



Winter School on Corpus Data for the Analysis of Discourse, Interactions and Arguments (February 7-9, 2023)

In the context of *LAND: Doctoral Programme in Applied Linguistics: Managing Languages, Arguments and Narratives in the Datafied Society*

- **Place:** A-22 (Red building at USI campus: via Giuseppe Buffi 13, Lugano 6900)
- **Lecturers:** Noah Bubenhofer, Aleksandra Gnach, Annette Hautli-Janisz, Julia Krasselt, Johanna Miecznikowski-Fuenfschilling, Chris Reed
- **Discussants:** Corina Andone, Steve Oswald, Daniel Perrin, Andrea Rocci, Olena Yaskorska-Shah
- **ECTS:** 2
- **Evaluation:** The evaluation of the participants' work will be based on two components: 1) attendance to the whole school and active participation in the *methodological seminars* and in the *interactive sessions*; 2) individual presentations in the *PhD Data sessions*, and the subsequent discussion with the senior researchers acting as discussants and their peers. See *Program description* below for details. Evaluation will be pass/fail.

Topic, focus

The main goal of the Winter School is to create a stage where PhD students can meet, interact and discuss their corpus data with Swiss and international senior researchers who are recognized leaders in the domains of corpus studies, discourse analysis and argumentation mining. The school involves three types of activities: 1) longer methodological seminars, 2) shorter interactive sessions guiding students through issues of data analysis, and 3) data sessions where PhD students present their ongoing research and discuss issues and solutions regarding the analysis of corpus data pertaining to their PhD theses.

Program description

The school includes the following activities:

Data-focused seminars (10h)

- *QT30 and beyond: Analysing argumentation in the wild*, Chris Reed (University of Dundee) and Annette Hautli-Janisz (University of Passau):
- *Collecting and analyzing data on social media*, Aleksandra Gnach (ZHAW):

Interactive seminars (3h)

- Noah Bubenhofer, Julia Krasselt: *Corpus Linguistics for Discourse Analysis: Data and Methods*
- Johanna Miecznikowski: *Transforming talk into data: the role of transcription*

Panel on Open Research Data in the area of linguistics (1h)

Projects discussed:

- *Upgrading the linguistic ORD-ecosystem Up-LORD* (main applicant: UZH, PI: Noah Bubenhofer)
- *Swiss-AL: Linguistic ORD Practices for Applied Sciences* (main applicant ZHAW, PI: Julia Krasselt)
- *Data-sharing skills in corpus-based research on talk-in-interaction* (main applicant: USI, PI: Johanna Miecznikowski-Fuenfschilling)

Data sessions (11h in total)

Data sessions occupy 45' slots, including 20' presentation and 25' discussion. They focus on the data analyzed in one's PhD thesis and can discuss a variety of problems including data collection, selection, transcription, annotation, interpretation, qualitative or quantitative analysis. The main focus is not to present a finished paper, but to open a window on one's PhD thesis laboratory in order to receive constructive feedback on the thesis' work in progress.



Detailed program

Tuesday, February 7th

- 8³⁰-10⁰⁰ Seminar: C. Reed, A. Hautli-Janisz:
QT30 and beyond: Analysing argumentation in the wild
- 10⁰⁰-10³⁰ Coffee
- 10³⁰-12³⁰ Seminar: C. Reed, A. Hautli-Janisz:
QT30 and beyond: Analysing argumentation in the wild
- 12³⁰-14⁰⁰ Lunch
- 14⁰⁰-14⁴⁵ DS: G. D'Agostino, *Pretty little lies. Uncovering dialectical and argumentative patterns emerging from irregular or unsatisfactory answers in the financial dialogue*
- 14⁴⁵-15³⁰ DS: R. Younis, *Experimental investigations into the rhetorical effects of rephrase*
- 15³⁰-16⁰⁰ Coffee
- 16⁰⁰-18⁰⁰ Seminar: C. Reed, A. Hautli-Janisz
QT30 and beyond: Analysing argumentation in the wild

Wednesday, Feb 8th

- 8³⁰-10⁰⁰ DS: Costanza Lucchini, *Epistemics and evidentials in ECCs*
DS: J. A. Lomelí Hernández, *Crossing data between corpora: from press conferences to twitter*
- 10⁰⁰-10³⁰ Coffee
- 10³⁰-12³⁰ Seminar: A. Gnach
Collecting and analyzing data on social media
- 12³⁰-14⁰⁰ Lunch
- 14⁰⁰-15³⁰ Seminar: A. Gnach
Collecting and analyzing data on social media
- 15³⁰-16⁰⁰ Coffee
- 16⁰⁰-16⁴⁵ DS: C. Mercuri, *Analysing conflicting frames and argumentation in the polylogical controversy surrounding fashion sustainability: preliminary findings and challenges*
- 16⁴⁵-17³⁰ DS: D. Barcella, *The strategic communication of the World Wide Web: a qualitative analysis of WWW's communication through the CERN archive*
- 17³⁰-18¹⁵ DS: D. de Oliveira Fernandes, *An experimental approach to the rhetorical effectiveness of insinuated ad hominem*

Thursday, Feb 9th

- 8³⁰-10³⁰ Seminar: Noah Bubenhofer, Julia Krasselt
Corpus Linguistics for Discourse Analysis: Data and Methods
- 10³⁰-11⁰⁰ Coffee
- 11⁰⁰-12⁰⁰ Seminar: Johanna Miecznikowski-Fuenfschilling
Transforming talk into data: the role of transcription
- 12⁰⁰-12⁴⁵ Panel on Open Research Data projects in linguistics
- 12⁴⁵-14⁰⁰ Lunch
- 14⁰⁰-14⁴⁵ DS: D. D'Agostino, *Heuristics towards an analysis of deictic practices in digital discourse*
- 14⁴⁵-15³⁰ DS: Narjes, *Studies in Syntactic complexity, Lexical Sophistication, and Argumentative patterns in a corpus of published journal articles written by native and non-native English writers*
- 15³⁰-16⁰⁰ Presentation: Cristina Grisot
CLARIN – Common Language Resources and Technology Infrastructure
- 16⁰⁰-16³⁰ Coffee
- 16³⁰-17¹⁵ DS: E. Angiolini, *Subdiscussions in children-adult argumentative interactions*
- 17¹⁵-18⁰⁰ DS: E. Battaglia, Elena Battaglia (USI): *Evidentiality in Italian talk-in-interaction. Incremental and interactive categorization of sources of knowledge.*



Seminar descriptions

Chris Reed (University of Dundee), Annete Hautli-Janisz (University of Passau): *QT30 and beyond: Analysing argumentation in the wild*

QT30 is the largest corpus of analysed dialogical argumentation ever created and contains 30 episodes of BBC's 'Question Time' from 2020 and 2021 (Hautli-Janisz et al. 2022), analysed with Inference Anchoring Theory (Budzynska et al. 2014), a framework for large-scale analysis of natural argumentation in dialogue. In this course we introduce students to the core properties of the corpus and have hands-on sessions on how IAT can be used to annotate well-versed political rhetoric as well as direct, often combative, justification-seeking of the general public. The course also features a session of applying a novel annotation layer for reframing, i.e., how the salience of pieces of information is adjusted in a debate.

Aleksandra Gnach (ZHAW): *Collecting and analyzing data on social media*

Social media platforms offer the opportunity to collect a wide range of data. However, the volume and variety of data are often overwhelming, and it is difficult to determine a focus of analysis or a research design with suitable methods. In this session, we first discuss how case studies can be used as a framework for social media research. Then, we determine what research questions lend themselves to social media case studies and how they can be answered by combining qualitative and quantitative data. In addition, we get to know tools that are useful for collecting quantitative social media data without programming knowledge. Finally, we will apply the acquired methodological knowledge by collecting data for a case study on social media.

Noah Bubenhofer (University of Zurich), Julia Krasselt (ZHAW), *Corpus Linguistics for Discourse Analysis: Data and Methods*

Working with large text corpora is part of the standard repertoire of empirical linguistic research. In this interactive session, Julia Krasselt and Noah Bubenhofer give insight into working with newspaper corpora.

The basis is the huge media archive `Swissdox@LiRI`: It contains the articles of more than 200 Swiss media, which can be downloaded free of charge for academic purposes. Using these data, exemplary corpus linguistic analyses are shown.

These topics are addressed:

- `Swissdox@LiRI` data basis: conditions of use, downloading the data (Noah).
- Basics of data analysis with the statistical software R (Julia)
- Example calculation of topic models with R (Julia)
- Basics of data analysis with Python using KoDuP (<https://gitlab.uzh.ch/noah.bubenhofer/kodup-germanistik/>) (Noah)
- Example calculation of word embeddings for semantic analyses (Noah)
- Discussion and Q & A

Johanna Miecznikowski Fuenfschilling (USI), *Transforming talk into data: the role of transcription*

Writing down somebody else's spoken discourse in order to keep a permanent, transportable and reusable record is widely practiced in literate societies, with diverse functions, and implies a material, formal, conceptual and pragmatic transformation of the inscribed discourse. In the case of transcription in linguistics, the resulting document is a piece of data, which acquires meaning within the scientific practices of the discipline's various fields. Specific choices are made according to the object and goals of investigation, technical aspects (e.g. the use of tools for creating, storing, viewing and analyzing recordings and transcripts), the available manpower, and the intended data users



(restricted or open access); transcribers may need to partly abandon entrenched writing and reading habits to make good transcription choices. In the workshop, these issues will be discussed (a) by comparing alternative transcription solutions regarding phenomena that are characteristic of spoken language such as regional variation, prosody, hesitation, repair, and overlap; (b) by practically exercising the transcription of prosody in a text editor and then examining its implications for syntactic analysis. Equipment required to participate in the exercise: computer and headphones.

Bio notes/ Profiles

Corina Andone, Assistant Professor of Language and Communication and senior researcher at the Amsterdam Centre for Language and Communication (ACLC)

<https://www.uva.nl/en/profile/a/n/c.andone/c.andone.html>

Noah Bubenhofer, Professor of German Linguistics at the University of Zurich

<https://www.ds.uzh.ch/static/cms/pfs/personen.php?detail=22>

Aleksandra Gnach, Professor of Media Linguistics with a focus on Social Media at ZHAW

<https://www.zhaw.ch/en/about-us/person/gnaa/>

Annette Hautli-Janisz, Professor of Computational Rhetoric und Natural Language Processing, University of Passau (DE)

<https://www.fim.uni-passau.de/en/research-professorships/lehrstuehle-professuren-und-fachgebiete/cornlp/team/annette-hautli-janisz/>

Julia Krasselt, Research Associate in Corpus Linguistics and Discourse Analysis at ZHAW

<https://www.zhaw.ch/en/about-us/person/krss/>

Johanna Miecznikowski-Fuenfschilling, Professor of Italian Linguistics at USI

<http://usi.to/hg8>

Steve Oswald, Senior researcher at the English Department, University of Fribourg

<https://www.unifr.ch/directory/en/people/19260/95d70>

Daniel Perrin, Professor at ZHAW school of applied linguistics

<https://www.zhaw.ch/en/about-us/person/pdan/>

Andrea Rocci, Professor at the Institute of Argumentation Linguistics and Semiotics, USI

<https://search.usi.ch/en/people/34244e1f31def74e55a4259bb8705bea/rocci-andrea>

Chris Reed, Professor of Computer Science and Philosophy at the University of Dundee (UK)

<https://arg.tech/~chris/>

Olena Yaskorska-Shah, Post Doc Researcher at the Institute of Argumentation Linguistics and Semiotics, USI

<https://search.usi.ch/en/people/c6b87def72819c3b3d6186d6673d37a2/yaskorska-shah-olena>