



The Virtual Language Observatory

Alexander König
CMDI workshop, Utrecht
2014-03-04



- VLO?
- What is behind it? – Relation to CMDI?
- How do I get my data in there?
- Demo + exercises

- Lots of resources somewhere out there:
 - Data collections
 - Corpora
 - Lexica
 - Grammars
 - Multimedia recordings
 - Software
 - Web applications / services
 - Old-school linguistic resources:
 - Books
 - Articles
 - CD-ROMs
- It's like a jungle, sometimes ...

- Metadata is everywhere
- Helping us to find things
 - A specific photo in our collection of thousands and thousands
 - The book we're looking for on Amazon
 - All Polish hip hop songs in our music collection

Metadata for music

Banshee Media Player

Media Edit View Playback Tools Help

Idle

Search your music

Now Playing

Play Queue 39

Libraries

Music 1967

- Recently Ad... 193
- Recently Play... 12
- Unheard 1775
- Audiobooks
- Videos
- Favorites
- Unwatched
- Podcasts
- Radio
- Online Media
 - Amazon MP3 Store
 - Miro Guide
 - Ubuntu One Music...
 - Internet Archive
 - Last.fm

All Artists (178)

- Aa'shiq al-Rasul
- Alan O'Bryant
- Alison Krauss
- Alison Krauss & Gillia...
- Andreas Dorau
- Aydee
- Bob Dylan
- Boris Gott
- Bracket
- Bran Van 3000
- Brett Dennen
- Cardigans
- Caro Emerald
- Nick Cave & The Bad S...
- Chris de Burgh
- Chris Thomas King
- Christian Jansen Band
- Clueso
- Colin Linden

All Albums (184)

- what's The S... Oasis
- 69 Love Songs The Magnetic...
- 69 Love Song... The Magnetic...
- A Thousand S... Linkin Park
- Abersowasvo... Dendemann
- Alles auf Einmal Ruben Cossani
- American III: S... Johnny Cash
- American IV: ... Johnny Cash
- Amputecture The Mars Volta
- An und für sich Clueso
- Besser zu zweit wingenfelder...
- Big Bad World Plain White T's
- Big Whoop godmanwho
- Bismillah Aa'shiq al-Rasul
- Bismillah Dawud Wharn...
- Bismillah Raihan
- Bismillah Yusuf Islam
- Bismillah Zain Bhikha
- Bismillah Ziad Sinnou
- Blood Money Tom Waits
- Blush Music Woven Hand
- Bring mich na... Wir sind Helden
- Bring mich na... Wir sind Helden
- Bulletproof La Roux
- Chocolate An... Ween
- Computers an... The Streets
- CONSOLERS ... Unknown Artist
- Cover Me Various Artists
- Crashkurs Son Goku
- Cyberspace a... The Streets
- Das sicherste ... Unknown
- Deichkind (Ar... Deichkind (Ar...
- Deleted Scen... Caro Emerald
- Demo Dawud Wharn...
- Der Mann, de... Unknown
- Die Pfütze de... Dendemann
- Donkey CSS
- Dortmund Song Aydee
- 9 Shes Electric
- 1 Roses
- 2 Love Is Like Jazz
- 3 When My Boy Walks Down the Street
- 4 Time Enough for Rocking When We're Old
- 5 Very Funny
- 6 Grand Canyon
- 7 No One Will Ever Love You
- 8 If You Don't Cry
- 9 You're My Only Home
- 10 (Crazy for You but) Not That Crazy
- 11 My Only Friend

Name	Artist	Album	Time
9 Shes Electric	Oasis	(what's The Story) Morning Glory	3:41
1 Roses	The Magnetic Fields	69 Love Songs	0:28
2 Love Is Like Jazz	The Magnetic Fields	69 Love Songs	2:57
3 When My Boy Walks Down the Street	The Magnetic Fields	69 Love Songs	2:38
4 Time Enough for Rocking When We're Old	The Magnetic Fields	69 Love Songs	2:04
5 Very Funny	The Magnetic Fields	69 Love Songs	1:27
6 Grand Canyon	The Magnetic Fields	69 Love Songs	2:29
7 No One Will Ever Love You	The Magnetic Fields	69 Love Songs	3:15
8 If You Don't Cry	The Magnetic Fields	69 Love Songs	3:07
9 You're My Only Home	The Magnetic Fields	69 Love Songs	2:18
10 (Crazy for You but) Not That Crazy	The Magnetic Fields	69 Love Songs	2:19
11 My Only Friend	The Magnetic Fields	69 Love Songs	2:01

1,967 songs - 4.5 days - 7.2 GB

Repeat Off

- Researcher: “where do I start?”
- Provide a single entry point giving access to all information
- Because of the large amount of data:
 - Drill-down paradigm (decrease search space gradually)
 - Multiple ways of exploring:
 - Full-text search
 - Facet browsing
 - Unified interface, links to the original context
 - <http://catalog.clarin.eu/vlo>

- A **simple** way to narrow down the search space, step by step
- Values offered are **dynamic**: they change with every previous selection made
- Purpose: **quickly navigating** through a huge amount of metadata

COUNTRY

[Netherlands](#) (15563)

[Germany](#) (8802)

[Japan](#) (4010)

[Belgium](#) (3956)

[Papua New Guinea](#) (3838)

[more...](#)

- Purpose: **quickly navigating** through a huge amount of resources
- Gives an idea of what is “out there”
- Useful too for **metadata curation**
- **Not** the tool to answer research questions!

VLO Faceted Browser (1)

CLARIN Virtual Language Observatory - Resources

CLARIN-D tutorial... x About opt-in and ... x eduGAIN current s... x About opt-in and ... x http://w...nce.cmdi x Taking Screenshot... x CLARIN Virtual La... x +

http://catalog.clarin.eu/ds/vlo/?jsessionid=9142406B23A992213594182B10FAD0CB?fq=genre:language_description&fq=language:Japa

MPI watch wiki taaltech clarin books tmp drupal PID dict CiteULike iGoogle Calendar Tel dev Bookmarks



Virtual Language Observatory

Explore the world of language resources and technology from different perspectives



VLO Home >> Faceted Browser Resources

search

COLLECTION

- [Graduate Institute of Applied Linguistics Library](#) (26)
- [Centre de Ressources pour la description de l'Oral \(CRDO-Aix\)](#) (4)
- [OLAC Metadata Providers](#) (4)
- [Speech and Language Data Repository](#) (4)
- [SIL Language and Culture Archives](#) (1)
- [WALS Online](#) (1)

CONTINENT

COUNTRY

LANGUAGE : [all](#) > Japanese

GENRE : [all](#) > language_description

SUBJECT

- [japanese language--phonology](#) (6)
- [japanese language--grammar](#) (4)
- [morphology](#) (3)
- [syntax](#) (3)
- [japanese language](#) (2)
- [japanese language--grammar, generative](#) (2)
- [japanese language--textbooks for foreign speakers--english](#) (2)

Showing 1 to 10 of 40

<< < 1 2 3 4 > >>

Results

- [A generative phonology of standard colloquial Japanese](#)
- [An introduction to Japanese phonology](#)
- [Complex verbs and the lexicon](#)
- [Contrastive studies Hungarian-Japanese](#)
- [Essential Japanese grammar](#)
- [Issues in Japanese linguistics](#)
- [Japanese](#)
- [Japanese Digit String](#)
- [Japanese Digit String](#)
- [Japanese Digit String](#)

<http://catalog.clarin.eu/ds/vlo>

Field	Value
name	kleve-route
description	This recording was made to generate a freely available test resource including speech and gestures. The annotations were created by Peter and Kita who is gesture researcher at the MPI for Psycholinguistics., Diese Aufnahme wurde erzeugt, um eine frei verfügbare Test Resource zur Verfügung stellen zu können, die Sprache und Gestik umfasst. Die Annotationen wurden von Peter und Kita, dem Gestik Researcher am MPI erzeugt.
language	English
country	Netherlands
continent	Europe
year	2002-10-30
id	test-hdl:1839/00-0000-0000-0009-294C-9
genre	unspecified
organisation	Max Planck Institute for Psycholinguistics
origin	MPI corpora
resourceType	video, video, audio, image, image, annotation

Resources:



<hdl:1839/00-0000-0000-0009-294E-5>



<hdl:1839/00-0000-0000-0009-294F-C>



<hdl:1839/00-0000-0000-0009-2950-E>

- Metadata analyzed is CMDI format
- Metadata sources
 - CMDI files harvested from CLARIN centres
 - CMDI'fied OLAC records (from CLARIN centres and others)
 - CMDI'fied LRT inventory records
- You can get to resources directly from search results

- <http://catalog.clarin.eu/vlo/>
- **Find some resources in the catalogue:**
 - Corpus Gysseling
 - Telephone conversation recordings in Nepal

- **Find some resources in the DoBeS archive of endangered languages which are:**
 - (Spoken) discourse with at least two consultants in Asia
 - Or (spoken) discourse with at least two consultants in a Face to Face conversation

- Inherent limit: Simple search
 - no OR combinations possible
 - no sophisticated search operations

- Links to [language](#) information: WALS, Wikipedia, Ethnologue, LinguistList ... and the VLO
- Descriptions in the record listing
- National Project facet
- Feedback link

- A faceted browser is as good as its data, so further curation steps are needed
- more CMDI metadata
- some more facets e.g.: year
- Interface improvements

- ... ask them now
- ... or send a mail to vlw@clarin.eu
- More information:
 - www.clarin.eu/vlo
 - www.clarin.eu/cmdl



