

Teaching with CLARIN

Moderators:
Iulianna van der Lek and Francesca Frontini

CLARIN Annual Conference
Tuesday, 28 September 2021
13:30 - 14:30



Schedule

Intro

Three-Minute Presentations

Teaching with CLARIN Award

Next Steps

Q&A

[#CLARIN2021](#)

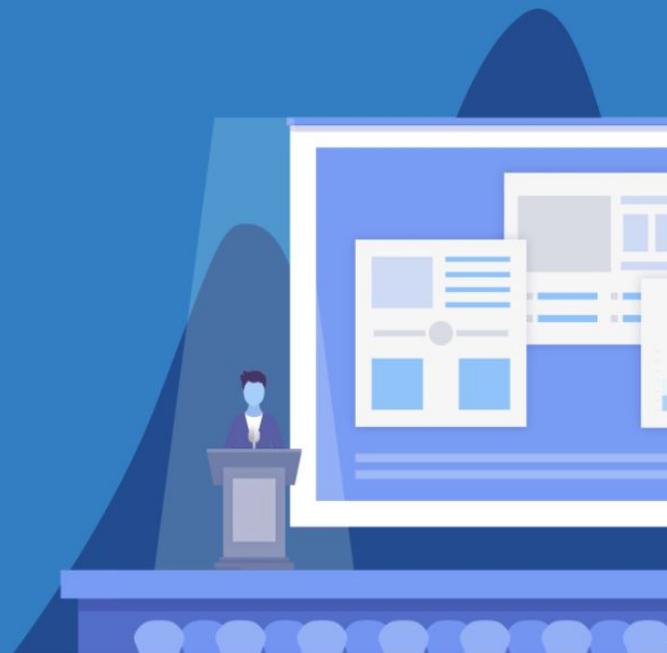
CLARIN Training and Education Strategy

- Better visibility of the training and educational materials & activities
- A sustainable network of CLARIN trainers
- Funding instruments not only for research, but also for teaching
- Support the integration of CLARIN into the curricula

UPSKILLS Project

UPgrading the SKIlls of LInguistics and LAnguage Students

The UPSKILLS project is an Erasmus+ strategic partnership for higher education that seeks to identify and tackle the gaps and mismatches in skills for linguistics and language students through the development of a new curriculum component and supporting materials to be embedded in existing programmes of study.



<https://upskillsproject.eu/>

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



TECHNICAL

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

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

ORGANISATIONAL

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

UPSKILLS - First Deliverables

Needs Analysis

[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](#)

"Teaching CLARIN in Times of Corona" call

- Launched last winter to **acknowledge** and **showcase** the efforts made by teachers, lecturers and trainers throughout the CLARIN network to meet the education needs during the Covid-19 pandemic.
 - 9 submissions received from 6 different countries
- 3 awards (2 by jury, one by public vote)
 - 350 euros each!
- After the presentation of these materials, the award ceremony will take place.

Three-minute Presentations

| | |
|---|---|
| 1. Applied Language Technology | 6. Introduction to Speech Analysis |
| 2. Archilochus of Paros: Elegiac Fragments - XML Archive | 7. Oral Archives for Sociolinguistic Research |
| 3. Computational Morphology with HFST | 8. Privacy by Design in Research |
| 4. GATE, an Open-Source Toolkit for Natural Language Processing | 9. Voices of the Parliament |
| 5. Introduction to Digital Humanities | |

Applied Language Technology

Tuomo Hiippala

Faculty of Arts, University of Helsinki, Finland

<https://www.clarin.eu/content/applied-language-technology>



Applied Language Technology

What? A two-course module (10 ECTS) for teaching the basics of language technology and its applications – no prerequisites

Why? To show that language technology is accessible and applicable

How? Layperson explanations; gradual buildup of vocabulary – interactive programming; explanation videos; openly available learning materials

What next? A massive online open course

Core libraries and resources



Learning materials



Archilochus of Paros: Elegiac Fragments - XML Archive

Anika Nicolosi and Beatrice Nava

University of Parma, Italy

<https://www.clarin.eu/content/archilochus-paros-elegiac-fragments-xml-archive>



Experience



- practical skills in XML/TEI coding of ancient literary texts: the poems of Archilochus of Paros (VII century B.C.)
- an overview of the state of the art in the field of scientific and digital critical editions
- test the potential of digital philology (text and critical apparatus)

Computational Morphology with HFST

Erik Axelson

Faculty of Arts, University of Helsinki, Finland

<https://www.clarin.eu/content/computational-morphology-hfst>



Background

An adaptation of the teaching material of the course “Computational Morphology” taught by **Mathias Creutz** in the Master’s programme “Linguistic Diversity and Digital Humanities” at the University of Helsinki.

The course and its exercises have been created by Mathias Creutz and co-developed by **Senka Drobac**. The examples and exercises of the course use command line tools of **HFST** toolkit.

- **HFST – Helsinki Finite-State Technology** (<https://hfst.github.io/>): a toolkit intended for processing natural language morphologies. Can be used via command line tools or a Python API.

The web course

Using an earlier version of the course material, we have developed an interactive version implemented as **Jupyter Notebooks**. It uses the same examples and exercises, but HFST command line tools have been replaced with HFST Python interface.

- **Jupyter Notebooks** (<https://jupyter.org/>): an interactive computational environment combining code execution, rich text, mathematics, plots and rich media

About the course

- Offered as part of FIN-CLARIN / SAFMORIL (CLARIN Knowledge Centre for Systems and Frameworks for Morphologically Rich Languages) resources.
- **Self-study course** containing seven lectures, tutorial-type lectures with assignments at the end.
- Prerequisites: foundations of general linguistics, some programming experience is desirable, knowledge of Natural Language Processing (NLP) is also a plus.
- The course demonstrates **how HFST tools can be used for generating finite-state morphologies.**
- Suitable as a complement to a more theory or linguistics oriented course on morphology.

Goals of the course

After successfully completing the course, you

- understand basic theory on finite-state automata and transducers,
- are able to design morphological lexica using finite-state technology,
- know how to write morpho-phonological rules in a finite-state framework,
- understand the diversity of morphological structure in different languages
- and know how to take these differences into account when designing computational models of morphology.

Accessing the course

- Running it on your computer: get source code from Github repo, install Jupyter software, Python version 3, Python packages *hfst-dev* and *graphviz*.
- The course is accessible also at CSC (Finnish IT Center for Science) Notebooks service: <https://notebooks.csc.fi>.
- Logging to the service requires HAKA (authentication federation of Finnish Universities) or CSC account or visitor account via CSC/SAFMORIL helpdesk plus a join code for registering.
- We are working on getting the course more easily accessible.

Screenshots 1/2

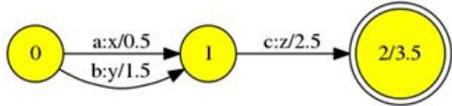
jupyter Part1 Last Checkpoint: 14 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Help Trusted Python 3

Run

6.4. Weighted finite-state transducer (WFST)

A weighted finite-state transducer (WFST) is a finite automaton for which each transition has an input label, an output label, and a weight.

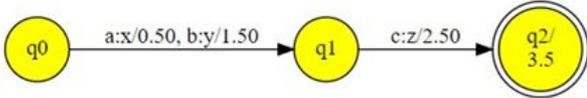


```
graph LR; 0((0)) -- "a:x/0.5" --> 1((1)); 0 -- "b:y/1.5" --> 1; 1 -- "c:z/2.5" --> 2(((2/3.5)))
```

The initial state is labeled 0. The final state is 2 with final weight of 3.5. Any state with non-infinite final weight is a final state. There is a transition from state 0 to 1 with input label "a", output label "x", and weight 0.5. This machine transduces, for instance, the string "ac" to "xz" with weight 6.5 (the sum of the arc and final weights).

```
In [9]: tr = HfstIterableTransducer()
tr.add_transition(0, 1, 'a', 'x', 0.5)
tr.add_transition(0, 1, 'b', 'y', 1.5)
tr.add_transition(1, 2, 'c', 'z', 2.5)
tr.set_final_weight(2, 3.5)
tr.view()
```

Out[9]:



```
graph LR; q0((q0)) -- "a:x/0.50, b:y/1.50" --> q1((q1)); q1 -- "c:z/2.50" --> q2(((q2/3.5)))
```

```
In [10]: # Convert to HfstTransducer before Lookup.
from hfst_dev import HfstTransducer
TR = HfstTransducer(tr)
print(TR.lookup('ac'))

(('xz', 6.5),)
```

Theory part

Run example code

Visualize the result

Screenshots 2/2

Assignment

Write your own solution

Modify the code to make it work

Open questions to think about

The screenshot shows a Jupyter Notebook interface with the following content:

- 7. Assignments**
- Assignment 5.1: Analysis of vocabularies**
- The HFST project has produced open-source morphological analyzers for some languages. You can download them from [Kielipankki \(The Language Bank of Finland\) pages](#).
- a.** Decide what language you want to work on. You can choose one of the following analyzers: [English](#), [Finnish](#), [French](#), [German](#), [Swedish](#), or [Turkish](#).
- Download the zip package, extract it, and search for file with extension `.hfst`. This is the analyzer file.
- b.** Read the transducer from file and test it with `HfstTransducer.read_from_file` and `HfstTransducer.lookup`. Type in a few word forms and check what analyses you get for these word forms (the format of the tags is slightly different in the different languages).
- In []: `pass # <write your solution here>`
- Perform the same test in an interactive xfst shell (open it with `start_xfst`). Use commands `load stack` and `lookup` instead.
- In []: `pass # <write your solution here>`
- c.** Pick a text file containing words of the language you chose. There should be one word per line in the file. If you don't have any text file available, you can use one of the pre-prepared files `english-words.txt`, `finnish-words.txt`, `french-words.txt`, `german-words.txt` or `swedish-words.txt`. For reading, you can use Python's `open(path,mode)` and `<file>.read()`, `<file>.readline()`, or `<file>.readlines()`.
- d.** Run the entire text file through the analyzer, as follows:
- In []:

```
# <you must edit the code so that it will work for the Language you have chosen>
for line in some_text_file:
    some_analyzer.lookup(line)
```
- e.** Browse through the analyses of the words. Find a few words that are out-of-vocabulary, that is, the analyzer does not recognize them and cannot provide any analysis. Why aren't these words covered in the vocabulary of the analyzer?
- Assignment 5.2: Creating a translating guesser**
- Your task is to create a guesser that translates words between related languages or dialects. For instance, in the attached file

GATE, an Open-Source Toolkit for Natural Language Processing

Diana Maynard

Faculty of Engineering, University of Sheffield, UK

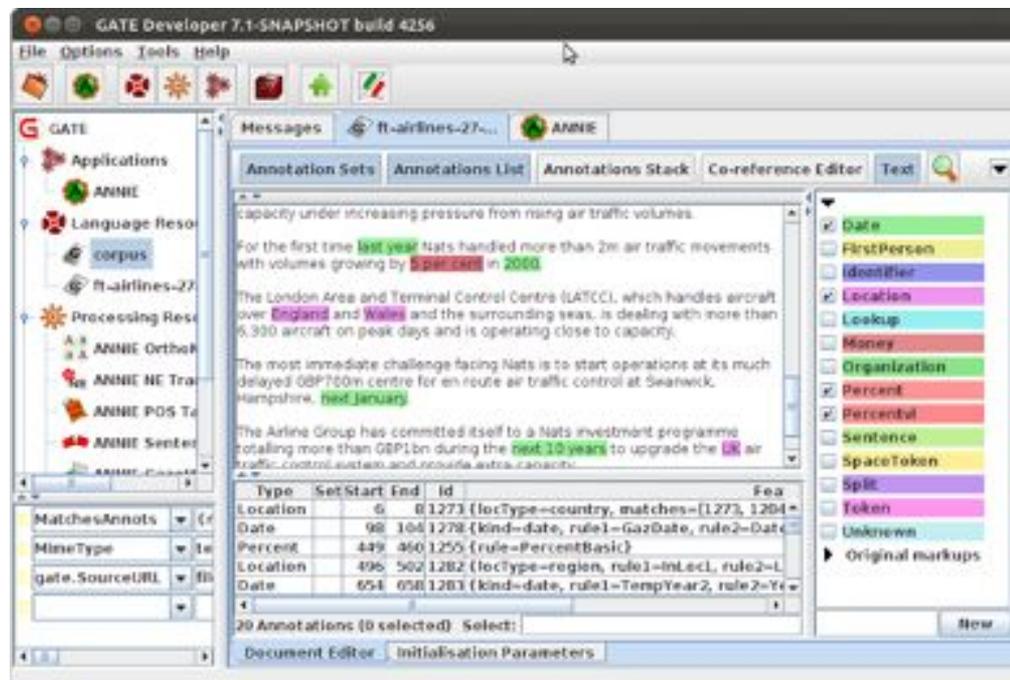
<https://www.clarin.eu/content/gate-training-course>



What is GATE?



- Freely available & open-source
- Over 20 years of development
- Plugins for both high and low-level NLP components
- Covers 25 languages
- 100 ready-made applications



Aims of the course

- Provide an **introduction to NLP** for those new to the area
- Show people how to perform NLP tasks for a wide variety of domains and applications
 - **Basic NLP tasks:** tokenisation, POS tagging, morphological analysis...
 - **Complex applications:** sentiment analysis; hate speech; misinformation...
- Provide instruction on how to use the GATE toolkit and applications in an **off-the-shelf** fashion for a variety of NLP tasks
- Enable people to **adapt and develop GATE tools** themselves for other tasks, domains, applications and languages
- Assist **software developers and programmers** to use GATE programmatically and to integrate it with other tools and systems. e.g. other CLARIN resources

Who is the course for?



Suitable for researchers from any discipline and skill level!

- **Programmers** who want to integrate it in their own software or extend its functionality (e.g. add new plugins)
- **Linguists** who want to develop new tools without needing programming expertise
- **Social scientists and humanities researchers** who want to use off-the-shelf tools to process their data
- **Anyone** who wants to learn about NLP in a simple hands-on way

12 Course Modules

can be mixed and matched as needed



Basic Content

Intro to GATE + NLP

JAPE – Write your own simple rules

GATE Cloud + Google Sheets apps

GATE Plugins

Applications

Social media analysis

Sentiment Analysis

Analysing Disinformation

Analysing Online Abuse

Specialist content

GATE for Programmers

GATE, NLP and Python

Machine Learning

Code Clinic

Format



- Can be taught as **in-person**, **online** or **self-study** materials
- Developed and taught by a team of **topic specialists**
- **12 mix-and-match** 3-hour modules
- Mixture of information, explanation, and exercises (with hints and answers)
- **Discussion** points throughout
- Further exercises for those who want to explore in more depth
- All tools, slides and corpora are **available for download** for free
- Course **videos** available on-demand
- **Personalised** courses available on-demand

Example - processing HTML

Test this pipeline

Type the content to
annotate:

```
<html>
<head>
<script language=JavaScript>
document.cookie='FTSection=hp/homepg;domain=.ft.com;path=/' ;document.cookie= 'FTPage=0hpart;domain=.ft.co
m;path=/' ;document.cookie='FTReferrer='+escape(document.referrer)+' ;domain=.ft.com;path=/track/';document.co
okie='FTURL='+escape(document.URL)+' ;domain=.ft.com;path=/track/';
</script>
```

Load an HTML file

Or select a text file:

Choose file ftairlines-27-jul-2001.html

Output type:

JSON

Set the format

Document format:

HTML

Customize annotations

Test Pipeline

download

Text parsed out of
the HTML

Annotation types: Date Location Organization

FT.com | TotalSearch | Global Archive | Print Return to Article | Print this Page Airlines take over running of air traffic control
FT.com site, Jul 27, 2001 BY KEVIN DONE, AEROSPACE CORRESPONDENT Seven UK airlines including British Airways,
Virgin Atlantic, BMI British Midland and EasyJet, on Friday took over control of the air traffic control system, completing
one of the government's most controversial public-private partnership deals. Completion of the National Air Traffic Services deal
comes at a critical time for the government as it tries to push through the PPP for the London Underground. The sale to a
strategic investor of a 46 per cent stake in Nats is the first time in Europe that management control of en route air traffic

Course usage

- Course has been running for **over 12 years** and constantly updated and improved
- COVID-19 led to the development and implementation of the **online taught course** in 2021
- Elements of the course are frequently used for separate specialised / smaller courses
- Elements have been used by others (with permission) in their own student courses

Further information

<https://gate.ac.uk/family/training.html>

Introduction to Digital Humanities

Zuzana Neverilova

Faculty of Arts, Masaryk University, Czech Republic

<https://www.clarin.eu/content/introduction-digital-humanities>



A practical course about DH



- presentation slides
- videos
- commented Jupyter Notebooks
- install and try a software
- online discussions

I never tried OCR on old digitized books before. I was surprised how easily we can get the results...

Introduction to Speech Analysis

Mietta Lennes

Faculty of Humanities, University of Helsinki, Finland

<https://www.clarin.eu/content/introduction-speech-analysis>



Corpus Linguistics and Statistical Methods (5 cr)

Word attributes

- baseform: kieli
- baseform (compound boundaries): kieli
- part-of-speech: noun
- msd: NUM_Sg/CASE_Nom
- dependency relatin: nominal subject
- word position in a name: outside (0)
- original thread
- original message
- Show Dependency Tree
- Näytä dependenssipuu



Introduction to Speech Analysis (5 cr)

tänne sun huon
no niin varmaa
ie sa neuvoo
isään roholi:
itä nä
pieninkatu e.
is sä tuut ni
o kei: n o o
ilvä



Translated in 2020

Data Clinic (5 cr)

| 1 | match | sem | job | dephased | med | hmm | lex | ref | deprel |
|---|-------|-----|-----|----------|-----------------|------------------|---------|----------------|--------|
| 2 | C | 0 | 0 | 2 | SUBCAT_Nu | (NUM_Sg/CASE_Nom | kuusi | [kuusi_0n_1] | 2:041 |
| 3 | C | 0 | 1 | 1 | NUM_P/CASE_Anom | | viisi | [viisi_0n_1] | 7:048 |
| 4 | C | 0 | 2 | 0 | POS_Ng | (VERB_Aut) | puotta | [puotta_0n_1] | 8:020 |
| 5 | C | 0 | 3 | 3 | NUM_Ng | (CASE_0n) | viisita | [viisita_0n_1] | 4:042 |
| 6 | C | 0 | 4 | 4 | | | | | 0 |
| 7 | C | 0 | 5 | 4 | SUBCAT_OC | | | | 0 |
| 8 | C | 0 | 6 | 5 | SUBCAT_Dom | (NUM_Sg | | | 0 |
| 9 | C | 0 | 7 | 6 | POS_Sg | (VERB_Aut) | | | 0 |



Reasons for translating the course materials

1. Speech analysis methods and workflows can be applied in many fields
2. Similar courses needed for international vs. Finnish-speaking students
3. Course already open to all universities
4. Reasonable amount of text, no restricted material
5. “well-established” content, used in online teaching since 2002 or so...
6. **Persistent access to speech resources available via CLARIN!**
7. The main tool used in the course is [Praat](http://www.praat.org/)* – open, well maintained and can be installed almost anywhere.

*Boersma, Paul & Weenink, David (2021). *Praat: doing phonetics by computer* [Computer program]. Available from <http://www.praat.org/>

The live collection is now available via Library of Open Educational Resources

<http://urn.fi/urn:nbn:fi:lb-2021063021>



Puheen analyysin perusteet - Introduction to Speech Analysis



Mietta Lennes

Created: 30.6.2021 17:51 Published: 30.6.2021 19:31

Edited: 1.7.2021 18:51

Tämä kurssi tarjoaa yleiskuvan puheaineistojen käsittelystä ja puheen akfoneettisista analyysimenetelmistä. Kurssi koostuu kuudesta oppitunnin joiden aikana tutustutaan erityisesti Praat-puheanalysohjelman käyttömahdollisuuksiin ja opitaan soveltamaan ohjelman ominaisuuksia joustavasti omassa työssä. Kurssilla opetellaan käyttämään myös ELAN-ohjelmaa, jolla voi litteroida ja annotoida äänen lisäksi videoita. — The course offers a general picture of managing speech corpora and of methods that are available for the acoustic-phonetic study of speech. The course consists of six lectures during which you will use a speech analysis program called Praat and learn to apply the main features of the program to your work with speech recordings. In addition, you will learn the basics of another program called ELAN that can be used for transcribing and annotating as well as video material.

Keywords

fonetiikka kokeellinen fonetiikka akustinen fonetiikka perustaaajuus

intonaatio (puhe) tutkimusmenetelmät äänteet puheentuotto

artikulaatio (puhe) kielentutkimus

Route to A wing Corpus, Downloadable Version

Reitti A-siipeen -korpuksen ladattava versio
reittidemo-dl

Persistent Identifier of this resource: <http://urn.fi/urn:nbn:fi:lb-2021063021>

Access location: <http://urn.fi/urn:nbn:fi:lb-2021063021>

This corpus is available via the download service in Kielipankki - the Language Store of Finland.

This corpus also includes the same content that was originally available via the LAT system (Lähtökielinen) (see Relations). Further details about the most recent version are available on the info page.

Choose language:
fi en

Annotating with ELAN



Tällä tavalla on kätevä litteroida esimerkiksi pitempää pätkää kerralla

Choose a language:



Choose a part of the learning resource:

Lesson 3: Exercises

How does pitch analysis work?

Creating pictures

Lesson 3: Exercises

Learning goals

- Fundamentals of speech production (recap): phonation and articulation
- Various ways of taking measurements from sound in Praat
- Visualizing and measuring fundamental frequency / pitch
- Drawing and plotting in the Picture window
- Praat scripts: History

Exercises

1. Read/Watch the material *How does pitch analysis work?*
2. Make sure that you know how to find the internal manual pages (**Help**) about acoustic analyses in Praat (see, e.g., the keywords **Pitch, Spectrum, Spectrogram, Intensity**). In Praat, the manual can be accessed in almost all windows under the Help menu or button. Take a look at the manual and/or Praat website and try to find out what kinds of acoustic analyses are available in Praat.
3. In Praat, open a sound file and the corresponding TextGrid file – preferably the same files you worked with in the previous lesson. Open a TextGrid editor with the sound and the TextGrid. Switch on the spectrogram, the pitch curve, and the intensity curve. Switch off the formant analysis in case it is currently displayed.
4. If required, adjust the pitch analysis settings in the editor window, in order to make them more appropriate for the speaker in question (**Pitch:Pitch settings...**).
5. In the TextGrid editor, change the displayed scale of the pitch curve so that the peaks and troughs are clearly visible (**Pitch:Advanced pitch settings...**).

Static download package of the same content on Zenodo, DOI:[10.5281/zenodo.5506969](https://doi.org/10.5281/zenodo.5506969)

zenodo

September 14, 2021

Lesson **Open Access**

Puheen analyysin perusteet – Introduction to Speech Analysis

 Lennes, Mieta

This course offers a general picture of managing speech corpora and of methods that are available for the acoustic-phonetic study of speech. The course consists of six lectures during which you will use a speech analysis program called Praat and learn to apply the main features of the program to your own work with speech recordings. In addition, you will learn the basics of another program called ELAN that can be used for transcribing and annotating audio as well as video material.

The material has been translated and can be used both in English and in Finnish.

[Preview](#)

| Files (193.5 MB) | |
|---|---|
| Name | Size |
| oppitunti-1_lesson-1.zip | 5.8 MB <input type="button" value="Preview"/> <input type="button" value="Download"/> |
| md5:afabdac1b39e88b5fd80f25e636d1c02  | |
| oppitunti-2_lesson-2.zip | 355.4 kB <input type="button" value="Preview"/> <input type="button" value="Download"/> |
| md5:ace03621369a2005adea48e9b75b737e  | |
| oppitunti-3_lesson-3.zip | 871.9 kB <input type="button" value="Preview"/> <input type="button" value="Download"/> |

12

 views

18

 downloads

[See more details...](#)

Indexed in

OpenAIRE

Publication date:

September 14, 2021

DOI:

[DOI 10.5281/zenodo.5506969](https://doi.org/10.5281/zenodo.5506969)

Related identifiers:

Published in
<http://urn.fi/urn:nbn:fi:ilb-2021063021> (Lesson)

References
<http://urn.fi/urn:nbn:fi:ilb-2020112929>
(Video/Audio)
[10.5281/zenodo.376094](https://doi.org/10.5281/zenodo.376094)

License (for files):

 Creative Commons Attribution 4.0 International



KIELIPANKKI

The Language Bank of Finland

<https://www.kielipankki.fi/training-materials/>



LANGUAGE BANK ACCESS CORPORA TOOLS ORGANIZATION SUPPORT

SUOMEKSI PÅ SVENSKA

Training materials

Suomeksi

Search:

| Author | Title | Description | Citation | Organization |
|-----------------|--|--|---|--------------|
| Axelsson, Erik | Computational Morphology with HFST | <p>The web course "Computational Morphology with HFST" is an adaptation of the teaching material of the course "Computational Morphology" taught by Mathias Creutz in the Master's programme "Linguistic Diversity and Digital Humanities" at the University of Helsinki. The original course and its exercises have been created by Mathias Creutz and co-developed by Senka Drobac. Using an earlier version of the course material, Erik Axelsson has developed a Jupyter notebook interactive version.</p> <p>The web course uses the same examples and exercises as the original course, but HFST command line tools have been replaced with HFST Python interface.</p> | Axelsson, Creutz, Drobac (2019). Computational Morphology with HFST – an adaptation of the teaching material of the course "Computational Morphology" taught by Mathias Creutz in the Master's programme "Linguistic Diversity and Digital Humanities" at the University of Helsinki. [Learning material]. Language bank of Finland. Accessible at http://urn.fi/urn:nbn:fi:lb-2021053001 | |
| Hiippala, Tuomo | Applied Language Technology | Together, these two courses provide an introduction to applied language technology for audiences who are unfamiliar with language technology and programming. The learning materials assume no previous knowledge of the Python programming language. | Hiippala, Tuomo (2021) Applied Language Technology: NLP for the Humanities. In David Jurgens, Varada Kolhatkar, Lucy Li, Margot Mieskes and Ted Pedersen (eds) Proceedings of the Fifth Workshop on Teaching NLP. Association for Computational Linguistics, 46–48. DOI: 10.18653/v1/2021.teachingnlp-1.5 | |
| Lenes, Mieta | Introduction to Speech Analysis (Puheen analyysin perusteet) | <p>This course offers a general picture of managing speech corpora and of methods that are available for the acoustic-phonetic study of speech. The course consists of six lessons during which you will use a speech analysis program called Praat and learn to apply the main features of the program to your own work with speech recordings. In addition, you will learn the basics of another program called ELAN that can be used for transcribing and annotating audio as well as video material.</p> <p>The materials include background readings, videos and exercises that can be used for self-study and completed independently. The content can be embedded into other courses, or students may choose to participate in one of the online instances of the Introduction to Speech Analysis course offered by FIN-CLARIN via the University of Helsinki (see Courses and training).</p> | <p>Current version in the <i>Library of Open Educational Resources</i>:</p> <p>Lenes, Mieta (2021) <i>Introduction to Speech Analysis (Puheen analyysin perusteet)</i> [Online course material]. Available at: http://urn.fi/urn:nbn:fi:lb-2021063021.</p> <p>Downloadable version:</p> <p>Lenes, Mieta. (2021, September 14). Puheen analyysin perusteet – Introduction to Speech Analysis. Zenodo. https://doi.org/10.5281/zeno-66.5506969</p> | |

General issues to solve

- How to maintain and use the content in a flexible way?
- Live access for real-time learning in different contexts
- Stable releases for citation and future reference (PIDs, long-term storage)
- Access management for teachers-only content
- How to share quizzes, exams and assessment criteria?
- Potential for sharing terminologies across courses

Oral Archives for Sociolinguistic Research

Silvia Calamai and Rosalba Nodari

Faculty of Languages for Intercultural and Business Communication,
University of Siena, Italy

<https://www.clarin.eu/content/oral-archives-sociolinguistic-research>





Oral archives for sociolinguistic research

Silvia Calamai, Rosalba Nodari
Siena University

Privacy by Design in Research

How You Can Do a Data Protection Impact
Assessment for an Innovative Research Scenario
Involving Speech Data

Esther Hoorn and Henk van den Heuvel

University of Groningen, the Netherlands

<https://www.clarin.eu/content/privacy-design-research>

<https://delad.ruhosting.nl/wordpress/dpia-role-play-with-video/>





university of
 groningen

Voices of the Parliament: A Corpus Approach to Parliamentary Discourse Research

Darja Fiser and Kristina Pahor de Maiti

Faculty of Arts, University of Ljubljana, Slovenia

<https://www.clarin.eu/content/voices-parliament-corpus-approach-parliamentary-discourse-research>



STRUCTURE & CONTENT

THEORY

- Corpus linguistics
- Parliamentary records
- Gender & language



HANDS-ON

- MPs' production
- Topics
- Women-related issues
- Subcorpora
- Frequency lists
- Keywords
- Collocations
- Manual annotation
- siParl 3.0 corpus
- NoSketch Engine



TABLE OF CONTENTS

[Training Materials](#)[Related Topics](#)[Home](#) / [Learn & Exchange](#) / [Teaching with CLARIN](#)

Teaching with CLARIN

Welcome to our new CLARIN platform, which encourages the integration of CLARIN resources, tools and services into the curricula of SSH-related disciplines. Here, teachers and lecturers are able to publish training materials in open access, share best practices in teaching, and reach out to those who would like to reuse the work prepared by their colleagues.

With many courses having moved online, the richness and robustness of the CLARIN's distributed virtual infrastructure have been a lifeline for many lecturers, researchers and students. This section showcases some early examples of courses and teaching materials collected via the [Teaching CLARIN in Times of Corona](#) call. It aims to acknowledge and present the efforts made by teachers, lecturers and trainers throughout the CLARIN network, who have worked hard to meet educational needs, especially during the Covid-19 pandemic. The materials below can be downloaded from CLARIN repositories or GitHub and are suitable for reuse.

If you like to stay informed about new training materials or exchange your teaching experience, please [subscribe](#) to our dedicated mailing list for trainers.

<https://www.clarin.eu/content/teaching-clarin>

Teaching with CLARIN Award

Materials evaluated by jury from CLARIN ERIC

Criteria for evaluation:

- Accessibility
- Adaptability and Modularity
- Clarity, Comprehensibility, and Readability
- Content and Technical Accuracy
- Appropriateness and Relevance for CLARIN

The Jury Awards go to...

Introduction to Speech Analysis

Mietta Lennes

Faculty of Humanities, University of Helsinki, Finland



The Jury Awards go to...

Voices of the Parliament: A Corpus Approach to Parliamentary
Discourse Research

Darja Fiser and Kristina Pahor de Maiti

Faculty of Arts, University of Ljubljana, Slovenia



The Public Award goes to....

GATE, an Open-Source Toolkit for Natural Language Processing

Diana Maynard

Faculty of Engineering, University of Sheffield, UK



Next steps

Interviews

Teaching with CLARIN -> ongoing call (Sept-June)

Best practices guidelines

Inventory of reusable learning materials

Creation of new learning materials

Increase the [Trainers' Network](#)

[CLARIN-UPSKILLS Cafe](#): 1 Dec 2021, 14:00 - 16:00

Q&A

Stay in contact

Subscribe to our mailing list: training@lists.clarin.eu

Email at training@clarin.eu

[CLARIN Events](#)

[@CLARINERIC](#)