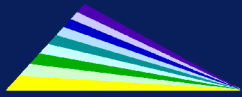




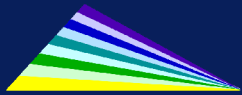
# Language Archiving at the MPI for Psycholinguistics

Paul Trilsbeek  
MPI for Psycholinguistics  
DoBeS Archive

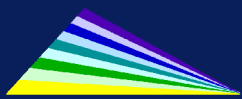
---



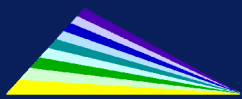
- Institute in Nijmegen, the Netherlands
- 4 research groups:
  - Language Acquisition
  - Language Comprehension
  - Language & Cognition (“field linguists”)
  - Language Production
- Large technical group: support & software development



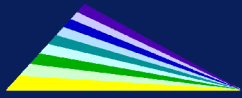
- Started creating a digital archive about 10 years ago for a number of reasons:
  - Preventing “data cemeteries”: central, well organized data archive with rich metadata descriptions
  - Taking care of long-term preservation of data: continuous migration to latest storage technology, making use of “archivable” file formats, data distribution
  - Data access: Making data discoverable and accessible in various ways over the internet



- Contains a large number of languages:
  - Languages from all over the world studied by MPI field linguists
  - First and second language acquisition studies
  - Endangered languages documented for the VolkswagenStiftung DoBeS programme
  - Spoken Dutch corpus
  - Sign languages
  - etc.



- Kinds of data:
  - video recordings
  - audio recordings
  - images
  - transcriptions/annotations
  - lexica
  - various other types of written documents
- All described using IMDI metadata scheme



# MPI “LAT” Archiving Framework & Tools

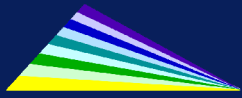
Preparation



**Primary resources**



**ELAN/LEXUS/SYNPATHY** Annotation, Lexicon & Syntax



# MPI "LAT" Archiving Framework & Tools

## Preparation



**Primary resources**



**ELAN/LEXUS/SYNPATHY** Annotation, Lexicon & Syntax

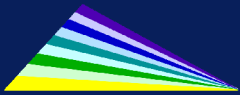
## Integration



**IMDI:** Metadata descriptions, data organization



**LAMUS:** Data Uploading and Management, Access Management



# MPI "LAT" Archiving Framework & Tools

## Preparation



**Primary resources**



**ELAN/LEXUS/SYNPATHY** Annotation, Lexicon & Syntax

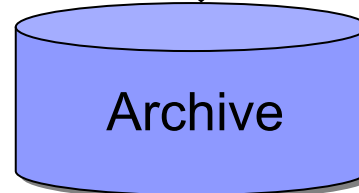
## Integration



**IMDI:** Metadata descriptions, data organization



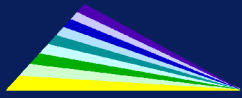
**LAMUS:** Data Uploading and Management, Access Management



Archive

>30 TeraBytes  
>10.000 hours A/V





# MPI "LAT" Archiving Framework & Tools

## Preparation



**Primary resources**



**ELAN/LEXUS/SYNPATHY** Annotation, Lexicon & Syntax

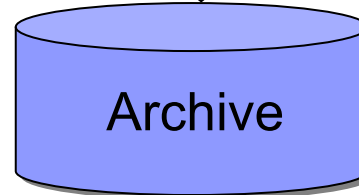
## Integration



**IMDI:** Metadata descriptions, data organization



**LAMUS:** Data Uploading and Management, Access Management



Archive

>30 TeraBytes  
>10.000 hours A/V

## Utilization



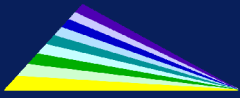
**IMDI/GIS:** Metadata Browsing & Searching, Google Earth



**ANNEX/LEXUS/SEARCH:** Complex Access via the Web

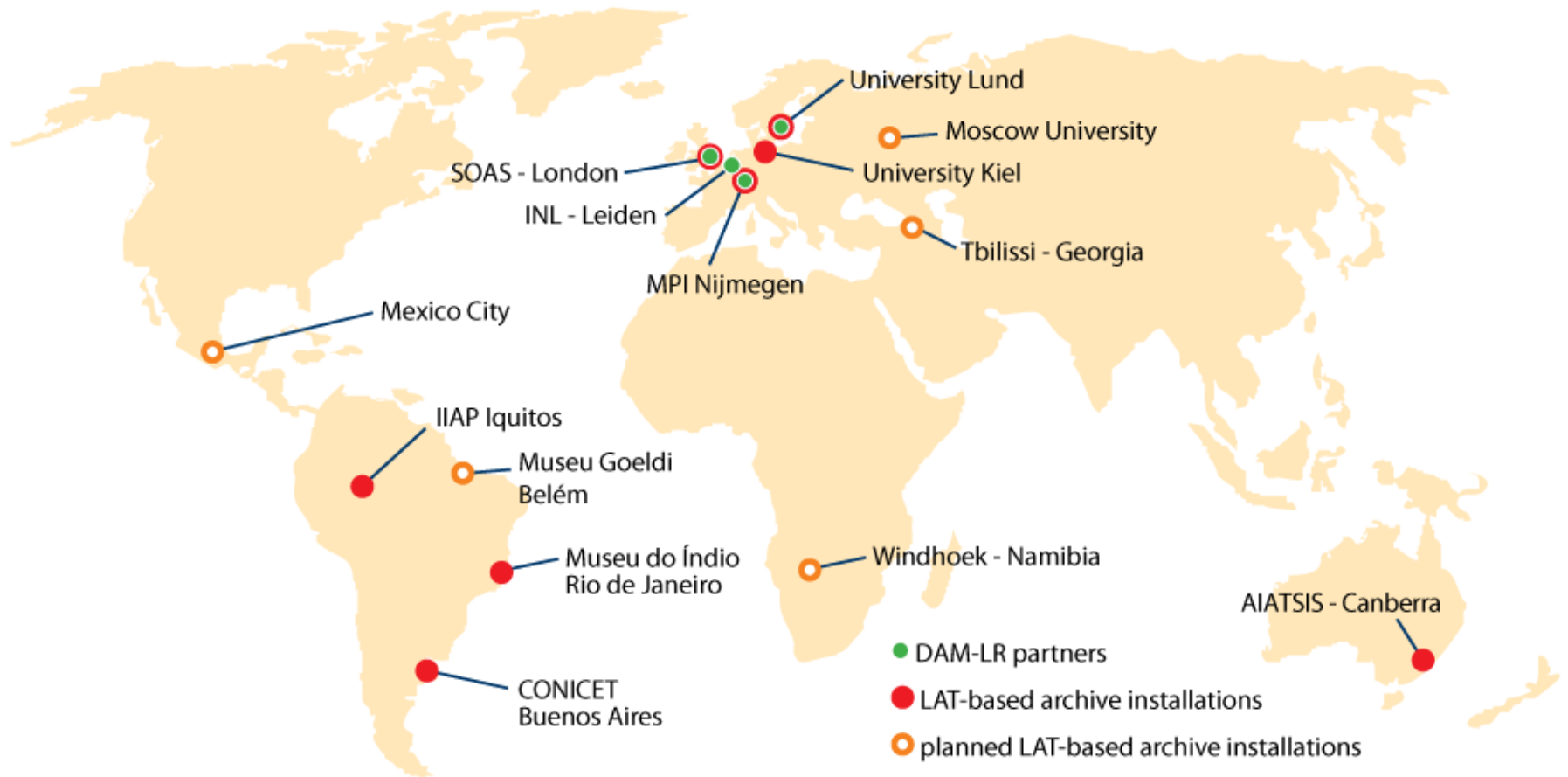


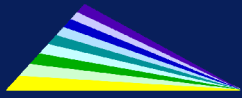
**ADDIT/VICOS:** Enrichments/Views



# Grid of Language Archives

## Distribution of MPI “LAT” archive installations

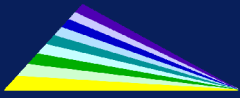




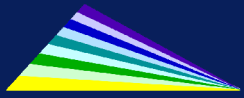
## Services Offered

- Digitizing of linguistic data
- Archiving of linguistic data using web-based upload and management system
- Installing LAT archiving server
- Giving training courses on archiving and linguistic tools

(all depending on available capacity and priorities)

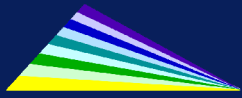


- DAM-LR: Distributed Access Management for Language Resources  
[www.dam-lr.eu](http://www.dam-lr.eu)
- DELAMAN: Digital Endangered Languages and Musics Archives Network  
[www.delaman.org](http://www.delaman.org)
- CLARIN: Common Language Resources and technology Infrastructure  
[www.clarin.eu](http://www.clarin.eu)



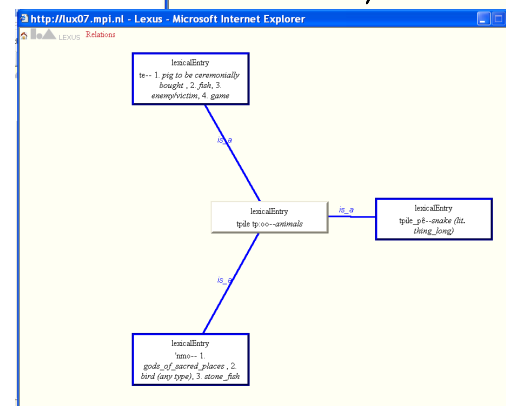
- Archive browser & metadata search
- Content search & annotation viewer
- Google Earth overlay & lexicon tool



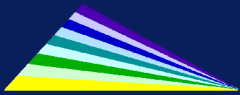


## More Information

- [www.lat-mpi.eu](http://www.lat-mpi.eu)
- [www.mpi.nl/dobes](http://www.mpi.nl/dobes)







# ANNEX Annotation Viewer

<http://corpus1.mpi.nl/ds/annex/runLoader?jsessionid=F0593...36656D81E0AEE99EB82B12&transferuid=1&nodeid=MPI533878%23>
<http://corpus1.mpi.nl/ds/annex/runLoader?jsessionid=F05935103936656D81E0AEE99EB82B12&>

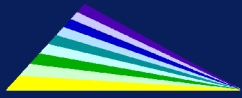
ANNEX
Workspace
user: anonymous
logout

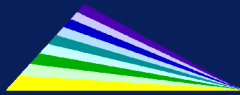
File: phidang\_talk.eaf

madum akka paīyaṅ khasiṅna?ā ah hani ah yoba ayuiyaktihē ah multēi yoi  
ayuyaktihē tele tola ayuiyaktihē yoba khimce baddhe uyuwakte bhai?nina  
maiyuṅyaktakteni naṅ bago ah bhai?ni athabihē athabihē athabihē yoba utti  
uttikhera bagoi? ā bakhai? bai? ā ā ā yoba dui nambar wadbe hungo teī teli  
tola waliṅ tolabe?ṅa bhai?ni athabi garda kheri bakhai? them them pukuwakte

S 00:00:00.000  
00:00:00.000 - 00:00:00.000 00:00:00.000

tx@SR	tele	tola	ayuiy	yob	khi	ba	uyu	bhai	maiu	naṅ	bago	ah
gw@SR	tele	tola	ayuiy	yob	khi	ba	uyu	bhai	maiu	naṅ	bago	ah
mph@SR	tele	to	a			k	c	ba			bago	ah
mgl@SR	a	cla	s	-				h	-	ver	DEM	FIL
lg@SR	C	N						C	-	C	C	C
rt@SR												
eng@SR	You	were	stayin			there	were	man			this	You
nep@SR	टेलेटोलमा	बखुहुन्थ्यो	पर	धेरै	घरहरू	थिए					यो !	यता

[illegible]



# ANNEX Content Search

ANNEX Annotation Content Search

http://corpus1.mpi.nl/ds/annex/search.jsp?sessionId=B1455ED460AE826016F920E5CF686A77&tr Google

ANNEX SEARCH help user: pautri@mpi.nl logout

Substring Search **Single Layer Search** Multiple Layer Search

**Domain:** ,â†Akhoe Hai//om  
☒ EAF<sup>45</sup>

**Query History:** < >

**Mode:** N-gram within annotation | case insensitive | exact match

**Find** # ge All Tiers

#hits : 830  
#annotations with a hit : 704  
#annotations investigated : 24871

**Ready**

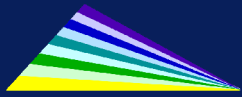
>

frequency 1 - 23 of 263

Annotation	Percentage	Count
i ge	29.97%	211
ta ge	5.26%	37
tita ge	3.27%	23
nëb ge	1.99%	14
naub ge	1.28%	9
khoes ge	1.28%	9
khoeb ge	1.28%	9
llnāb ge	0.99%	7
ob ge	0.99%	7
dī ge	0.99%	7
lgóas ge	0.85%	6
du ge	0.85%	6
llnās ge	0.71%	5
lgaus ge	0.71%	5
xūs ge	0.71%	5
gu ge	0.71%	5
llnas ge	0.57%	4
tsūs ge	0.57%	4
nēsas ge	0.57%	4
ii ge	0.57%	4
llnati ge	0.43%	3
lhūs ge	0.43%	3
ös ge	0.43%	3

Applet SearchApplet started





# LEXUS Lexicon Tool

Lexus

Back Forward Reload Stop Home <http://lux07.mpi.nl/mpi/lexus/lexiconOverview.htm> Google

Lexus — Language Archiving T... Lexus

LEXUS Word List View Search Relation types Help demo

## lexicon: Word list

Results 201-225 of 351 for all entries

<< first < previous Page 9 next > last >>

Lexical Entry
<b>mywongo</b> cuttle fish
<b>ndapí nté pipí pyu</b> bird sp, ? <i>Dicaeum pectorale rosseli</i>
<b>ndapí nté pipí pyu</b> fish sp. (a wrasse, ? <i>Slxbar wrasse</i> , <i>Thalasomma hardwicke</i> )
<b>ndéédí</b> bird sp. 'black pigeon' ( <i>Ducula pinon salvadorii</i> )
<b>ndéla chedé</b> fish type (Butterfly type e.g. <i>Chaetodon baronessa</i> )
<b>ndiki we</b> fish sp. (Orbicular cardinal fish <i>Spaeramia orbicularis</i> ??)
<b>ndikiwe</b> fish sp. Orbicular cardinalfish
<b>ndíla chelé</b> fish sp, butterfly fish
<b>'ne ké</b>
<b>'ne skirt</b>
<b>'ne</b> your fish (2Poss form of te) - probably should be 'né
<b>'né</b> your fish (2Poss of te) Spelling uncertain, but contrasts with 'ne 'skirt'.
<b>'n:ee</b> fish sp
<b>ngépa</b> call an animal to you
<b>ng:il</b> crying flock of birds
<b>ng:il</b> traditional medicine
<b>nípa pu</b> fish sp, ?giant flathead
<b>njépi</b> fish sp, Yellow margin Triggerfish ( <i>Pseudobalistes flavimarginatus</i> )
<b>njéte we</b> fish sp, small spotted grouper, ? <i>Ephinephelus merra</i>
<b>njini</b> fish sp
<b>njini</b> trap door spider
<b>njini</b> tree sp. <i>Castanopsis acuminatissima</i>

**ndéla chedé** (N)

fish type (Butterfly type e.g. *Chaetodon baronessa*)

