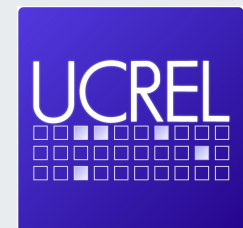# What can you do with the CLARIN research infrastructure?

@ Corpus Linguistics 2023 conference 2nd July 2023

Prof Paul Rayson

Director of UCREL research centre and CLARIN ambassador

School of Computing and Communications

Lancaster University, UK

@perayson

@perayson@fediscience.org

THE QUEEN'S ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2015

UCREL

What is this workshop about?

- Focus on practical issues for how corpus linguists can benefit from the well-developed 'Common Language Resources and Technology Infrastructure' (CLARIN) for language as social and cultural data

- Overview of CLARIN infrastructure including language resources, knowledge infrastructure, participating consortia, resource families

- Aims:
  - Participants will learn how to access and search the existing corpora and tools in the CLARIN research infrastructure and how to embed or deposit their own resources and tools in one of the centres of the infrastructure (with reference to annotation, formats and standards, metadata, licencing and documentation).

- Showcase key resources and tools such as those from the ParlaMint project

Workshop team

- Paul Rayson and Martin Wynne plus Dario Del Fante (in person)

- Darja Fišer, Francesca Frontini (co-organisers not attending)

Schedule for the workshop …

Time	Activity
09:30	Welcome, introductions, and aims for the workshop (Paul Rayson)
09:45	Introduction to CLARIN and Resource Families (Martin Wynne)
10:15	Introduction to tutorial on ParlaMint project (Dario Del Fante)
10:30	Break
10:45	Hands-on practical with ParlaMint (Dario Del Fante)
11:15	A menu to choose from hands-on practical worksheets … to explore:
	Virtual Language Observatory https://vlo.clarin.eu/
	Switchboard https://switchboard.clarin.eu/
	Federated content search https://contentsearch.clarin.eu/
12:15	Wrap up, final questions and feedback
12:30	End

Things you will need (and not need) …

- A login to the lab computers
  - Either from the registration desk or from us now
- A web browser for the practical elements
- Follow the workshop link to find all the hands-on materials
- No prior programming experience is assumed and we will guide you through how to use the resources

Any questions before we start?

- Link to slides and materials:

- Relevant websites:
  - https://www.clarin.eu/
  - https://www.clarin.eu/resource-families
  - https://www.clarin.eu/parlamint

- Contact:
  - Email: p.rayson@lancaster.ac.uk
  - Twitter: @perayson
  - Mastodon: @perayson@fediscience.org

Schedule for the workshop …

Time	Activity
09:30	Welcome, introductions, and aims for the workshop (Paul Rayson)
09:45	Introduction to CLARIN and Resource Families (Martin Wynne)
10:15	Introduction to tutorial on ParlaMint project (Dario Del Fante)
10:30	Break
11:15	A menu to choose from hands-on practical worksheets … to explore:
	Virtual Language Observatory https://vlo.clarin.eu/
	Switchboard https://switchboard.clarin.eu/
	Federated content search https://contentsearch.clarin.eu/
12:15	Wrap up, final questions and feedback

Materials to explore

1. SSHOC handouts
   - Use Case 1 Gender in Parliamentary Discourse
   - Use Case 2 Creating a linguistically annotated corpus of 19th century English novels
   - Yankelevich, Tanya, Fiser, Darja, Lenardic, Jakob, Gorgaini, Elisa, & Braukmann, Ricarda. (2020, June 24). LIBER 2020 - Workshop: SSHOC Train-the-Trainer Bootcamp for Librarians. Zenodo. https://doi.org/10.5281/zenodo.3970799
2. Handout: "How to search for linguistic patterns across corpora" by Iulianna van der Lek and Martin Wynne
3. Using CLARIN Resources for Corpus Linguistics
   - https://app.lumi.education/flow/64919efb1269a946079f1677
   - Need to register on moodle, course developing until release in August
   - Lek, I. & Fišer, D. (2023). "Introduction to Language Data: Standards and Repositories". In UPSKILLS Learning Content. August 2023 (forthcoming). https://upskillsproject.eu/project/standards_repositories/