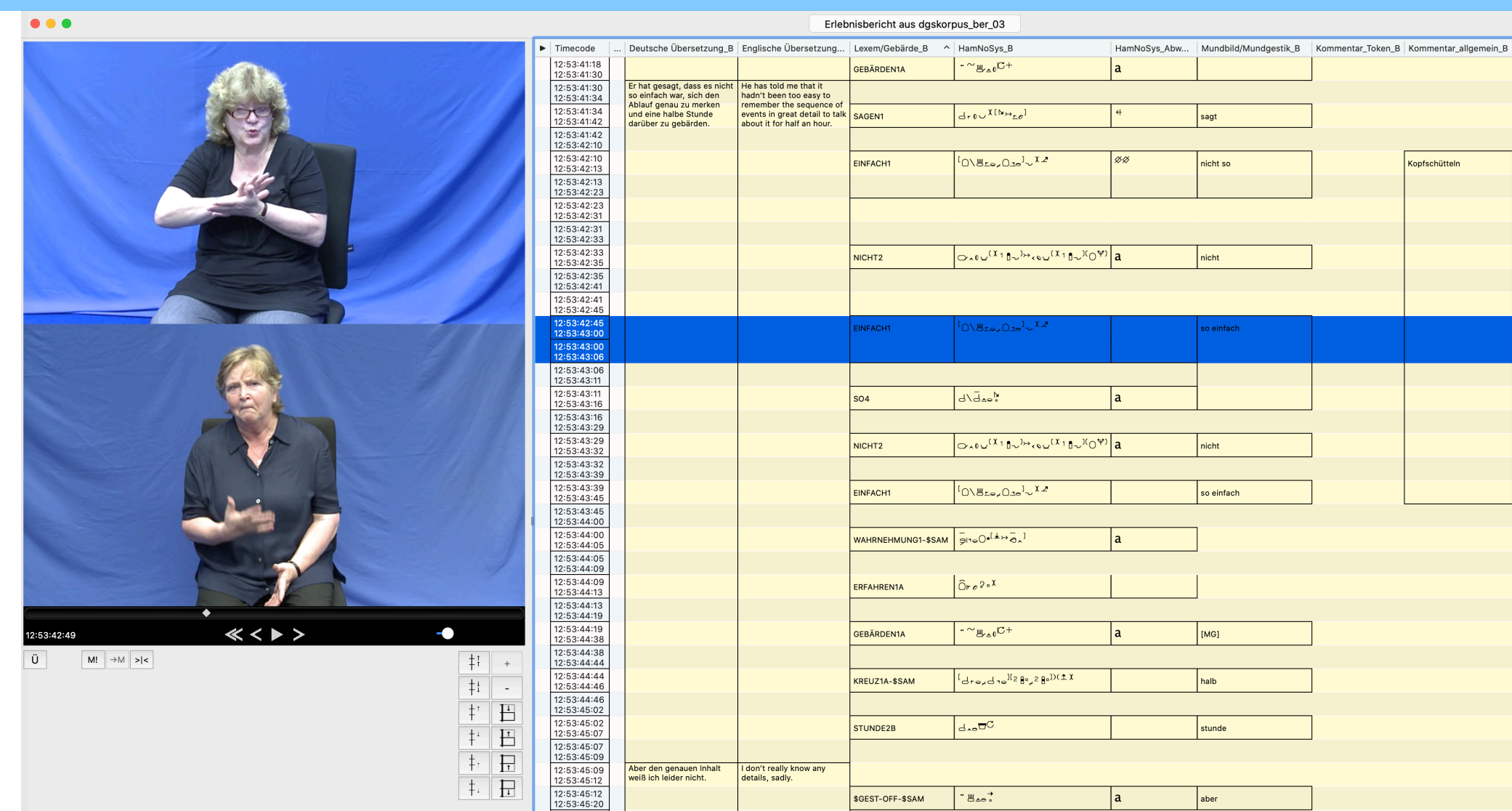


## DGS Corpus

- DGS Corpus:** collection (560 hrs) of dialogues between 330 (near-)native signers from 13 regions across Germany, 92.5 hrs lemmatized with 615,000 tokens; basis for:
  - The **Public DGS Corpus** (a subset of 50 hrs) on three different portals:
    - MY DGS**, for the DGS language community
    - MY DGS – annotated**, the research portal
    - MY DGS – ANNIS**, the corpus query portal
  - Corpus-based dictionary:** Digitales Wörterbuch der Deutschen Gebärdensprache (**DW-DGS**)



### New releases

- Products are extended progressively, by addition of new data, corrections to subtitles and annotations, as well as new features, e.g. new data formats, body pose information, persistent identifiers (DOIs), cross-references to other resources, collocation views and more

### Release history

- 2015: pre-release of **MEINE DGS**
- 2016/2017: addition of further videos
- 2018: Release 1, more videos & **My DGS – annotated**
- 2019: Release 2, videos completed (~ 50 hrs)
- 2020: Release 3, adding depth
- 2022: Release 4, adding functionality & **MY DGS – ANNIS**

### Public DGS Corpus Selection Process

- Include all participants and all tasks
- Prioritize elicitation tasks according to their importance for the language community
- Balance across region, age, gender
  - ➔ 7 core tasks:
    - Experience of deaf individuals, Free conversation,

- Discussion, Subject areas, Experience reports (world events), Region of origin, Deaf events
  - ➔ 14 other tasks
- + Jokes
- + retellings, description of procedures, staged negotiations ...

### Further Literature:

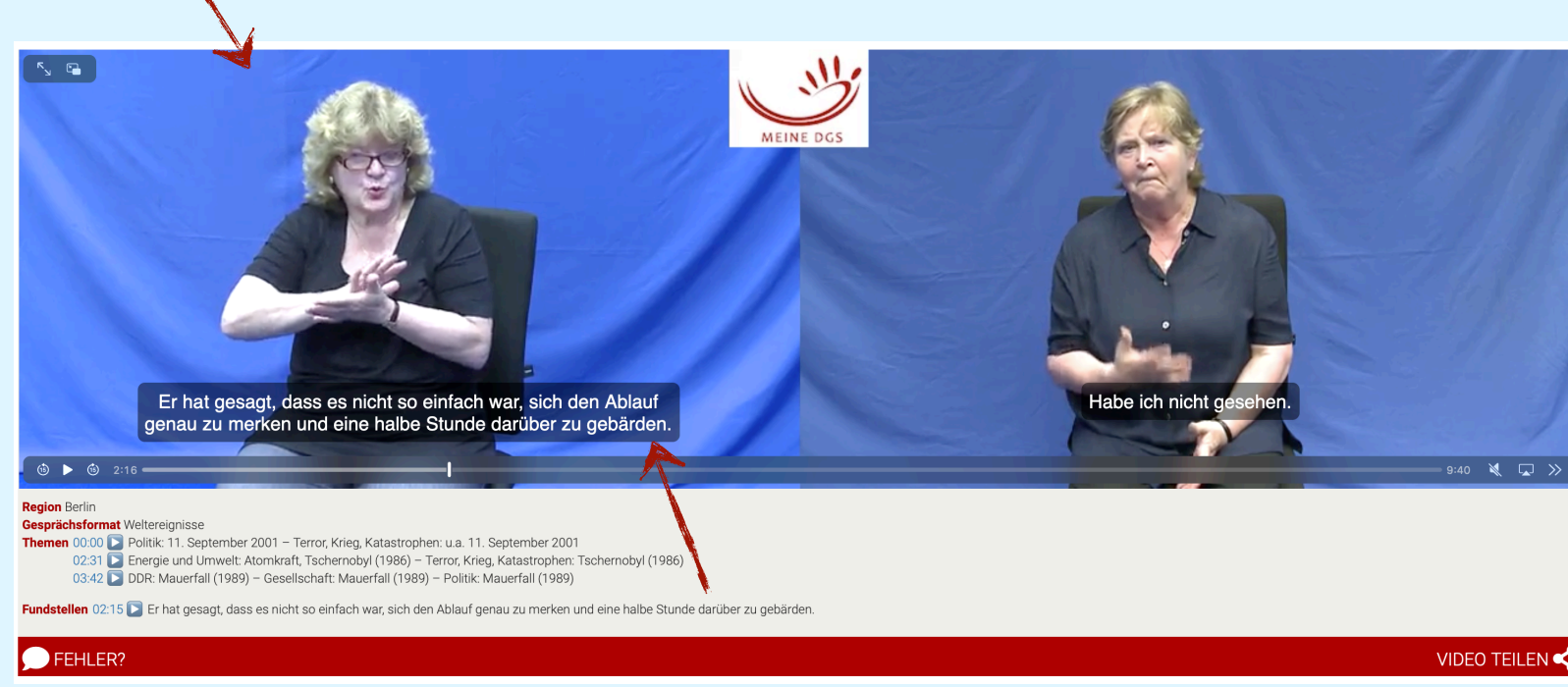
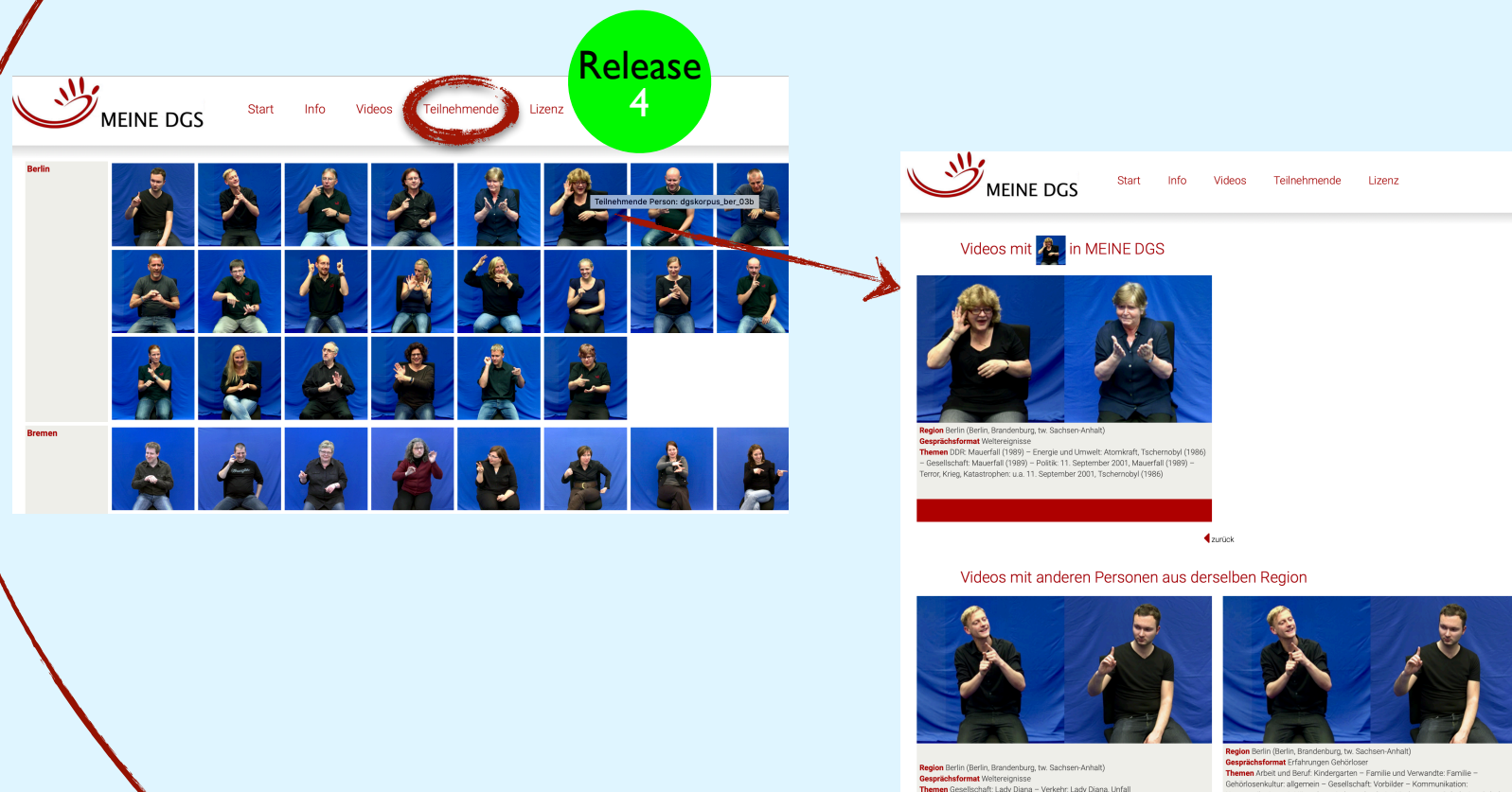
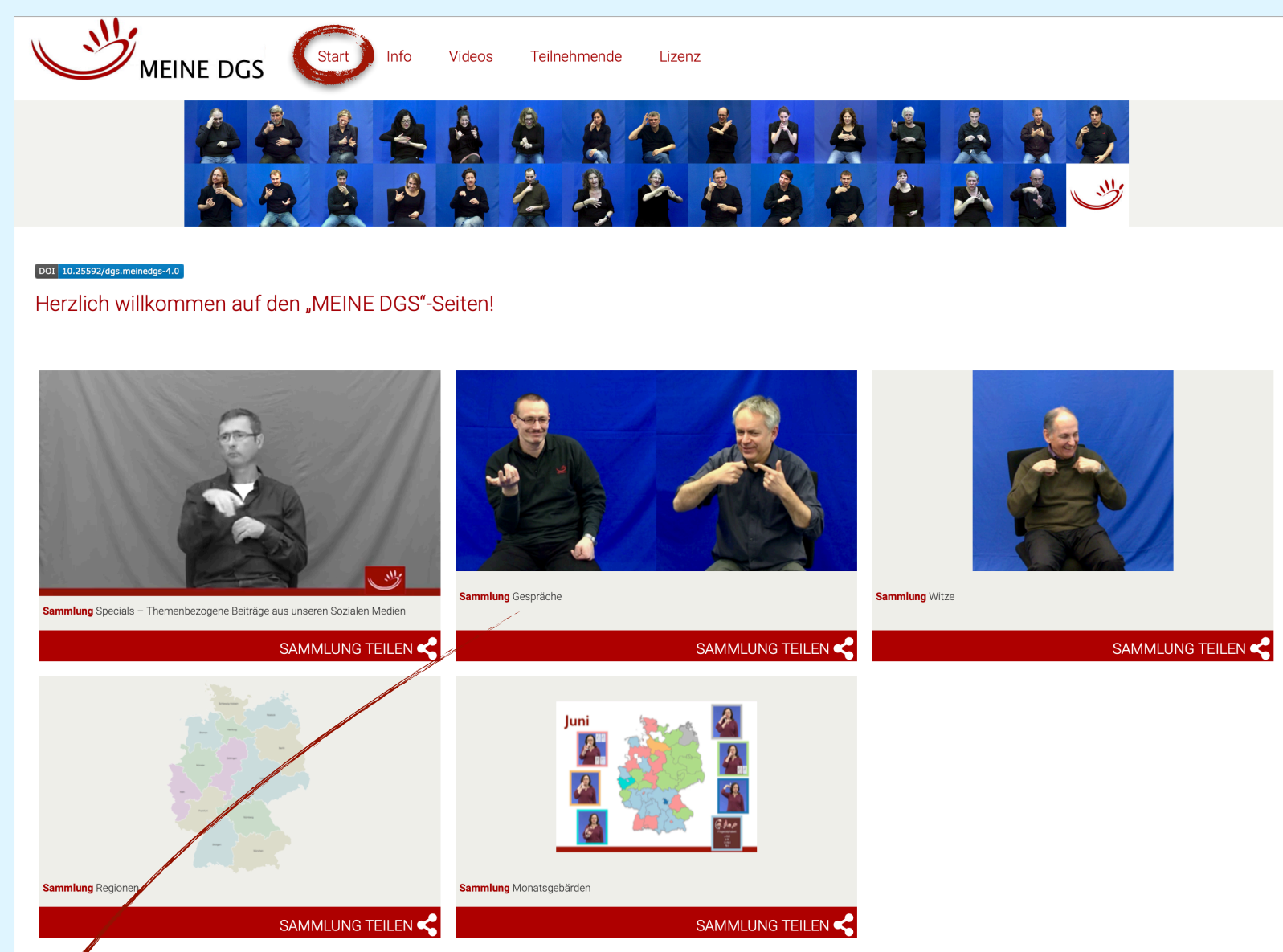
- Isard, Amy / Konrad, Reiner (2022): "MY DGS – ANNIS: ANNIS and the Public DGS Corpus". sign-lang@LREC
- Jahn, Elena / Konrad, Reiner / Langer, Gabriele / Wagner, Sven / Hanke, Thomas (2018): "Publishing DGS Corpus Data: Different Formats for Different Needs". sign-lang@LREC
- Müller, Anke / Hanke, Thomas / Konrad, Reiner / Langer, Gabriele / Wähl, Sabrina (2020): "From Dictionary to Corpus and Back Again – Linking Heterogeneous Language Resources for DGS". sign-lang@LREC
- Schulder, Marc / Blanck, Dolly / Hanke, Thomas / Hofmann, Ilona / Hong, Sung-Eun / Jezorski, Olga / König, Lutz / König, Susanne / Konrad, Reiner / Langer, Gabriele / Nishio, Rie / Rathmann, Christian (2021): "Data Statement for the Public DGS Corpus". DOI: <https://doi.org/10.25592/uhhfdm.9700>

## Public DGS Corpus

### MEINE DGS

- Only in German
- 7 core tasks (47.3 hrs; translation as subtitles)
  - + jokes (2.4 hrs, without translation)
  - + task "description of procedures" (0.9 hrs; translation as subtitles)

Release 4

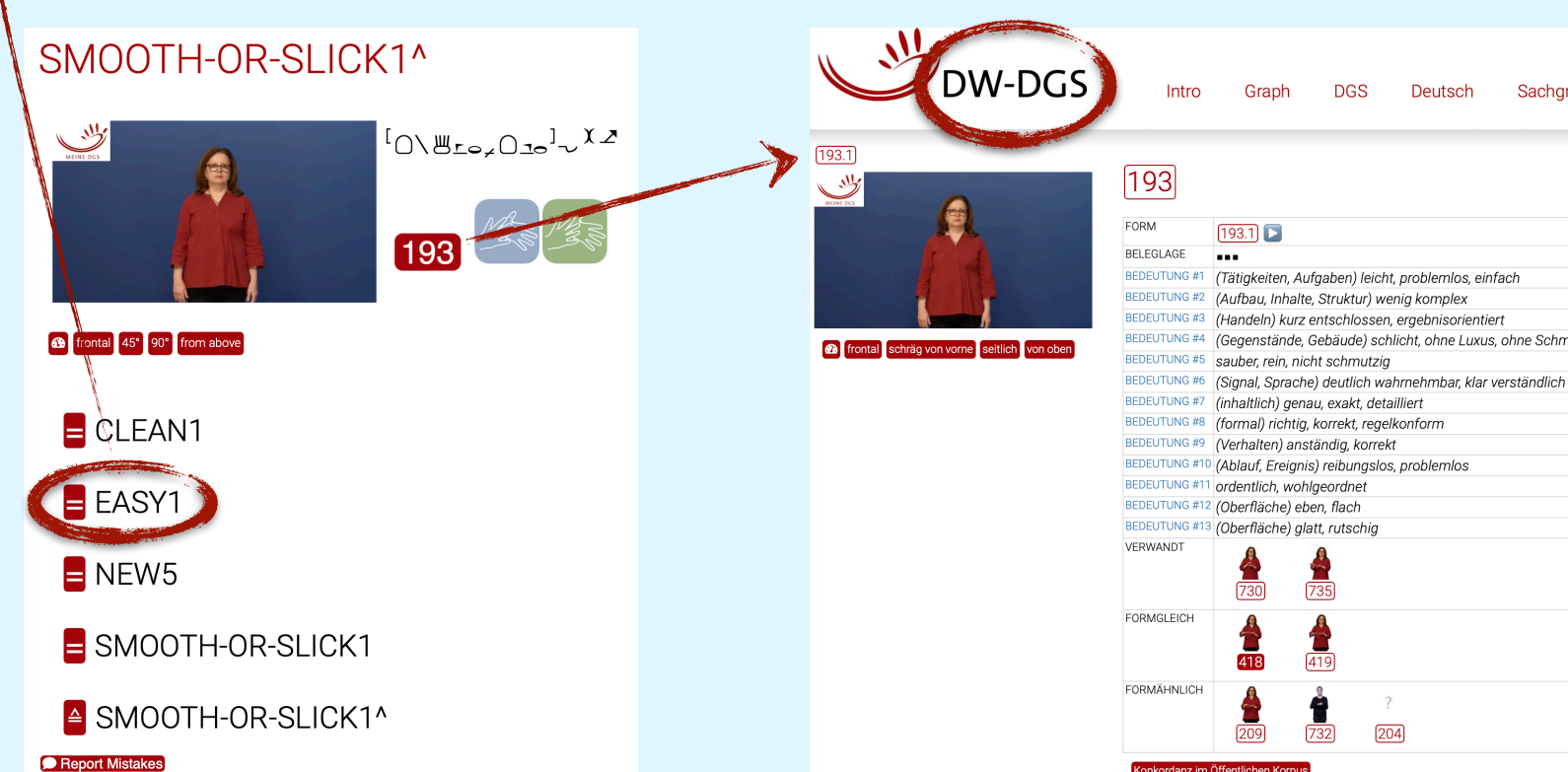
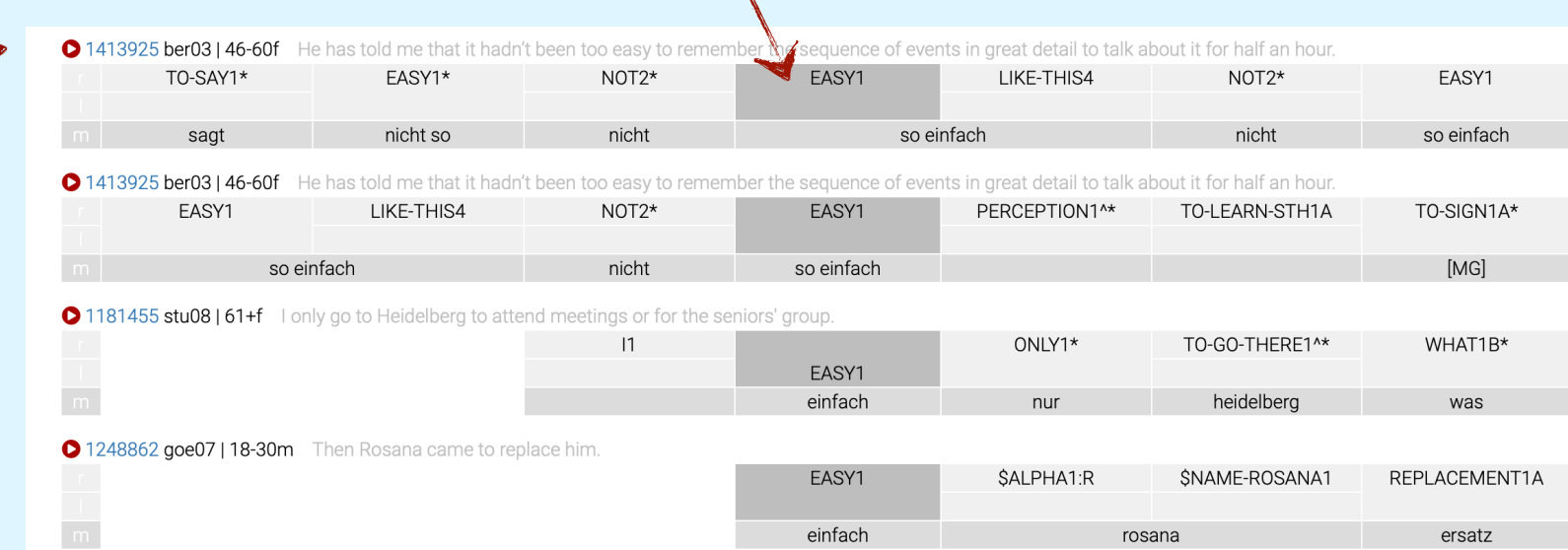
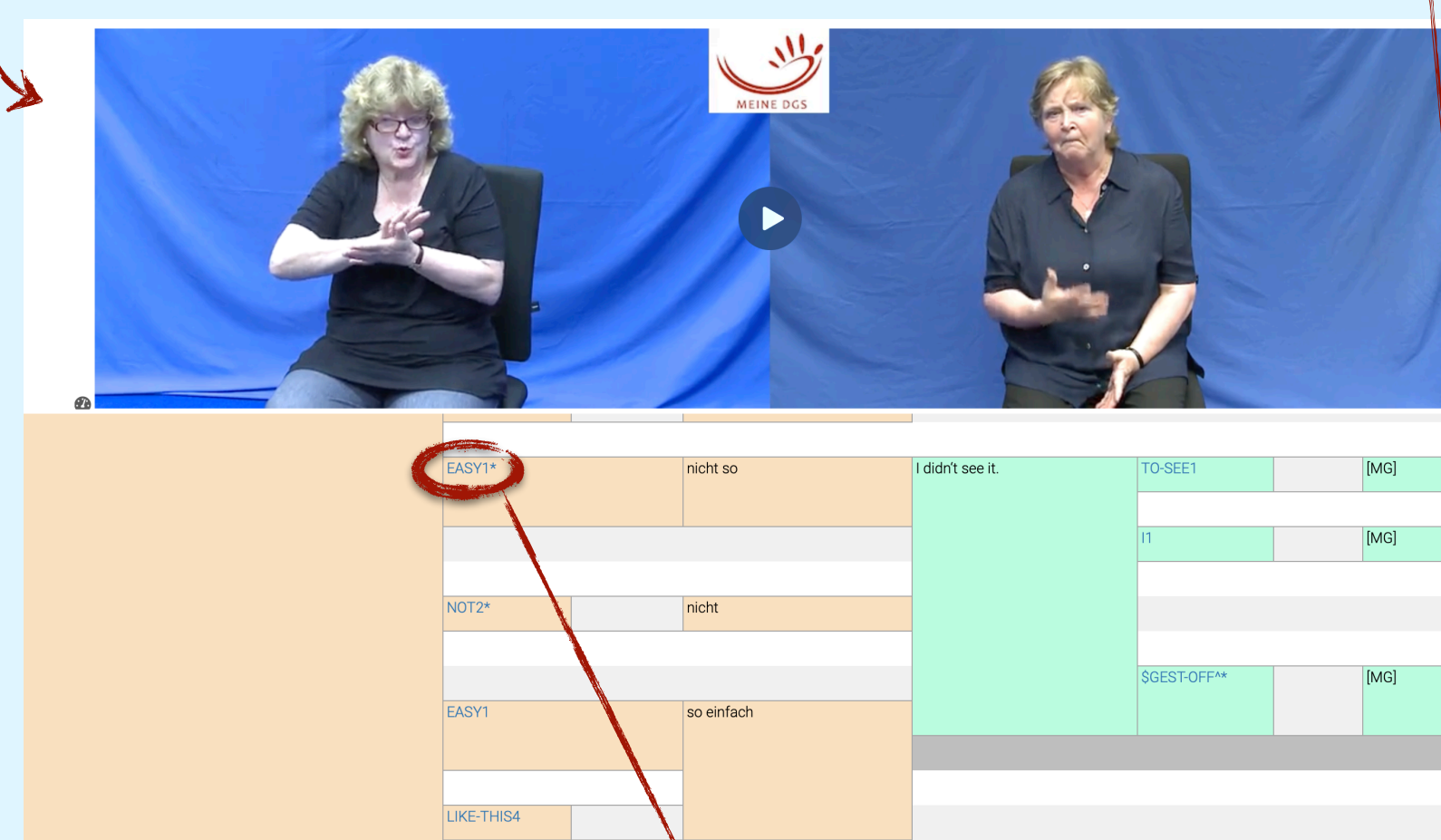
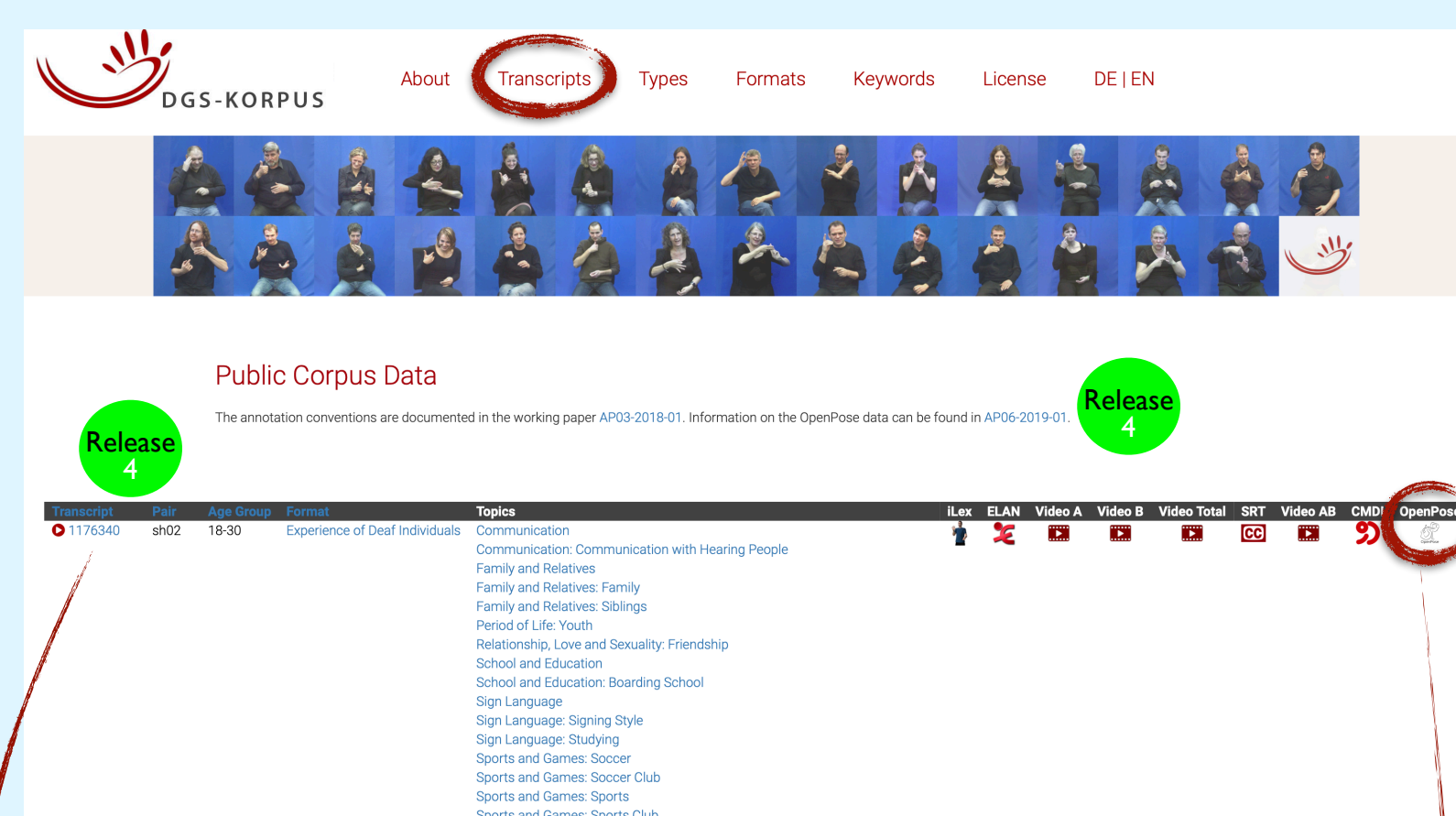


<https://meine-dgs.de>

### MY DGS – annotated

- Websites and annotations in German and English
- 7 core tasks (47.3 hrs; translated & lemmatized)
  - + jokes (2.4 hrs, without annotation)
  - + illustrative samples of 14 tasks (1.8 hrs; transl. & lemm.)
  - + task "description of procedures" (0.9 hrs; translated)

Release 4

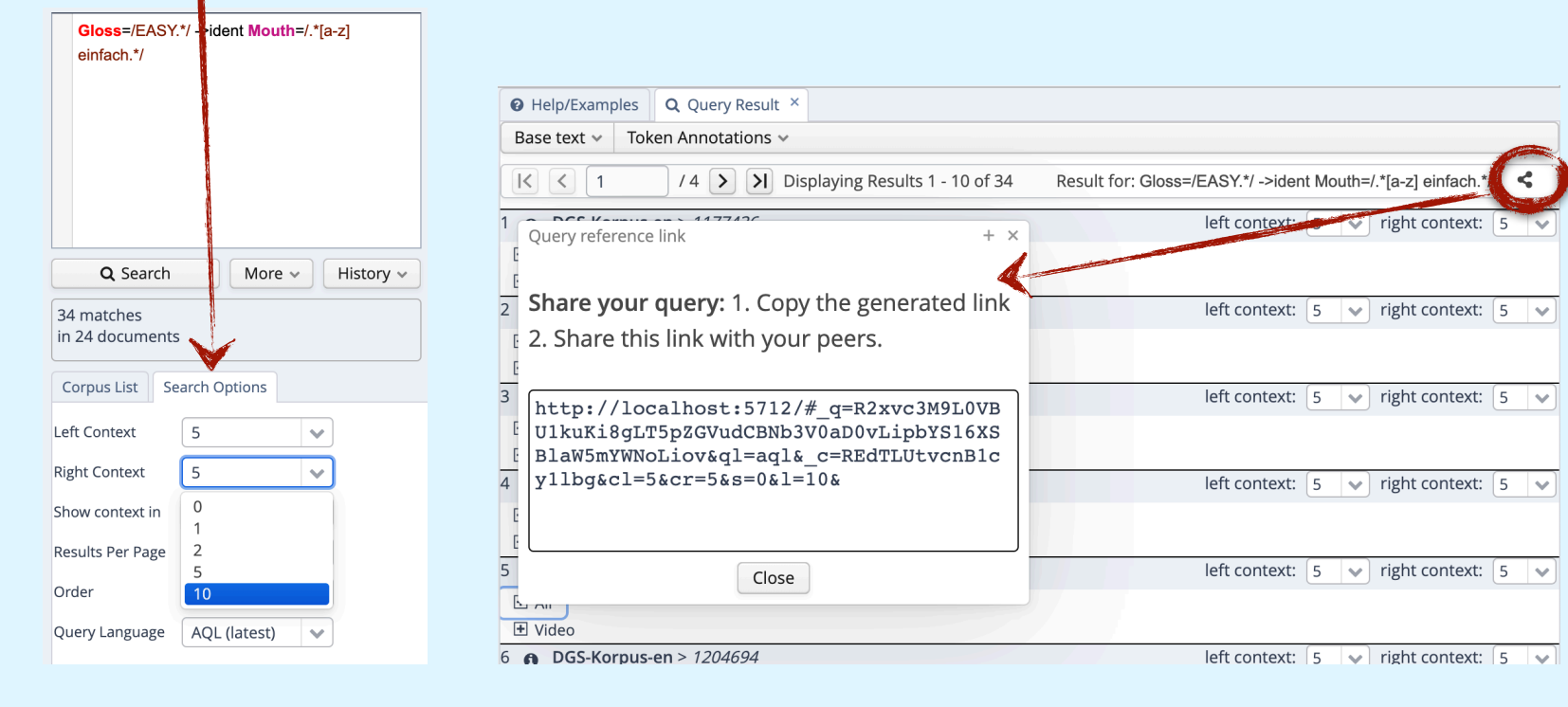
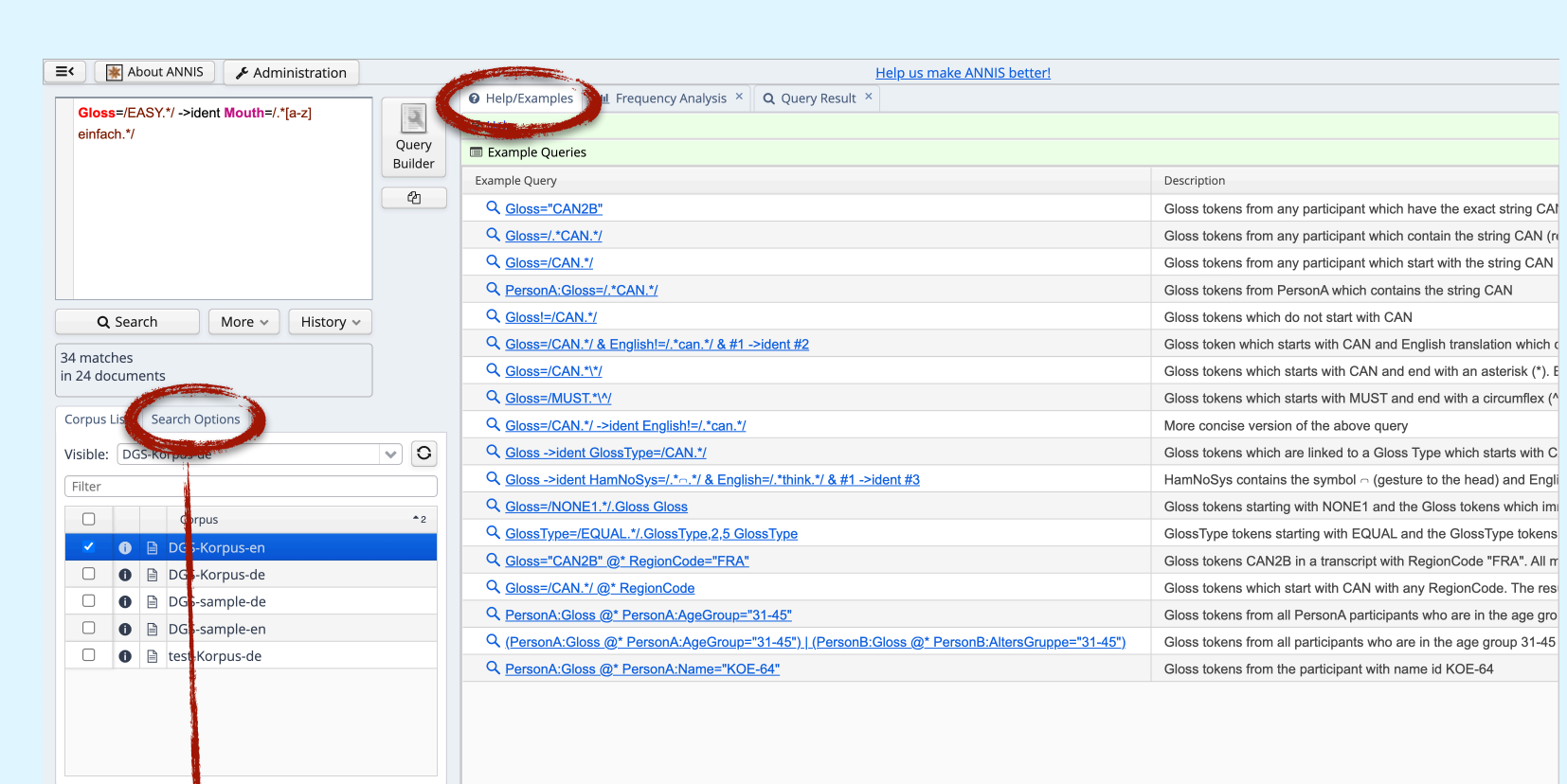
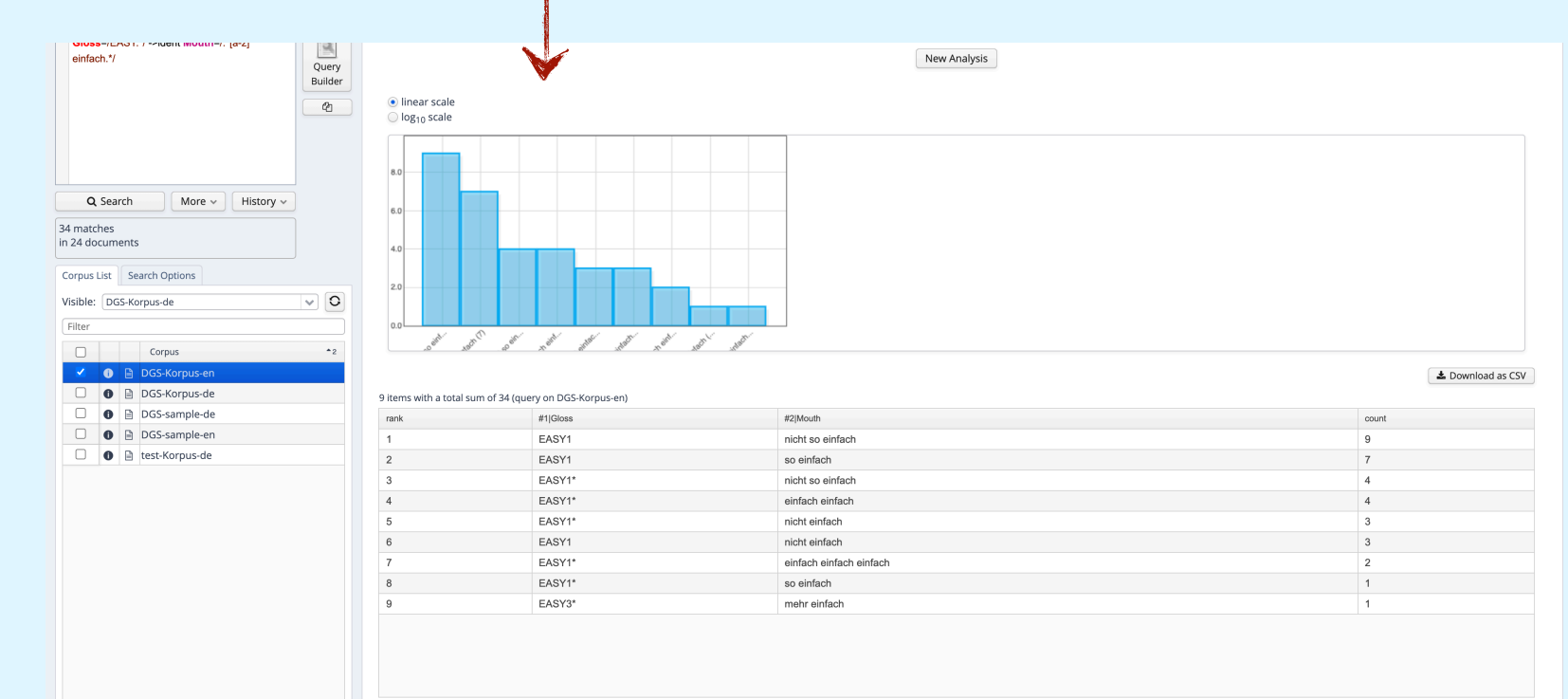
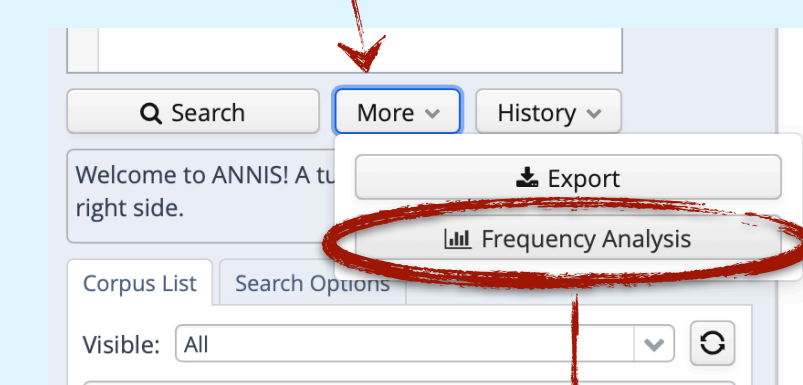
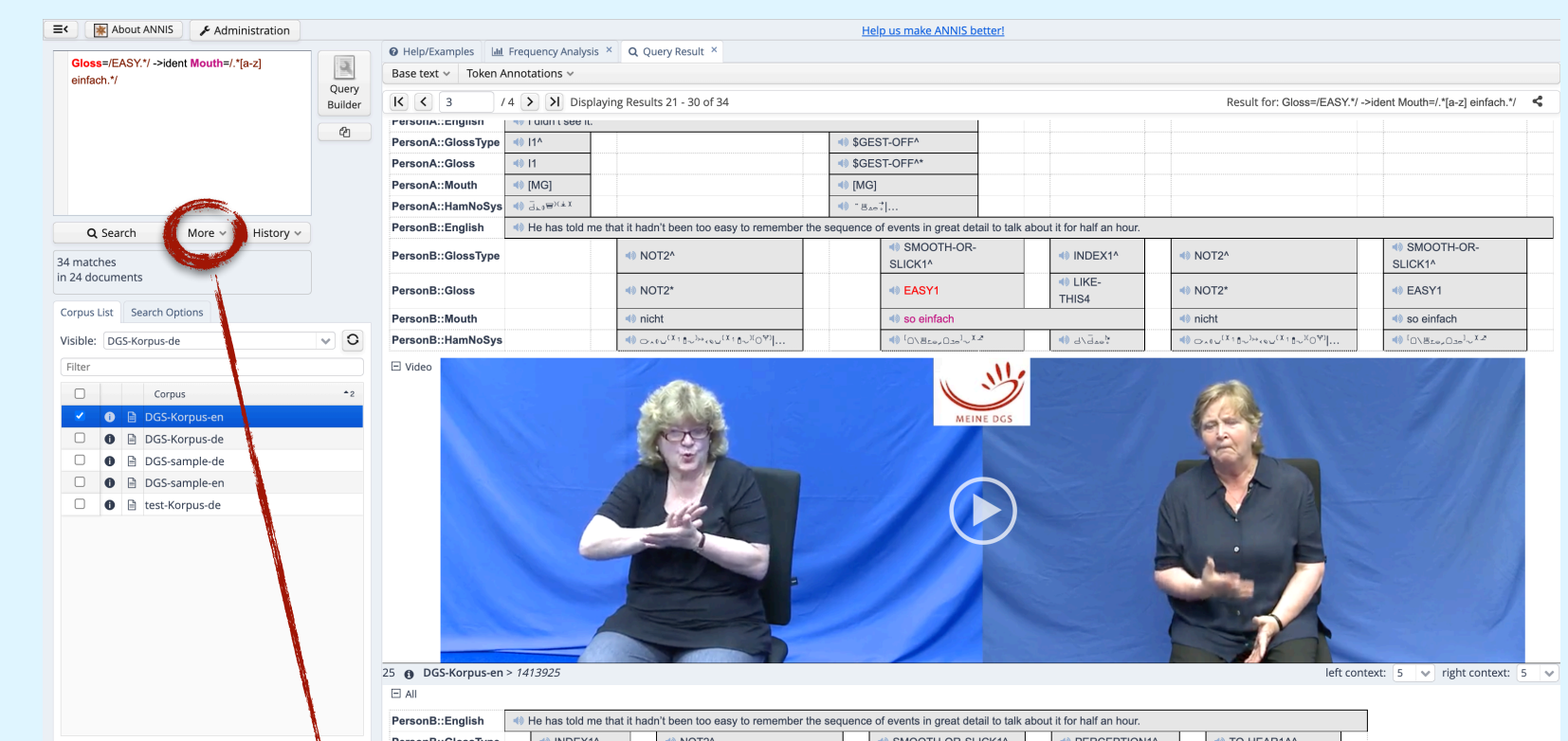


<https://ling.meine-dgs.de>

Release 4

### MY DGS – ANNIS

- Website in English, with annotations in German and English (DGS-Corpus-de, DGS-Corpus-en)
- 7 prioritised tasks (47.3 hrs.; translated & lemmatized)
  - + 4 retellings (0.1 hrs; translated & lemmatized)



<https://annis.meine-dgs.de>