1 Survey of curricula: Linguistics and language-related degrees in Europe

1.2 Toward a new profile for twenty-first century language specialists: Industry, institutional and academic insights

1.3 Competences, skills and tasks in today’s jobs for linguists: Evidence from a corpus of job advertisements

1.4 Survey of business sectors hiring linguists and language professionals

1.5 Graduate skills and employability: Focus interviews with selected job market stakeholders

1.6 Language data and project specialist: A new modular profile for graduates in language-related disciplines
UPSKILLS Learning Content

COMING FALL 2022!

3-6 ECTS block units in Moodle with mix-and-match activities

RESEARCH SKILLS
Introduction to scientific research
Analytical thinking and problem solving
Project management

DATA ACQUISITION SKILLS
Introduction to programming
Text and speech processing
Collecting data from human participants

DATA HANDLING SKILLS
Introduction to machine learning
Data standards and repositories
Language data science

CROSS-CUTTING COMPONENTS
Linguistic theory
Research management software

Illustration by Freepik Storyset, https://storyset.com/education
UPSKILLS Learning Content

Accessing the materials:

- All the content will be available on Moodle: https://upskills.fil.bg.ac.rs/
- You will be able to export and import the courses
- To request a guest account, please contact Nađa Soldatić at nadja@fil.bg.ac.rs
Main Expected Results

- Prepare students for the reality of the job market
- Sensitise academics with respect to what skills employers are looking for
- Raise awareness among employers about the skills and aptitude of graduates of linguistics and language-related degrees
- Create engaging modular learning content that will be freely accessible to everyone
- Promote active- task- and research-based learning
Dissemination Events

Dissemination events planned for each intellectual output:

○ **Bologna ✓, Graz ✓, Malta ✓**

○ **Save the Dates:**
  - *Integrating Research and Research Infrastructures into Teaching*, Utrecht, 4 November 2022
  - Summer School in Serbia - July 2023
Today

- Course demo: *Introduction to Language Data: Standards and Repositories*.
- Hands-on activity: [Lesson 3.1 - How to Annotate Text Collections in WebLicht](#)
Introduction to Language Data: Standards and Repositories

**Workload:** 4-5 ECTS (estimation, needs to be piloted)

**Target audience:** BA, but also suitable for MA

By the end of this unit block, **learners will be able to:**

- Explain the main concepts related to research infrastructures for language resources and technology and the role they play in the research data lifecycle in the context of Open Science and FAIR
- Use certified research data repositories to search, find and access language resources and datasets
- Process, annotate, and analyse different types of corpora in online environments according to standards and formats used by the community
- Archive, share and cite language resources
Introduction to Language Data: Standards and Repositories

**Unit 1** Introduction to Research Infrastructures of Language Resources and Technology

**Unit 2** Finding, Accessing and Using Language Resources

**Unit 3** Linguistic Processing, Annotation and Analysis

**Unit 4** Archiving and Sharing Language Resources
Introduction to Language Data: Standards and Repositories

A few lessons (still under review, need piloting):

- Lesson 2 - Intro to Research Infrastructures for Language Resources and Technology
- Lesson 2.2 - Finding Language Resources in the CLARIN Repositories
- Lesson 2.3 - How to Search within Collections of Resources
- Lesson 2.4 - How to Collect and Cite Language Resources from Data Catalogues
- Lesson 3.1 - How to Annotate Text Collections in WebLicht -> demo
Feedback and Discussion

How does UPSKILLS and the piloted lessons fit into your curriculum?

● What did you find useful?
● How would you adapt the lessons?
Stay tuned

https://upskillsproject.eu
Sign up to our mailing list!

https://twitter.com/UpskillsProject
https://www.facebook.com/upskillsproject