

# Building and maintaining online courses in digital research methods

Mietta Lennes  
University of Helsinki / FIN-CLARIN  
*[mietta.lennes@helsinki.fi](mailto:mietta.lennes@helsinki.fi)*



## 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 relation: nominal subject
- word position in a name: outside (O)
- original thread
- original message
- Show Dependency Tree
- Näytä dependenssipuu



## Introduction to Speech Analysis (5 cr)



(2020-)

## Data Clinic (5 cr)

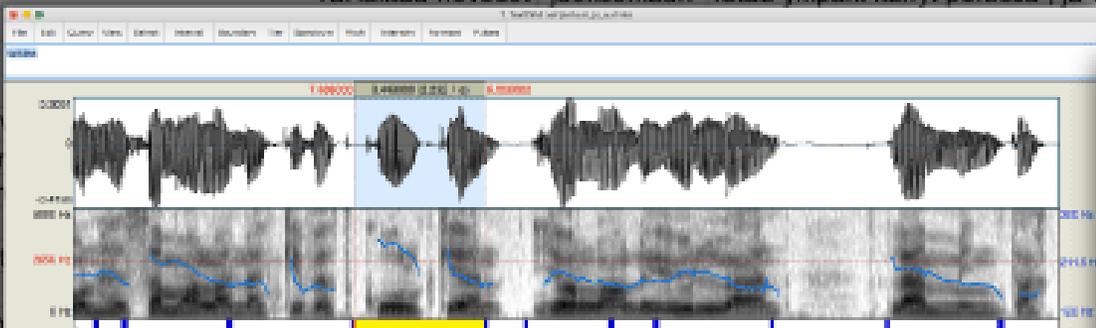
A	B	C	D	E	F	G	H	I
1	_match	_sen_tok	dephead	msd	lemma	lex	ref	deprel
2	0	0	0	2 SUBCAT_Rel NUM_Sg CASE_ine	mikä	[mikä_pn.1]		1 det
3	0	0	1	3 NUM_P CASE_Nom	eläin	[eläin_nm.1]		2 noun
4	0	0	2	0 PHS_Sg3 VOICE_Act TENSE_Prs MOOD_Ind	juosta	[juosta_vb.1]		3 ROOT
5	0	0	3	3 NUM_Sg CASE_Ess CMP_Pos	vapaa	[vapaa_la.1]		4 nommod
6	0	0	4	4				5 punct
7	0	0	5	4 SUBCAT_CC				6 cc
8	0	0	6	9 SUBCAT_Dem NUM_Sg				7 noun
9	0	0	7	9 PHS_Sg3 VOICE_Act TENSE_Prs MOOD_Ind				8 ROOT

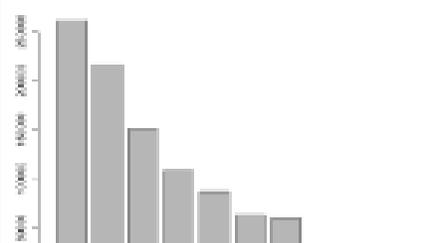


# Example course: Data Clinic (5 ECTS)

- **Support in language data management** for students who are starting their MA or PhD projects.
- During the course, each student composes a detailed *Data Management Plan (DMP)*.

	C	D	E	F	G		
1		left context	match	right context	left context lemmas	match lemmas	right context
2		Pentu saisi kodin maalta jossa on pihaa jossa	juosta	vapaana koko päivän :) ja peuhata koirakaverin ka	pentu saada koti maa joka olla piha joka	juosta	vapaa koko
3		Jos olisin silloin tarttunut sinuun , olisitko	juossut	karkuun , vai tarttunut myös minuun .	jos olla silloin tarttua sinä , olla	juosta	karkuun , ve
4		Tai laittaa hevoset	juoksemaan	rataa ympärit kämyt perässä , ja vielä ravia .	tai laittaa hevonen	juosta	rata ympärit
5	Meillä on ollu			ää takia .	olla olla ravuri pitkä aika ja se ei olla pakottaa	juosta	vaan se olla
6					ravuri ei saada	juosta	pakottaa !
7					heittää väite ja	juosta	karkuun kur
8					s 6 vuotias kun	juosta	into päin las
9	itykseni käy				vaan siksi että	juosta	juuri portat y
10	Älä ann				o elämä , vann	juosta	jeesus luo ja
11						juosta	se seinäjoki
12	ne kiertelevät				.Kaapelia pitkin	juosta	naapuri tam
13					kun se saada	juosta	ja merkkailla
14	perkkailia lurn				a se lähteä heti	juosta	.





# Stages of the Data Clinic

1. Contextualize your data; submit initial Data Management Plan
2. Organize and pre-process your material
3. Annotate and search
  1. text
  2. speech (audio and video)
4. Wrangle and visualize data
5. Discover suitable tools
  - meet a Data Advisor
6. Submit your final DMP

Data Management Plan

using metadata repositories (VLO, META-SHARE)

identifying variables and factors

file naming practices

plain text files

character encodings

using a text editor

regular expressions for searching text/annotations

using and understanding concordances

copyright issues

personal data issues

pseudonymization

understanding version control

recording systematic measurements or observations

annotating or parsing text

annotating speech in Praat

annotating speech in ELAN

importing tabular data to MS Excel

importing tabular data to R

filtering data

cleaning (not polishing) data

basic data visualization

exploratory visualization methods & tools

creating public metadata

etc...

applying statistical tests

# Another example course: Corpus Linguistics and Statistical Methods (5 ECTS)

The screenshot displays a corpus linguistics software interface. At the top, there is a list of text excerpts with the word 'kieli' highlighted. Below this, a detailed analysis window for the word 'kieli' is shown, listing various linguistic attributes such as baseform, compound boundaries, part-of-speech, and dependency relations. The interface also includes a dependency tree and a dependency network visualization.

**Word attributes**

- baseform: kieli
- baseform (compound boundaries): kieli
- part-of-speech: noun
- msd: NUM\_Sg|CASE\_Nom
- dependency relation: nominal subject
- word position in a name: outside (O)

[original thread](#)  
[original message](#)

[Show Dependency Tree](#)  
[Näytä dependenssipuu](#)

” After passing this course, you will understand what a corpus is and how corpora can be used for solving small questions concerning language use. You will be familiar with the fundamental terminology in corpus linguistics...”



# KIK-404. Corpus Linguistics and Statistical Methods (spring 2019)

[Dashboard](#) / [Courses](#) / [korpling-2019K](#) / [Week 2. Concordance and meaning](#) / [Week 2: Mystery words \(5 p.\)](#) / [Preview](#)

You can preview this quiz, but if this were a real attempt, you would be blocked because:

This quiz is not currently available

**Question 1**

Not complete

Marked out of 1

Flag question

Edit question

Below, you can see an excerpt of a concordance. The example sentences were picked out from a real corpus.

Look at the sentences and try to recognize the exact word form that is represented by the concordance. Drag the correct word to the answer box below the concordance.

**Concordance:**

1. He thought about his Mama and how \_\_\_ she was.
2. The butterflies visit all the \_\_\_ flowers, she said.
3. About twenty years ago or maybe twenty-five Faith sat in the late afternoon light of her front window reading an article all about maternal responsible its lot some \_\_\_ bad personality and character traits in children.
4. Her \_\_\_ face, seen like that , seemed featureless.
5. " Linda stared down at her \_\_\_ suede pumps.
6. She could be \_\_\_ but less so.
7. That must mean he 's \_\_\_ guilty , " I said.
8. I didn't know that before the law you are either guilty or not ; " \_\_\_ " has nothing to do with it.
9. Jonathan and his sister had held long worried telephone powwows about him -- until he surprised them by falling in love with a young woman in his building, a \_\_\_ divorcee with a six-year-old daughter.
10. You look very \_\_\_ , by the way.
11. I am a woman of looks, \_\_\_ , ugly or plain, still I am a woman.
12. I had a \_\_\_ good hunch we were going to get laid off that day - the day I realized that alcohol and tool buddies don't mix.
13. "It's \_\_\_ sad really, " Dave was saying , " here I am 35 years-old and all I own is what 's in my car.
14. "Yeah my brother and I used to be \_\_\_ close.
15. "You know I think you have a \_\_\_ nice bod, "I said.
16. For even drunk as I was, I knew Dave was not my brother, not a gentle lover, not someone had gotten \_\_\_ close to and would probably never see again.
17. "Must be \_\_\_ interesting, Ben, "my grandfather said.
18. "You look \_\_\_ good yourself."
19. They all had \_\_\_ names.
20. Yours is a \_\_\_ one, too.
21. "After all , most of my life I have had the company of \_\_\_ women."

The word form in the concordance is .

find	play	table	none	cry	pet	affect	letter	pretty	fabric	form	look	usually
body	home	news	plastic	know	figure	here	fish	paper	number	tabloid	kind	now
once												

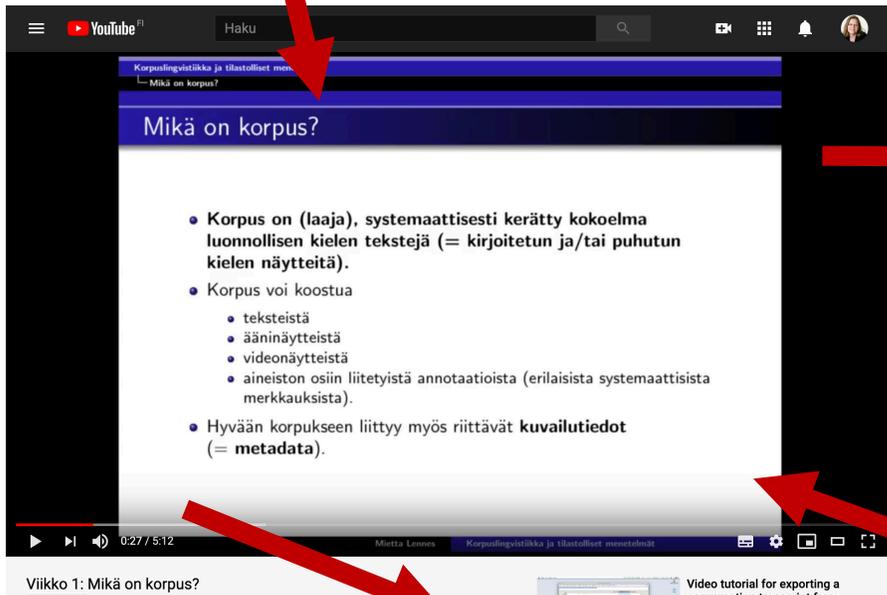
Quiz navigation

1 2

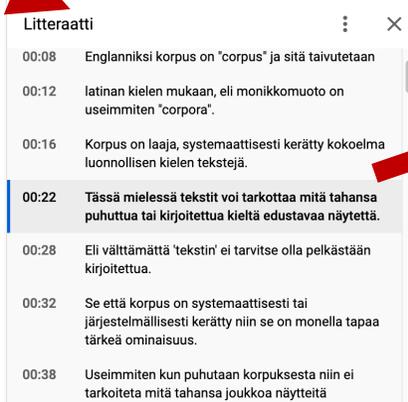
[Finish attempt](#)[Start a new attempt](#)

# Workflow for multilingual videos

1. Record and edit the original video, narration in Finnish, use parallel video tracks for slides in Finnish/English (hiding the one not in use)

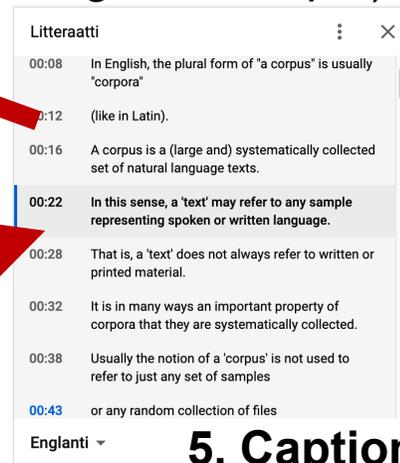


2. Finnish captions created locally (or can be done on YouTube)



3. Upload to YouTube

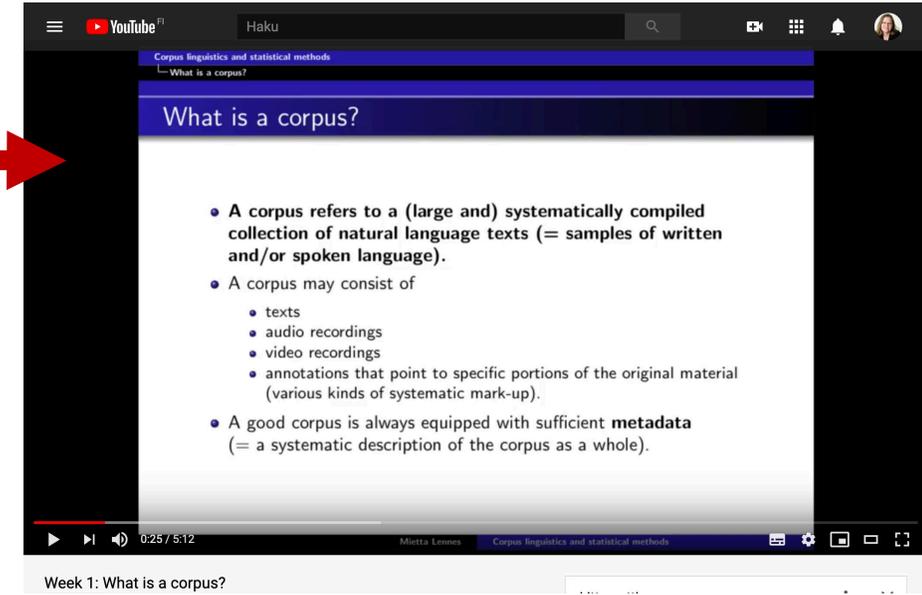
4. Translate the captions from Finnish to English on YouTube (Google MT helps!)



5. Captions can be stored separately and edited locally (e.g., as .srt files)

6. In the original video project file, just swap Finnish slides with English ones (hide and show the corresponding parallel video tracks).

7. Upload the English video on YouTube



8. Re-use/include the same multilingual captions that were created for the Finnish version of the video.

# Simple video: Lecture slides + narration

The screenshot displays the Camtasia 2018 software interface. At the top, the menu bar includes 'File', 'Edit', 'Modify', 'Text', 'View', 'Share', 'Window', and 'Help'. The main window title is 'viikko1\_johdanto\_1-3.cmproj'. The interface is divided into several sections:

- Record:** A red circle icon and a 'Share' button.
- Media Bin:** A central area containing various media assets, including a slide titled 'Kurssin aiheet' and several audio clips labeled 'Rec 02-23-15 1' through '12'.
- Slide Preview:** A large central window showing a slide titled 'Kurssin aiheet' with a bulleted list of topics and a paragraph of text.
- Empty Selection:** A panel on the right with the text 'Select an object on the timeline to see its properties.'
- Timeline:** A horizontal axis at the bottom showing time markers from 0:04:04:00 to 0:04:16:00. It features multiple tracks: Track 5, Track 4, Track 3 (with audio clips 'Rec 02-23-15 34' and 'Rec 02-23-15 33'), Track 2 (with 'korpusingvistiikka\_uento1 2'), and Track 1.

The slide content is as follows:

Korpuslingvistiikan johdantokurssi  
Kurssin suorittaminen  
Sisällöt

## Kurssin aiheet

- Erilaisia korpuksia
- Keskeisimmät korpusmenetelmät
- Korpusten kokoamisen periaatteet
- Mitä on tiedettävä tekijänoikeuksista ja henkilösuojusta
- Teksti- ja puhekorpuksen erot
- Tilastollisen analyysin mahdollisuudet erilaisten kielen ilmiöiden tutkimisessa
- Korpusten käyttömahdollisuudet ja rajoitukset
- Miten ja mistä aloitan oman korpustutkimukseni?

Jos sinulla on mielessä oma tutkimusaihe (gradu tms.) jossa tarvitaan korpusmenetelmiä, kysy rohkeasti!

Miitta Lennes Korpuslingvistiikan johdantokurssi



- Media
- Library
- Annotations
- Transitions
- Behaviors
- Animations
- Cursor Effects
- More

### Media Bin

Analysis tools:

- Kielipankki
- Korp search
- Korp data
- Relation algebra
- Simple statistics
- Advanced statistics
- Export formats
- Visualizations
- Morphological analysis
- Speech recognition
- Preprocessing
- Finle-State Technology
- Finle-State Transducers

Workflow:

```

graph TD
  text[teksti.txt] --> tool[Tool]
  
```

teksti.txt  
100 B, Tue Mar 13 14:24:27 EET 2018  
(Click here to add your note)  
Created with Chipster 3.12.2  
Analysis history

Import / Import data

### Captions

Hide Caption Track

Tällä kertaa tulos tuli nopeasti. Jäsennystulos tulee näkyviin sekä työvuokaavioon

02:56/07:16

Properties



Track 6

Track 5

Track 4

Track 3

Track 2

Track 1