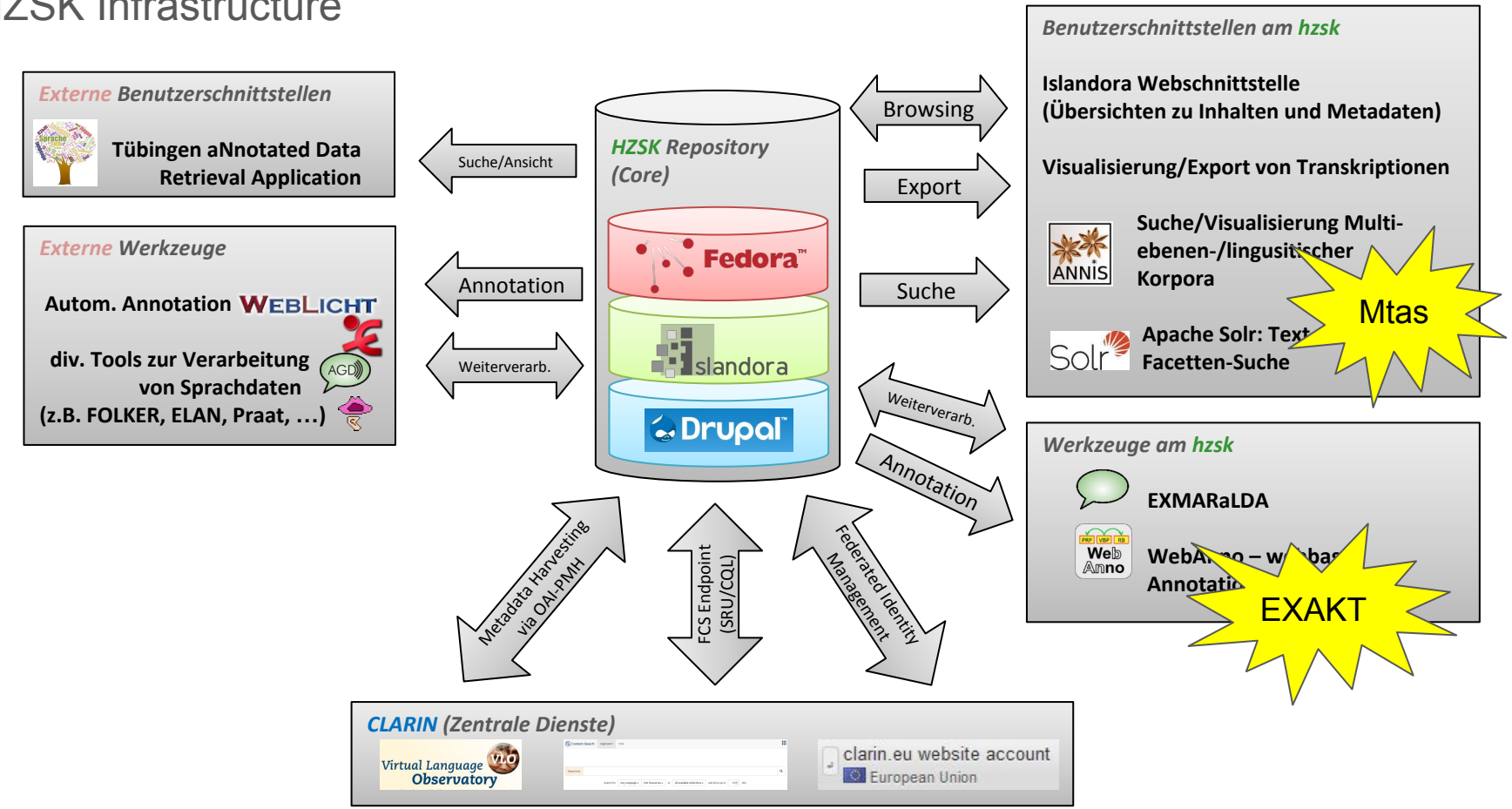
Experiences & Requirements

HZSK/INEL

MTAS Workshop, Meertens Instituut, Amsterdam, 25./26.09.2017

Anne Ferger, Tommi Pirinen, Daniel Jettka
HZSK / INEL, Universität Hamburg

HZSK Infrastructure



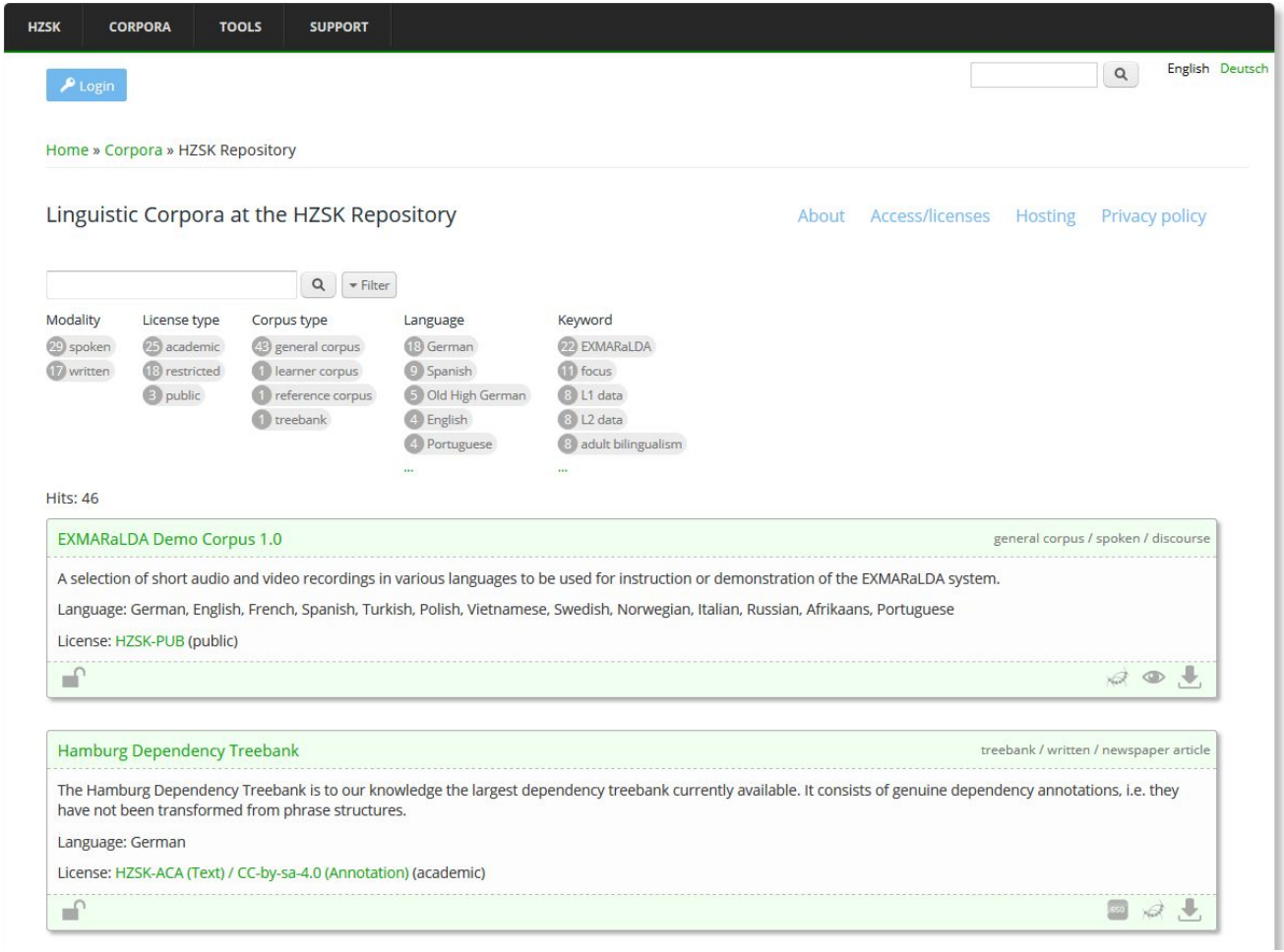
Search facilities

- EXAKT (Java, Desktop): EXMARaLDA corpora
- ANNIS
- SRU w/ MySQL: EXMARaLDA corpora (via CLARIN FCS)
- Solr: CMDI metadata

- DGD search (online): AGD corpora

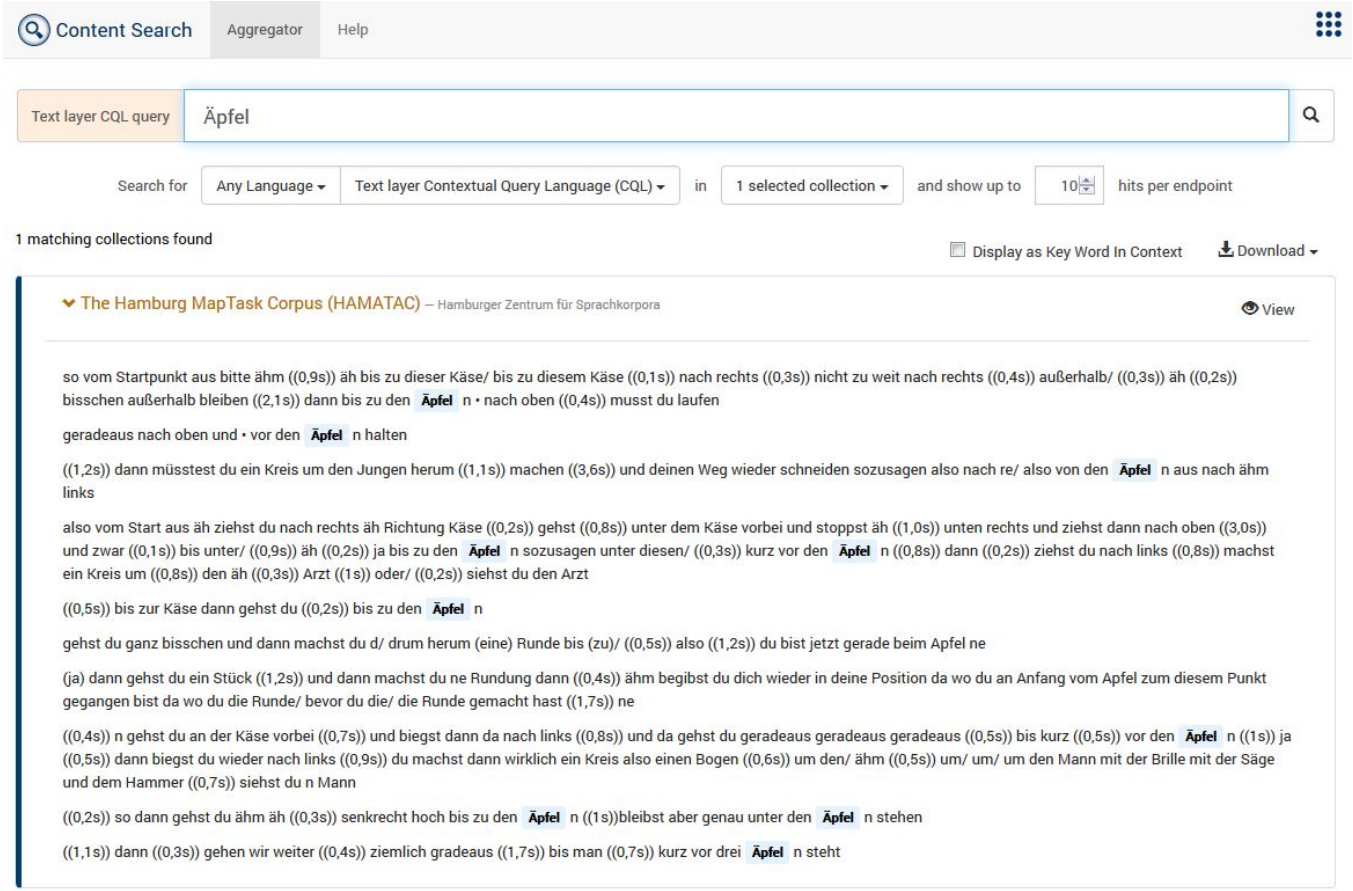
HZSK Repository

(Solr with CMDI)

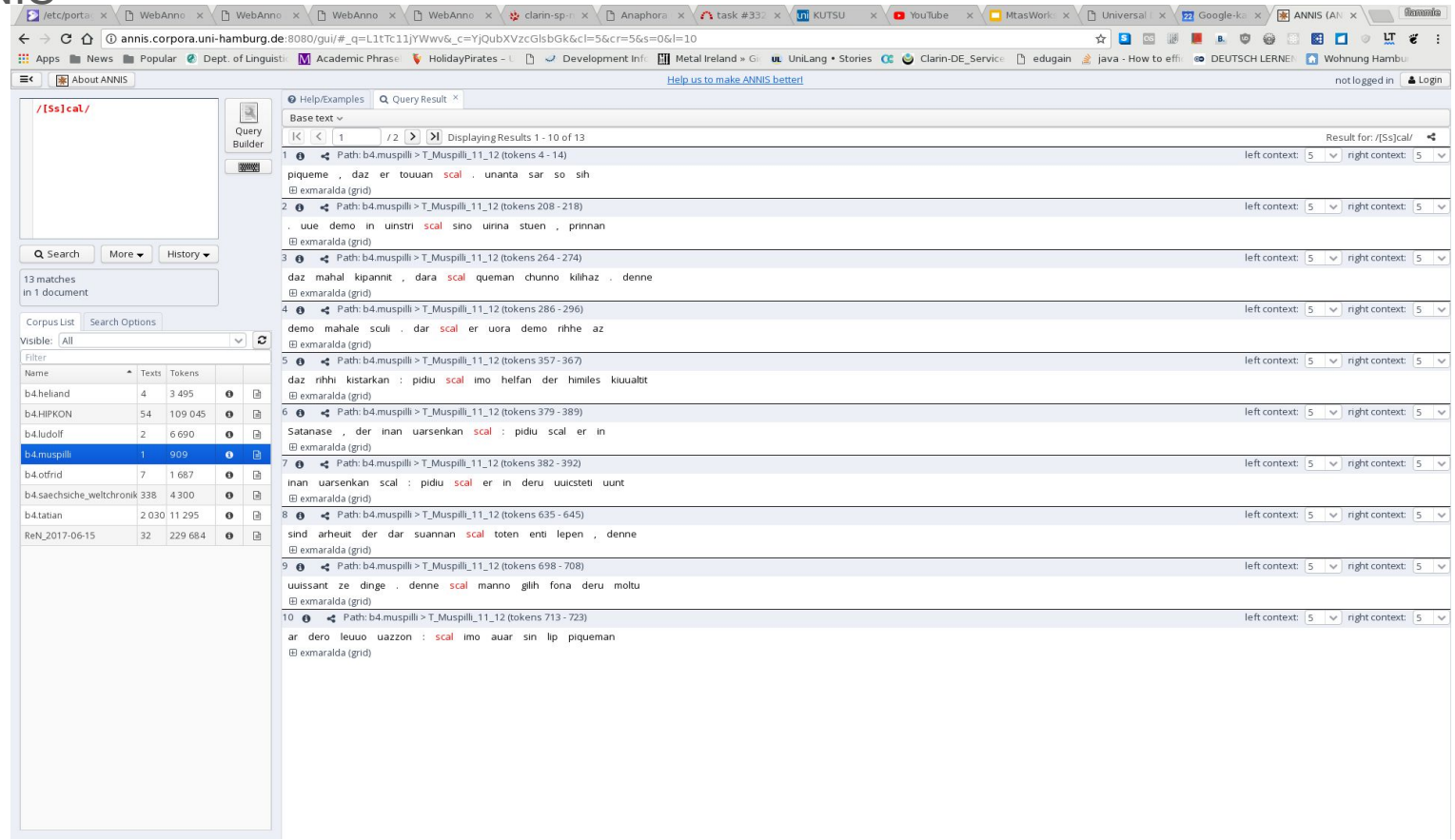


SRU

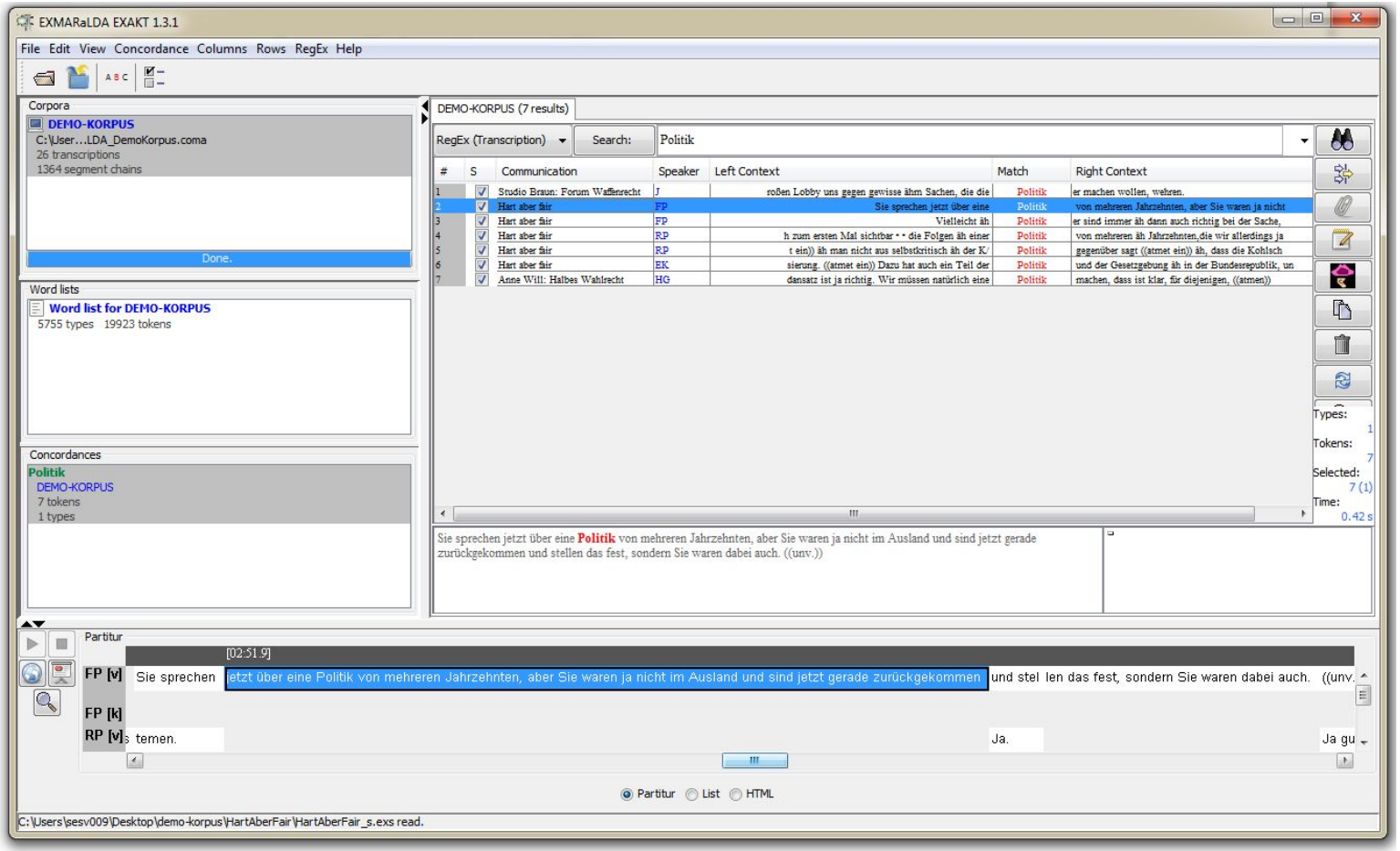
(CLARIN Aggregator)



ANNIS


annis.corpora.uni-hamburg.de:8080/gui/#_q=L1tTc11jYWwv&_c=YjQubXVzcGlsbGk&cl=5&cr=5&s=0&l=10
About ANNIS
Help us to make ANNIS better!
not logged in
Login
/[Ss]cal/
Query Builder
Search
More
History
13 matches
in 1 document
Corpus List
Search Options
Visible: All
Filter
Name
Texts
Tokens
b4.heliand 4 3 495
b4.HIPKON 54 109 045
b4.ludolf 2 6 690
b4.muspilli 1 909
b4.otfrid 7 1 687
b4.saechsische_weltchronik 338 4 300
b4.tatian 2 030 11 295
ReN_2017-06-15 32 229 684
Help/Examples
Query Result
Base text
Displaying Results 1 - 10 of 13
Result for: /[Ss]cal/
1 Path: b4.muspilli > T_Muspilli_11_12 (tokens 4 - 14)
piqueme , daz er touuan scal . unanta sar so sih
2 Path: b4.muspilli > T_Muspilli_11_12 (tokens 208 - 218)
. uue demo in uinstri scal sino uirina stuen , prinnan
3 Path: b4.muspilli > T_Muspilli_11_12 (tokens 264 - 274)
daz mahal kipannit , dara scal queman chunno kilihaz . denne
4 Path: b4.muspilli > T_Muspilli_11_12 (tokens 286 - 296)
demo mahale sculi . dar scal er uora demo rihhe az
5 Path: b4.muspilli > T_Muspilli_11_12 (tokens 357 - 367)
daz rihhi kistarkan : pidiu scal imo helfan der himiles kiuualtit
6 Path: b4.muspilli > T_Muspilli_11_12 (tokens 379 - 389)
Satanase , der inan uarsenkan scal : pidiu scal er in
7 Path: b4.muspilli > T_Muspilli_11_12 (tokens 382 - 392)
inan uarsenkan scal : pidiu scal er in deru uuicsteti uunt
8 Path: b4.muspilli > T_Muspilli_11_12 (tokens 635 - 645)
sind arheuit der dar suannan scal toten enti lepen , denne
9 Path: b4.muspilli > T_Muspilli_11_12 (tokens 698 - 708)
uuissant ze dinge . denne scal manno gilih fona deru moltu
10 Path: b4.muspilli > T_Muspilli_11_12 (tokens 713 - 723)
ar dero leuuo uazzon : scal imo auar sin lip piqueman
left context:
right context:
exmaralda (grid)

EXAKT


EXMARaLDA EXAKT 1.3.1
File Edit View Concordance Columns Rows RegEx Help
Corpora
DEMO-KORPUS
C:\User...LDA_DemoKorpus.coma
26 transcriptions
1364 segment chains
Done.
Word lists
Word list for DEMO-KORPUS
5755 types 19923 tokens
Concordances
Politik
DEMO-KORPUS
7 tokens
1 types
DEMO-KORPUS (7 results)
RegEx (Transcription)
Search:
Politik
S Communication Speaker Left Context Match Right Context
1 Studio Braun: Forum Waffenrecht J roßen Lobby uns gegen gewisse ähm Sachen, die die Politik er machen wollen, wehren.
2 Hart aber fair FP Sie sprechen jetzt über eine Politik von mehreren Jahrzehnten, aber Sie waren ja nicht
3 Hart aber fair FP Vielleicht äh Politik er sind immer äh dann auch richtig bei der Sache,
4 Hart aber fair RP h zum ersten Mal sichtbar * * die Folgen äh einer Politik von mehreren äh Jahrzehnten,die wir allerdings ja
5 Hart aber fair RP t ein)) äh man nicht aus selbstkritisch äh der K/ Politik gegenüber sagt ((atmet ein)) äh, dass die Kohlsch
6 Hart aber fair EK sierung. ((atmet ein)) Dazu hat auch ein Teil der Politik und der Gesetzgebung äh in der Bundesrepublik, un
7 Anne Will: Halbes Wahlrecht HG dansatz ist ja richtig. Wir müssen natürlich eine Politik machen, dass ist klar, für diejenigen, ((atmen))
Types: 1
Tokens: 7
Selected: 7 (1)
Time: 0.42 s
Sie sprechen jetzt über eine Politik von mehreren Jahrzehnten, aber Sie waren ja nicht im Ausland und sind jetzt gerade zurückgekommen und stellen das fest, sondern Sie waren dabei auch. ((unv.))
Partitur
[02:51.9]
FP [v] Sie sprechen etzt über eine Politik von mehreren Jahrzehnten, aber Sie waren ja nicht im Ausland und sind jetzt gerade zurückgekommen und stel len das fest, sondern Sie waren dabei auch. ((unv.
FP [k]
RP [v] temen. Ja. Ja gu
Partitur List HTML
C:\Users\sesv009\Desktop\demo-korpus\HartAberFair\HartAberFair_s.exs read.

HZSK/INEL Requirements

- Integration with Fedora/Islandora
- Audio/Video support (GUI)
- ISOtei
- CQL GUI libraries
- CLARIN requirements:
    a. FCS integration
    b. Shibboleth integration

User Requirements

- annotations following other annotations in different tiers
- sentence containing different annotations anywhere
- easy GUI
- KWIC table output
- musical score or other view of transcript at the match
- audio and video playback of match
- export or save search results

HZSK/INEL Experiences with Mtas

- relate the original word.id with the match (for jumping back into the transcriptions) ✓
- specific features of oral corpora (possibility to get the timepoints where the match can be played back)
- system requirements (re: sandbox installations)