

## Training in Corpus Linguistics

6 days

6 sessions á 4 hours: 2 hours theory/introduction + 2 hours training

4 topics (can be visited independently)

Training language is Bosnian

- |         |       |  |
|---------|-------|--|
| 31. May | 10-14 | Introduction CL and CoLABnet 4h <ul style="list-style-type: none"><li>• CoLABnet and Spoken Corpus</li><li>• What is a corpus</li><li>• What kinds of corpora are there</li><li>• What is corpus linguistics</li><li>• Applications for corpus linguistics</li><li>• How to get access to BCMS corpora</li><li>• Corpus Search</li></ul> |
| 1. June | 10-14 | Organization of a corpus (Sampling, Formatting): Intro + Hands-on training (App) 4h <ul style="list-style-type: none"><li>• Sampling and Representativeness</li><li>• Formatting</li></ul>   |
| 2. June | 14-18 | Annotation and Metadata: Rules and Tools 4h <ul style="list-style-type: none"><li>• Annotation schemas, Tagsets, Tagger</li><li>• Metadata in general and for spoken copora + CoLABnet</li></ul>   |
| 5. June | 10-14 | Transcription: Rules and Tools 4h <ul style="list-style-type: none"><li>• Tools for transcription (Exmaralda / ELAN)</li><li>• Transcription rules</li><li>• Training: Installation and use of tool</li></ul>  |
| 6. June | 10-14 | Transcription: Hands-on training 4h <ul style="list-style-type: none"><li>• Transcription of samples</li></ul>   |
| 7. June | 10-14 | Transcription: Putting the data together (+ metadata) 4h <ul style="list-style-type: none"><li>• Formatting transcripts</li><li>• Adding metadata</li><li>• Adding ORCID</li></ul>   |