

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

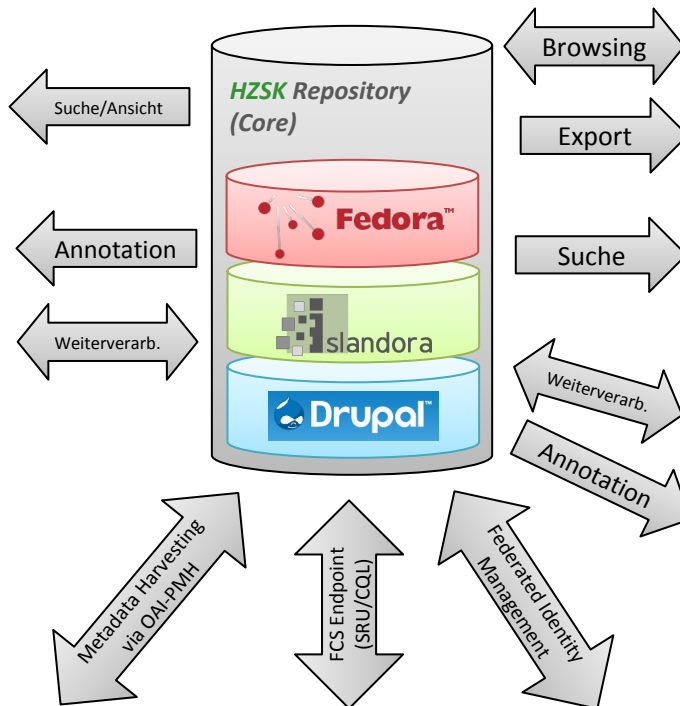
Externe Benutzerschnittstellen

 **Tübingen aNnotated Data Retrieval Application**

Externe Werkzeuge

Autom. Annotation **WEBLICHT**


div. Tools zur Verarbeitung von Sprachdaten
(z.B. FOLKER, ELAN, Praat, ...)




Benutzerschnittstellen am hzsk

Islandora Webschnittstelle
(Übersichten zu Inhalten und Metadaten)


Visualisierung/Export von Transkriptionen


 **Suche/Visualisierung Multi-ebenen-/linguistischer Korpora**

 **Apache Solr: Text Facetten-Suche**

Mtas

Werkzeuge am hzsk

 **EXMARaLDA**

 **WebAnno – webbasierte Annotation**

EXAKT

CLARIN (Zentrale Dienste)

 **Virtual Language Observatory**

 **clarin.eu website account**
European Union

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)

HZSKCORPORATOSTOOLSUPPORT

Login

EnglishDeutsch

Home » Corpora » HZSK Repository

Linguistic Corpora at the HZSK Repository

AboutAccess/licensesHostingPrivacy policy

QFilter

Modality

29 spoken

17 written

License type

25 academic

18 restricted

3 public

Corpus type

43 general corpus

1 learner corpus

1 reference corpus

1 treebank

Language

18 German

9 Spanish

5 Old High German

4 English

4 Portuguese

...

Keyword

22 EXMARaLDA

11 focus

8 L1 data

8 L2 data

8 adult bilingualism

...

Hits: 46





EXMARaLDA Demo Corpus 1.0

general corpus / spoken / discourse

A selection of short audio and video recordings in various languages to be used for instruction or demonstration of the EXMARaLDA system.

Language: German, English, French, Spanish, Turkish, Polish, Vietnamese, Swedish, Norwegian, Italian, Russian, Afrikaans, Portuguese

License: HZSK-PUB (public)







Hamburg Dependency Treebank

treebank / written / newspaper article

The Hamburg Dependency Treebank is to our knowledge the largest dependency treebank currently available. It consists of genuine dependency annotations, i.e. they have not been transformed from phrase structures.

Language: German

License: HZSK-ACA (Text) / CC-by-sa-4.0 (Annotation) (academic)



SRU

(CLARIN Aggregator)

Content Search

Aggregator

Help

Text layer CQL query

Äpfel

Search for

Any Language

Text layer Contextual Query Language (CQL)

 in

1 selected collection

 and show up to

10

 hits per endpoint

1 matching collections found

Display as Key Word In Context

Download

♥ The Hamburg MapTask Corpus (HAMATAC) – Hamburger Zentrum für Sprachkorpora

View

so vom Startpunkt aus bitte ähm ((0,9s)) äh bis zu dieser Käse/ bis zu diesem Käse ((0,1s)) nach rechts ((0,3s)) nicht zu weit nach rechts ((0,4s)) außerhalb/ ((0,3s)) äh ((0,2s)) bisschen außerhalb bleiben ((2,1s)) dann bis zu den **Äpfel** n • nach oben ((0,4s)) musst du laufen

geradeaus nach oben und • vor den **Äpfel** n halten

((1,2s)) dann müsstest du ein Kreis um den Jungen herum ((1,1s)) machen ((3,6s)) und deinen Weg wieder schneiden sozusagen also nach re/ also von den **Äpfel** n aus nach ähm links

also vom Start aus äh ziehst du nach rechts äh Richtung Käse ((0,2s)) gehst ((0,8s)) unter dem Käse vorbei und stoppst äh ((1,0s)) unten rechts und ziehst dann nach oben ((3,0s)) und zwar ((0,1s)) bis unter/ ((0,9s)) äh ((0,2s)) ja bis zu den **Äpfel** n sozusagen unter diesen/ ((0,3s)) kurz vor den **Äpfel** n ((0,8s)) dann ((0,2s)) ziehst du nach links ((0,8s)) machst ein Kreis um ((0,8s)) den äh ((0,3s)) Arzt ((1s)) oder/ ((0,2s)) siehst du den Arzt

((0,5s)) bis zur Käse dann gehst du ((0,2s)) bis zu den **Äpfel** n

gehst du ganz bisschen und dann machst du d/ drum herum (eine) Runde bis (zu)/ ((0,5s)) also ((1,2s)) du bist jetzt gerade beim Apfel ne

(ja) dann gehst du ein Stück ((1,2s)) und dann machst du ne Rundung dann ((0,4s)) ähm begibst du dich wieder in deine Position da wo du an Anfang vom Apfel zum diesem Punkt gegangen bist da wo du die Runde/ bevor du die/ die Runde gemacht hast ((1,7s)) ne

((0,4s)) n gehst du an der Käse vorbei ((0,7s)) und biegst dann da nach links ((0,8s)) und da gehst du geradeaus geradeaus geradeaus ((0,5s)) bis kurz ((0,5s)) vor den **Äpfel** n ((1s)) ja ((0,5s)) dann biegst du wieder nach links ((0,9s)) du machst dann wirklich ein Kreis also einen Bogen ((0,6s)) um den/ ähm ((0,5s)) um/ um/ um den Mann mit der Brille mit der Säge und dem Hammer ((0,7s)) siehst du n Mann

((0,2s)) so dann gehst du ähm äh ((0,3s)) senkrecht hoch bis zu den **Äpfel** n ((1s))bleibst aber genau unter den **Äpfel** n stehen

((1,1s)) dann ((0,3s)) gehen wir weiter ((0,4s)) ziemlich gradeaus ((1,7s)) bis man ((0,7s)) kurz vor drei **Äpfel** n steht

ANNIS

Web browser tabs: retc/port, WebAnno, WebAnno, WebAnno, WebAnno, clarin-sp, Anaphora, task #33, KUTSU, YouTube, MtasWorl, Universal, ANNIS (AI), ...

Address bar: annis.corpora.uni-hamburg.de:8080/gui/# q=L1tTc11jYWwv& c=YjQubXVzcGlzbgk&cI=5&cI=5&s=0&f=10

Navigation: Apps, News, Popular, Dept. of Linguistics, Academic Phrase, HolidayPirates, Development Info, Metal Ireland, UniLang, Stories, Clarin-DE_Service, edugain, java - How to effi, DEUTSCH LERNE!, Wohnung Hambu

Help/Examples Query Result

Base text

Path: b4.muspilli > T_Muspilli_11_12 (tokens 4 - 14)

Result for: /Ss]cal/

1 piqueme , daz er touuan scal , unanta sar so sih
[exmaralda (grid)

2 . uue demo in uinstri scal sino uirina stuen , prinnan
[exmaralda (grid)

3 daz mahal kipannit , dara scal queman chunno kilihaz , denne
[exmaralda (grid)

4 demo mahale sculi , dar scal er uora demo rihhe az
[exmaralda (grid)

5 daz rihhi kistarkan : pidu scal imo helfan der himles kiuaalit
[exmaralda (grid)

6 Satanase , der inan uarsenkan scal : pidu scal er in
[exmaralda (grid)

7 inan uarsenkan scal : pidu scal er in deru uucisteti uunt
[exmaralda (grid)

8 sind arheut der dar suannan scal toten enti lepen , denne
[exmaralda (grid)

9 uissant ze dinge , denne scal manno glih fona deru molu
[exmaralda (grid)

10 ar dero leuuu uazzon : scal imo auar sin lip piqueman
[exmaralda (grid)

Query Builder

Search

13 matches in 1 document

Corpus List Search Options

Visible: All

Name	Texts	Tokens		
b4.heland	4	3 495		
b4.HPKON	54	109 045		
b4.ludolf	2	6 690		
b4.muspilli	1	908		
b4.otfrid	7	1 687		
b4.saechische_welchronik	338	4 300		
b4.tatian	2 030	11 295		
ReN_2017-06-15	32	229 684		

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

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	<input checked="" type="checkbox"/>	Studio Braun: Forum Waffenrecht	J	roßen Lobby uns gegen gewisse ähm Sachen, die die	Politik	er machen wollen, wehren.
2	<input checked="" type="checkbox"/>	Hart aber fair	FP	Sie sprechen jetzt über eine	Politik	von mehreren Jahrzehnten, aber Sie waren ja nicht
3	<input checked="" type="checkbox"/>	Hart aber fair	FP	Vielleicht äh	Politik	er sind immer äh dann auch richtig bei der Sache,
4	<input checked="" type="checkbox"/>	Hart aber fair	FP	h zum ersten Mal sichtbar * - die Folgen äh einer	Politik	von mehreren äh Jahrzehnten, die wir allerdings ja
5	<input checked="" type="checkbox"/>	Hart aber fair	FP	t ein) äh man nicht aus selbstkritisch äh der K/	Politik	gegenüber sagt ((atmet ein)) äh, dass die Kohlsch
6	<input checked="" type="checkbox"/>	Hart aber fair	EK	sierung. ((atmet ein)) Dazu hat auch ein Teil der	Politik	und der Gesetzgebung äh in der Bundesrepublik, un
7	<input checked="" type="checkbox"/>	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 [M] Sie sprechen **jetzt 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 [M] temen.

Ja.

Ja gu

Partitur List HTML

C:\Users\psev009\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)