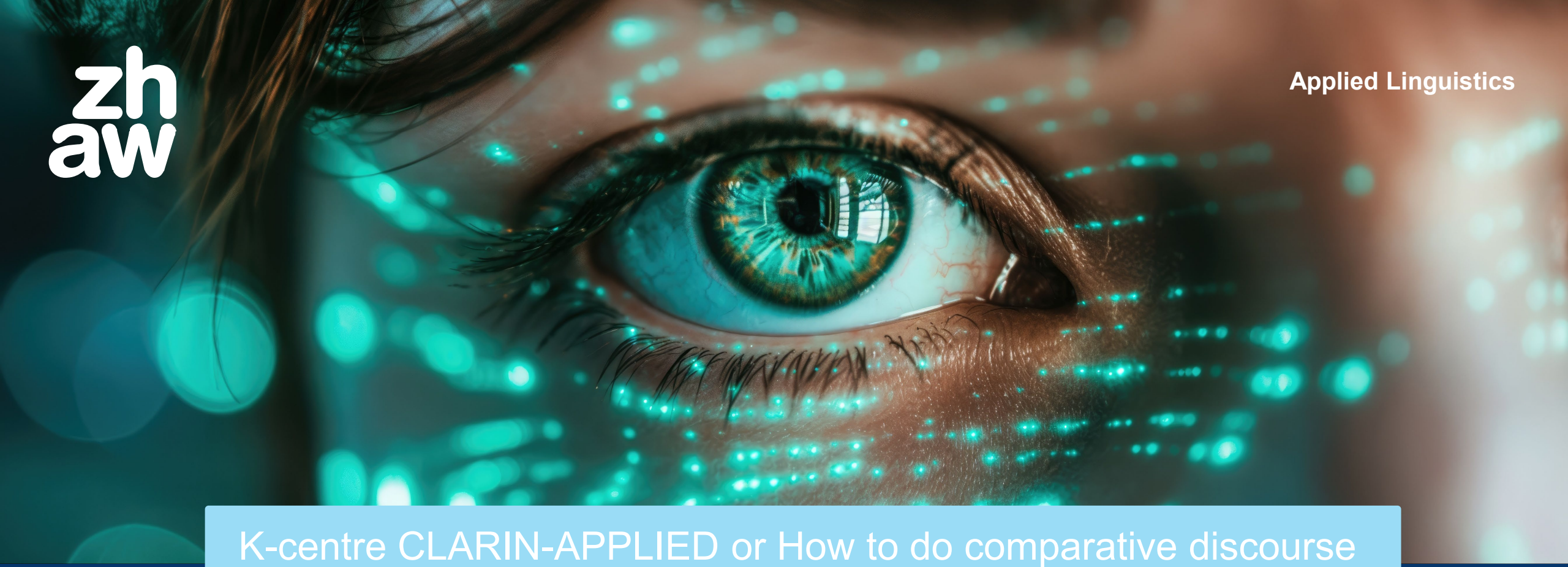


Applied Linguistics

K-centre CLARIN-APPLIED or How to do comparative discourse studies using corpus linguistics?

9.12.2025

Talk in CLARIN Café – SSH Research with CLARIN K-Centres: Expertise for Multimodal Data, Large Language Models and Discourse Analysis

by Philipp Dreesen & Dolores Lemmenmeier-Batinić

Outline

1) What is Applied Comparative Discourse Analysis?
2) Needs and Demands in the scientific community
3) K-centre: Website
4) Supportive Activities
5) K-centre: Next steps

1) What is Applied Comparative Discourse Analysis?

Discourse analysis
- does research on orders of language use and knowledge
- uses QUAL and QUAN (corpus linguistics) methods
- should not be limited to only one language/region/media type/time period/etc.

Comparative discourse analysis
- needs a tertium comparationis (e.g., a topic like ‚public service broadcasting', a concept like ‚east')
- compares elements to find similarities and differences between discourses
- can be multilingual but also monolingual

Desiderat: How to do comparative discourse studies using corpus linguistics?

2) Needs and Demands in the scientific community

Scenario of a corpus-based comparative discourse study

Research question: What similarities and differences exist in the discursive construction of public service broadcasting in European countries?

https://thenounproject.com/icon/reach-627744/

Scenario of a corpus-based comparative discourse study

Country	Public service broadcasting	terms
Germany	ARD, ZDF, Deutschlandradio	*ARD, ZDF, Dradio, Deutschlandradio, Das Erste, Dritte Programme, die Öffentlich-Rechtlichen, …*
Poland	Telewizja Polska (TVP), Polskie Radio	*TVP, PR, Telewizja Polska, Polskie Radio, …*
Italy	RAI	*RAI, Radiotelevisione Italiana, …*
UK	BBC	*BBC, British Broadcasting Corporation, Public Service Broadcasting, PSB, Auntie, …*
Hungary	MTVA	*MTVA, Médiaszolgáltatás-támogató és Vagyonkezelő Alap, M1, Kossuth Rádió*
Switzerland (DE)	SRF	*SRF, Schweizer Rundfunk- und Fernsehgesellschaft, SRG, …*

Typical questions in the community of comparative discourse studies

Methodological questions

- Where does the discourse on public broadcasting take place? e.g., press, radio, comments
- When/at what point in time is the discourse likely to be worth investigating? e.g., a discourse changing event
- In which languages? e.g., more than one language per country
- How is the subject of investigation verbalized? e.g., multi-word units
- To which degree a standardisation of data is useful? e.g., annotation
- Requires same tools same data and vice versa?

Practical research decisions

- Buy data? e.g., SketchEngine
- Freely accessible data/open research data? e.g., CLARIN VLO
- Crawl data yourself? e.g., Trafilatura
- Hybrid forms? e.g., process crawled data in LancsBox

What do they need?

Priority:

- Support in creating discourse models before searching for corpora or compiling corpora

- Support in transferring QUAL into QUAN

- for the tertium comparationis: ngram emeddings to find semantic spaces instead of search words

Welcome to CLARIN-APPLIED!

The CLARIN Knowledge Centre for Applied Comparative Discourse Analysis (CLARIN-APPLIED) offers expertise in applied discourse analysis, especially with multilingual corpora and corpora suitable for comparative analysis. This encompasses five areas of expertise and tools for all users:

1. Best practices in **applied comparative discourse analysis**;
2. access to the **Swiss-AL platform** with various multilingual (German, French, Italian) corpora for discourse analytical purposes;
3. access to **documentation** of corpora and tools;
4. access to multimodal **learning material**;

Philipp Dreesen

Dolores Lemmenmeier-Batinić

The K-centre is operated by the ZHAW Digital Discourse Lab.
https://www.zhaw.ch/en/linguistics/business-services/digital-discourse-lab

3) K-centre

Website: https://clarin-applied.zhaw.ch

Offered areas of expertise and tools for all users:

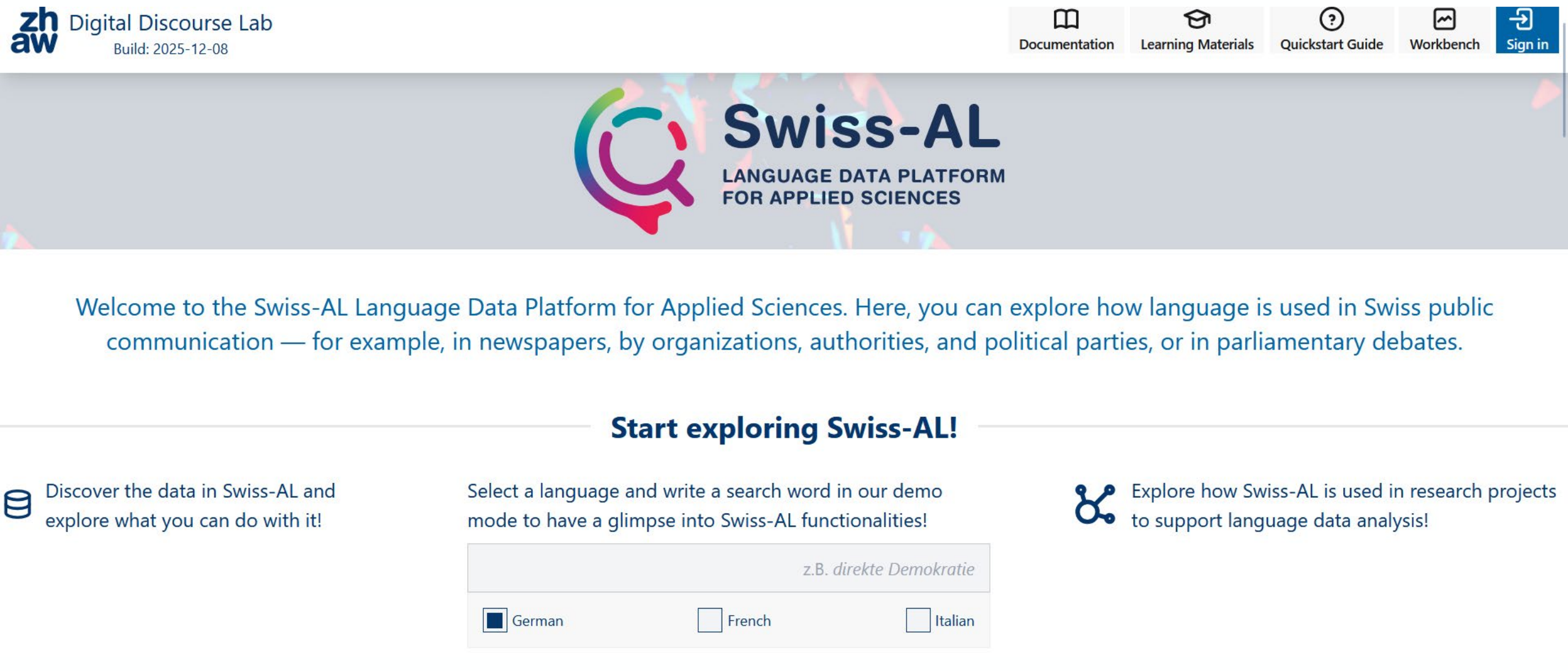
1. Best practices in **applied comparative discourse analysis**;
2. access to the **Swiss-AL platform** with various multilingual (German, French, Italian) corpora for discourse analytical purposes;
3. access to **documentation** of corpora and tools;
4. access to multimodal **learning material**;
5. curated list of relevant publications, as well as resources and tools **relevant for comparative discourse analysis**.

Resources

On the Swiss-AL Platform we provide access to corpora of Swiss public communication, which currently consist of:

- Journalistic corpora;
- Parliamentary corpora;
- Corpora of press releases and media texts from politics, economy, science, and civil society.

Swiss-AL Platform: https://swiss-al.zhaw.ch



Swiss-AL Platform as a corpus workbench

Swiss-AL
LANGUAGE DATA PLATFORM FOR APPLIED SCIENCES
Build: 2025-12-08

Documentation | Learning Materials | Quickstart Guide | Landing | User

Current Project
OER TM
Switch

Selected Corpus
FR Journalistic Corpus (high reach+regional, 2018-25)
Switch

Philipp's workspace

Distribution (Documents) - Musk | Overview | Semantic Space | Topics

- Explore Corpus
  - Overview
  - Topics
  - Keywords
  - Semantic Space
- Search Corpus
  - Distribution of Words
  - Context of Words
  - Distribution in Documents
  - Collocations

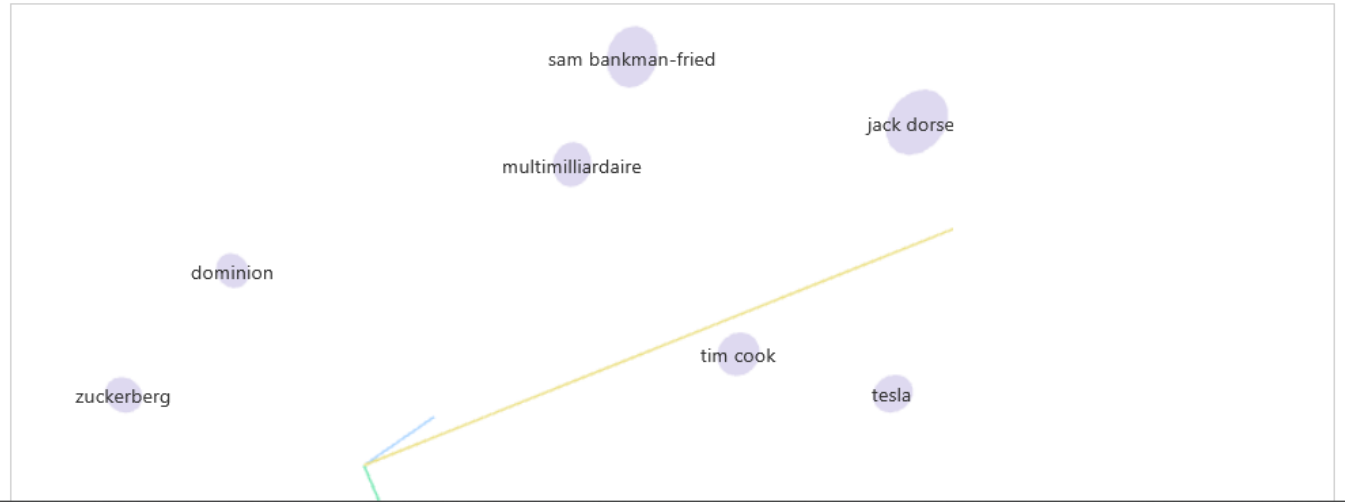
Semantic Space

musk

Submit

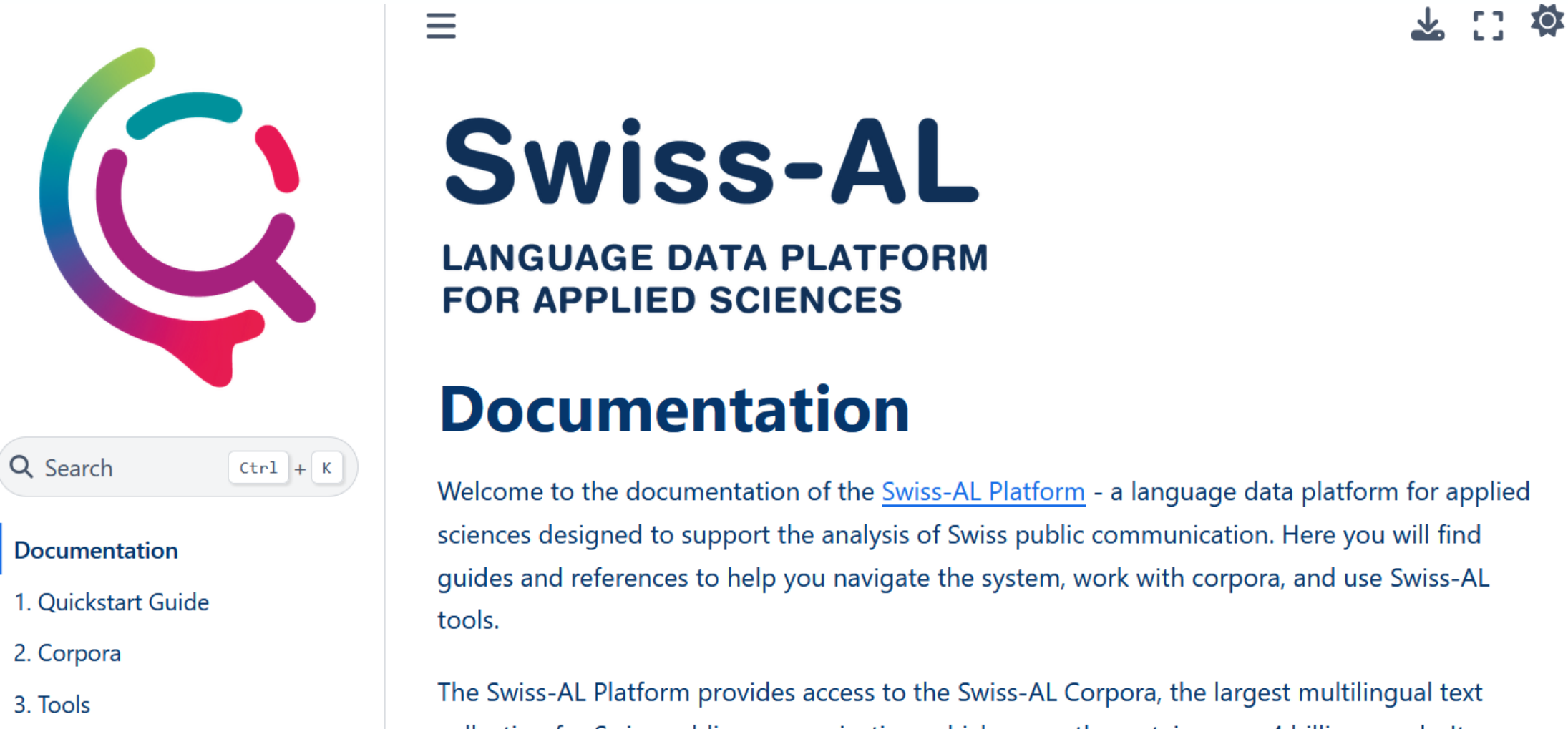
Display 40 terms

Activate 3D view

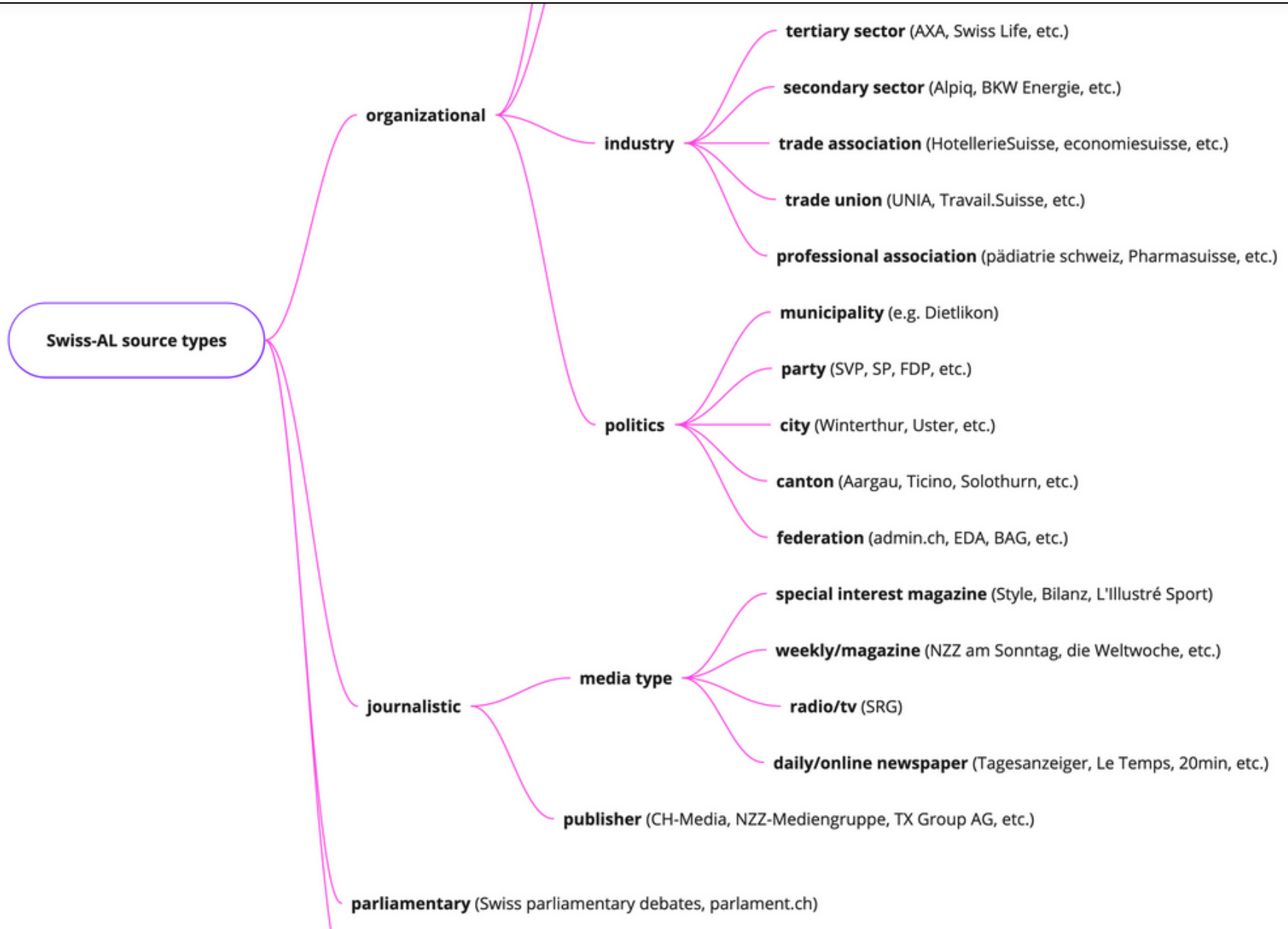


Lemma	Similarity	Absolute Frequency
musk	1	2827
elon musk	0.863	18115
zuckerberg	0.859	333
trump	0.76	62442
elon	0.757	548
mark zuckerberg	0.751	3396

Swiss-AL Documentation: https://github.zhaw.ch/pages/digitallinguistics/swiss-al-docu/intro.html



Corpora in DE, FR, IT from diferrent sources



Documentation with examples in DE, FR, IT

Search Ctrl + K

Search Type	Syntax	Explanation
One-term search	`"democrazia"`	Searches for the exact word *democrazia*.
Word search using parameter	`[word="democrazia"]`	Searches for the word *democrazia* using the **word** parameter.
Lemma search	`[lemma="democrazia"]`	Searches for all **inflected forms** of the lemma *democrazia*.
Part-of-speech (PoS) search	`[pos="NOUN"]`	Searches for **nouns** (*NOUN*) in the Italian PoS tagset.
Named Entity (NER) search	`<ne_entityType="LOC"> [];`	Searches for named entities labeled as *LOC* (locations). You can also specify *ORG* for

Contents

3) K-centre

Webinars:

- #1 ORD and Comparative discourse analysis (09.05.2025)
- #2 CLARIN Ressources and digital Tools for corpus linguistics in the context of comparative discourse analysis (17.06.2025)

Support:

- Swiss Parliament Data (DE, FR, IT) in TSV
- cooperation with opendata.ch
- preparing data for ParlaMINT

4) Supportive activities

OSCARS-project MORCDA

Network comparative discourse studies

OSCARS
Open Science Clusters' Action for Research & Society

Funded by the European Union

OSCARS funded Project

Making Open Research Data Suitable for Comparative Discourse Analysis (MORCDA)

Project Start:
01 Jan. 2025

Project End:
31 Dec. 2026

Field:
SSHOC

Researchers
Stephanie Evert DE
Waldemar Czachur PL
Goranka Rocco IT
Cristina Grisot CH
Philipp Dreesen CH (PI)

Università degli Studi di Ferrara

network comparative discourse studies

Digital Discourse Lab

CLARIN-CH

SWISS-AL

Friedrich-Alexander-Universität Erlangen-Nürnberg

zh aw

https://www.oscars-project.eu/projects/morcda-making-open-research-data-suitable-comparative-discourse-analysis

OSCARS
Open Science Clusters' Action for Research & Society

Funded by the European Union

Why using Open Data Research in Comparative Discourse Analysis?

1. The promotion of a European public sphere is essential for the cultivation of democratic discourse and opinion formation. In this context, discourse analysis offers profound insights into transnational media discourses, supported by multilingual and comparative perspectives.

2. Growing gap between non-commercial science and Big Tech: In terms of public discourse, the role of science is to help societies in Europe observe and understand their own opinion-forming processes and those of their neighbours in a data-based way.

3. Growing (?) gap between the mostly qualitative working community of comparative discourse researchers and the community of open research data

What is the solution for this gap?

Bridging the gap through communities of qualitative and quantitative research in the field of comparative discourse studies, e.g., webinars on:

1. **Sharing knowledge:** Information on ORD and CLARIN infrastructures, corpora and tools.
2. **Community Engagement and Training:** Develop targeted training materials to elevate ORD skills among comparative discourse researchers, promoting the integration of digital language resources.
3. **Best Practices and Tool Development:** Identify needs within the comparative discourse analysis community and develop best practices and digital tools tailored for multilingual and comparative research.

https://comparative-discourse-studies.net

network comparative discourse studies

The *network comparative discourse studies* connects researchers and lecturers who are interested in discourse analysis and comparative issues. We are focused on language-related discourse analysis, i.e., how language use generates collectively shared knowledge orders and positionings in societies. Comparative discourse analysis aims at comparing discourses as conditions, mechanisms, and objects. It addresses questions such as:

- What are the comparative dimensions to be considered in *intra*lingual discourse analyses?
- Which research designs are suitable for *inter*lingual comparisons?

https://comparative-discourse-studies.net/en/publications

network comparative discourse studies

About Bibliography Members Events Corpora Teaching

EN

Bibliography

Show 10 entries

Search:

Author	Year	Title	Language of Investigation	Studied Language
Münchow Patricia von	2012	Élever l'enfant entre tradition et modernité: une analyse du discours contrastive de guides parentaux français et allemands	FR	FR, DE
Schröter Melani, Veniard Marie, Taylor Charlotte, Blätte Andreas	2019	A comparative analysis of the keyword multicultural(ism) in French, British, German and Italian migration discourse	GB	FR, GB, DE, IT
Lewis Diana M	2017	Coherence Relations and Information Structure in English and French Political Speeches	GB, FR	GB
Mattfeldt Anna	2021	Risk in Discourses around Fracking: A Discourse Linguistic Perspective on the UK, the USA and Germany	GB	GB, DE

5) K-centre: Next steps

1) Dissemination, e.g., to members of network comparative discourse studies
2) Link to curated bibliography on https://comparative-discourse-studies.net/en/publications/
3) Publish webinar #1 and #2 as learning material
4) Planned workshop in Sarajevo in 2026 to generate learning material (scenario based best practices)
5) Planned webinar with K-centre CROATINA in Jan/Feb 2026
6) Planned exchange with the translation studies on the meaning of *equivalent*
7) Special Issue on comparative discourse analysis in Journal of Corpora and Discourse Studies (JCaDs), publishing in 2027; including the paper *Corpus-based Comparative Discourse Analysis between Needs and State of the Arts*