

# Multilevel Annotation of Historical Documents

Cristina Vertan

[cristina.Vertan@uni-hamburg.de](mailto:cristina.Vertan@uni-hamburg.de)



Handwritten text in a cursive script, likely from a historical document or manuscript.

አኑሃ፡ለአሙ፡ወሰዳግ፡ዮዕቆብ፡
ለራሱል፡ወጸርኝ፡በቃሉ፡ወበ
ክዩ፡ወዳድአ፡ለራሱል፡ካመ፡
ወልጁ፡እገቱ፡ለላባ፡ወእቱ፡ወ
ክመ፡ወልጁ፡ርብቃ፡ወእቱ፡
ወሮጸት፡ራሱል፡ወአጸድዳዲ
ለአቡሃ፡ዘጸት፡ነገራ፡ወሰባ፡
ስመዳ፡ግቃ፡ስመ፡ዮዕቆብ፡ወ
ልጁ፡ርብቃ፡እገቱ፡ሮጸ፡ወተቀ
ቢሎ፡ወሐቀሮ፡ወሰዳግ፡ወወሰ
ድ፡ቢቶ፡ወነገሮ፡ለላባ፡ካሉሎ፡
ዘጸት፡ነገራ፡ወደቢሎ፡ለባ፡
ዮዕቆብ፡እመኑ፡ዐጸዎ፡ወእመ
ኑ፡ሠጋግዮ፡አጸት፡ወነገራ፡መስሊ
ሁ፡ሠላሳ፡መዋዕል፡።
ወደቢሎ፡ለባ፡ለዮዕቆብ፡እ
ስመ፡አኑዮ፡አጸት፡አተቀቀኝ፡
ለተ፡በክ፡፤ግረ፡፤ዐስባክ፡መ
ጸት፡ወእቱ፡ወቦቱ፡ለላባ፡ክል
ኢ፡አዋልጁ፡ስግ፡ለአጸት፡ቶል



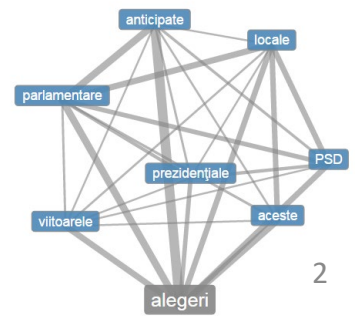
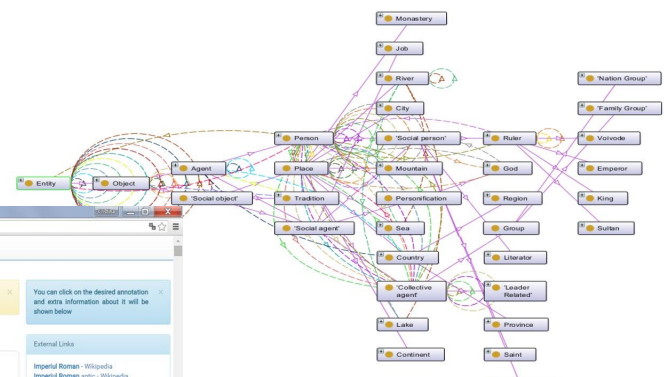
kerpens, Kynle, canis &c. Im plurali feken sie
hinten an die Wörter; die eine lebendige Sache bes
deuten, den Artikel ij; als: Căij, Oamenij, equi,
homines: leblose Kreaturen aber endigen sich im Plu
rali auf ele, als Scamele, Văsele, u. s. w. Auch
haben die Moldauer zween Articulus mulmini ge
neris, e und a, als: mujere, gaima, furnini, galli
na. Wörter, die sich auf e endigen, haben im Plu
rali ile, als: mujere, mujerile, die sich aber auf a
endigen, haben im Plurali ele, als gaima, gaimele.

3) Kan man vielleicht wahrscheinlicher mutmaß
fen, daß diejenigen Wörter, welche mehr mit der Ita
liänischen als mit der alten römischen Sprache überein
kommen, aus dem langen Umgange, welchen die Mol
dauer mit den Senuestern während ihres Besizes der
Küsten des schwarzen Meers hatten, sich in unsere
Sprache mit eingeschlichen haben.

Denn auf gleiche Weise haben die Moldauer,
nachdem sie mit den Griechen, Türken und Polzen
zu handeln anfingen, auch Wörter aus der Sprache
dieser Bölker in die ihre aufgenommen; zum Exem
pel, von den Griechen Pedepja, παιδευσις, Kivernis
jve, κιβήρισσις, Procopie, προκοπή, Blafsem, βλασφημία,
azyma, άζυμοι, Drum, δρέμυς, Pizma, πιζμα. Da
mir nun also beyder Partheyen Meinungen vorgetra
gen haben, so getrauen wir uns nicht zu bestimmen,
welche von beyden der Wahrheit am nächsten sey? aus
Surcht,

```
<phrase>
  <ne> <person personT="f" first=",Dragosch"
    relationT=",son" refT=",w123"/> </ne>
  <phraseParts constituentT="NP" mult="false">
    <tok idT="" langT="ro" styleTranscr="de">
      <pos posT="N"/>
      <string> Dragosch</string>
    </tok></phraseParts>
  </phrase>
  , ein Sohn ihres
  <phrase>
    <ne> <person personT="tl"/></ne>
    <phraseParts constituentT="NP" mult="true">
      <tok idT="" langT="de" styleTranscr="de">
        <pos posT="N"/>
        <string> Königs </string>
      </tok>
      <tok idT="w123" langT="ro" styleTranscr="de">
        <pos posT="N"/>
        <string> Bogdan</string>
      </tok> </phraseParts>
    </phrase>
```

Web interface for semantic annotation of the text 'Dimitrie Cantemir, Descrierea Moldovei'. It shows a list of semantic classes on the left, the annotated text in the center, and a sidebar with external links and class information on the right.



# Annotation Particularities

- Often several Layers of Annotation (e.g. linguistic, editorial, text structure, domain specific). Annotation layers are sometimes interconnected
- Sometimes synchronisation between different text variants (e.g. original, transliteration, translation)
- Non-continuous annotation segments
- Changes on the base text during the annotation required
- Often more linguistic categories as for modern data
- Need of user-friendly annotation interfaces
- Modular Architecture flexible at changes (new layers, new annotation categories)
- Often need of manual annotation

# TraCES - From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages

- ERC Advanced Grant 2014–2019
- **aim:** reliable and extensive linguistic **data** based on annotated texts for a **diachronic analysis** of Gəʿz (lexicography, morphology and style)
- **corpus:** several texts belonging to **different periods and genres** of Ethiopic literature (**text-critical editions**)

Ge'ez Text –  
Example  
from Genesis  
29.11

እኑሃ፡ለእመ፡ወሰዳ፡ዎሶቆ፡ከ፡  
ለራሔል፡ወደርኝ፡በቀሉ፡ወበ  
ክዩ፡ወዳድኦ፡ለራሔል፡ከመ፡  
ወልደ፡እኅቱ፡ለላባ፡ወእቱ፡ወ  
ከመ፡ወልደ፡ርከቀ፡ወእቱ፡  
ወሮክት፡ራሔል፡ወአጸደዳ፡  
ለአቡሃ፡ዘኝተ፡ከገረ፡ወሰባ፡  
ሰዎዳ፡ፊቀለ፡ከመ፡ዎሶቆ፡ከ፡  
ልደ፡ርከቀ፡እኅቱ፡ሮኦ፡ወተቀ  
በሎ፡ወሐቀሮ፡ወሰዳዎ፡ወወሰ  
ደ፡ቤቶ፡ወከገሮ፡ለላባ፡ከሎ፡  
ዘኝተ፡ከገረ፡ወደቤሎ፡ለባ፡ለ  
ዎሶቆ፡እዎኑ፡ዐጽዎሃ፡ወእዎ  
ኑ፡ሠጋሃ፡አኝተ፡ወከበረ፡ዎሰሌ  
ሁ፡ሠላሳ፡መዋሶለ፡  
ወደቤሎ፡ለባ፡ለዎሶቆ፡እ  
ከመ፡እኑሃ፡አኝተ፡ኢተተቀኝደ፡  
ለተ፡በከ፡ኝግረኑ፡ዐሰበከ፡ዎ  
ኝተ፡ወእቱ፡ወቦቱ፡ለላባ፡ክል  
ኤ፡አዋልደ፡ከግ፡ለእኝተ፡ተል

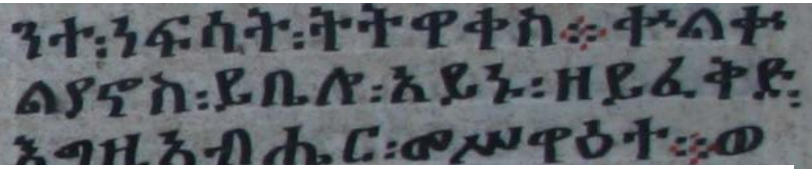


Word Separator

Section mark



# The Classical Ethiopic (Gə‘əz)



- Each symbol represent a syllable (Fidäl)
- Wowels are represented.
- LR-Writting
- Dedicated symbols for Digits
  - For a number representation and value fora number are different
  - 𐌀𐌁 Transcription 1 wa 5 Value 6 (1+5)
- Dedicated interpunction: (Space) ∷ (. or ,) ∷ (,) ∷ (:) ∷ (;) ∷ (?) ∷ (Paragraph)
- Laryngals and Sibilants are changeblae from the phonemisc and graphemic point of view.
  - ሃሌሉያ (häleluyā)
  - ሀሌሉያ ሐሌሉያ ጎሌሉያ ሓሌሉያ ኃሌሉያ
- Nonconcatenative Morphology

Schrifttabelle

	a	u	i	ā	e	a/ø	o
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
ḥ	ሐ	ሑ	ሒ	ሓ	ሔ	ሐ	ሑ
m	መ	ሙ	ሚ	ማ	ሜ	ሞ	ሟ
ś	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
r	ረ	ሩ	ሪ	ራ	ሮ	ሮ	ሾ
s	ሰ	ሱ	ሲ	ሳ	ሴ	ሰ	ሱ
q	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ
b	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
t	ተ	ቱ	ቲ	ታ	ቲ	ት	ቶ
ḥ	ተ	ቱ	ቲ	ታ	ቲ	ት	ቶ
n	ነ	ኑ	ኒ	ና	ኔ	ን	ኖ
ʾ	አ	ኡ	ኢ	ኣ	ኤ	አ	ኡ
k	ከ	ኩ	ኪ	ካ	ኤ	ከ	ኩ
w	ወ	ዉ	ዊ	ዋ	ዌ	ወ	ዉ
c	ዐ	ዑ	ዒ	ዓ	ዔ	ዐ	ዑ
z	ዘ	ዙ	ዚ	ዛ	ዜ	ዘ	ዙ
y	የ	ዮ	ዹ	ያ	ዼ	ዩ	ዮ
d	ደ	ዱ	ዲ	ዳ	ዴ	ደ	ዱ
g	ገ	጑	ጒ	ጓ	ጔ	ገ	጑
t	ጠ	ጡ	ጢ	ጣ	ጤ	ጠ	ጡ
p	አ	አ	አ	አ	አ	አ	አ
ś	አ	አ	አ	አ	አ	አ	አ
q	ፀ	ፁ	ፂ	ፃ	ፄ	ፀ	ፁ
b	ፈ	ፉ	ፊ	ፋ	ፅ	ፈ	ፉ
f	ፎ	ፏ	ፐ	ፑ	ፒ	ፆ	ፏ
q <sup>w</sup>	ቁ		ቁ	ቁ	ቁ	ቁ	ቁ
ḥ <sup>w</sup>	ኅ		ኅ	ኅ	ኅ	ኅ	ኅ
k <sup>w</sup>	ከ		ከ	ከ	ከ	ከ	ከ
g <sup>w</sup>			ገ	ገ	ገ		

2018.05.2018

√ qds × 'asta1a22a3a → 'astaqaddasa

# Language particularities

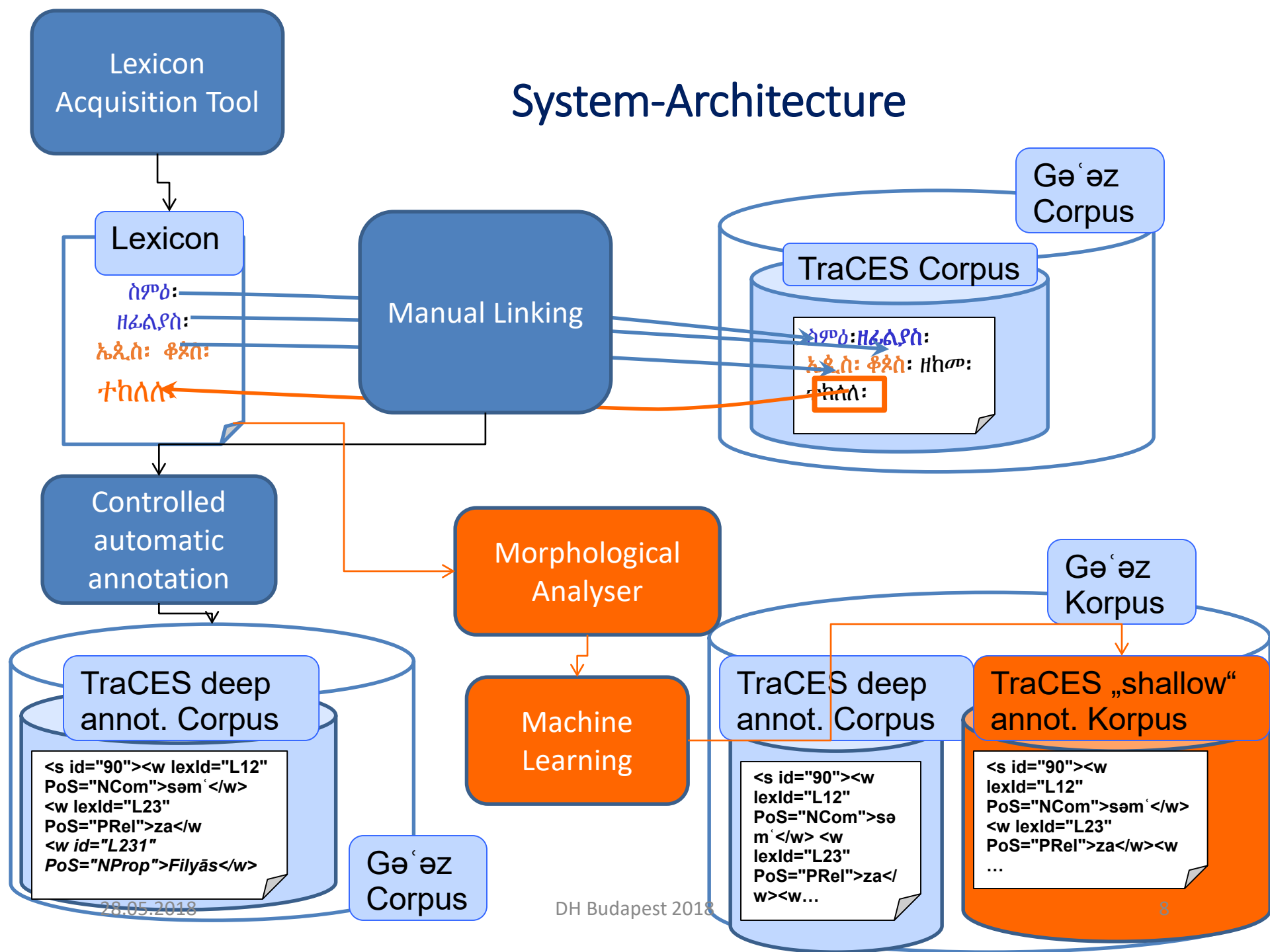
- Vowels can be themselves part-of-speech

and	House	his
ω	ᵒ	ᵒ
wa	be	tu
wa Conj	bet N	u Pr

Letter compression in Originalscript (Fidäl), but not in the transcription

before		the days		
ᵒ	ᵒ	ᵒ	ᵒ	ᵒ
'ə	ma	wā	'ə	l
'əm Prep		mawā'əl N		

# System-Architecture





# User requests and Challenges for the annotation

- Automatic transcription
- Synchronisation between original and transcription during the annotation
- CoOrrection of the text during the annotation (while kkeeping the annotation)
- Controlled automatic Annotation:
  - Tokenising
  - Change of the text
  - Linguistic Annotation
  - Sentence Annotation
  - NE-Annotation
- Possibility of very flexible text divisions (not necessary hierarchical)
- Multilevel annotation (flexible change of annotation level)
- Approx. 30 linguistic categories (PoS)(e.g. Number following 3 categories : Nature, Pattern and Syntax)
- User-friendly GUI
- Possibility of adapting the system for other scripts and transcriptions

# Textdatenbank und Wörterbuch des Klassischen Maya



Prof. Dr. Nikolai Grube<sup>1</sup>

Dr. Christian Prager<sup>1</sup>

Dr. Sven Gronemeyer<sup>1,2</sup>

Elisabeth Wagner<sup>1</sup>

Katja Diederichs<sup>1</sup>

Franziska Diehr<sup>3</sup>

Maximilian Brodhun<sup>3</sup>

<sup>1</sup> Rheinische Friedrich-Wilhelms-Universität, Bonn

<sup>2</sup> La Trobe University, Melbourne

<sup>3</sup> Niedersächsische Staats- und Universitätsbibliothek, Göttingen

Sign Catalogue

Text-Image-Link-Editor  
(Documentation Layer)

Text Markup  
(Text Layer)

Text Annotation  
(Analysis Layer)

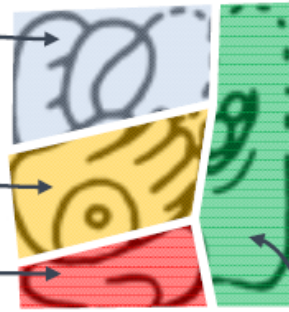
Bsp.: PAL: PMI1, A3

60st  
hasLogographicValue: **HUN**

713bb  
hasLogographicValue: **K'AL**

24st  
hasSyllabicValue: **li**

181br  
hasSyllabicValue: **ja**



[60st:713bb:24st] .181br

[60:713:24] .181

[HUN:K'AL:li] .ja






K'AL-li-HUN-ja  
K'AL-li-ja HUN

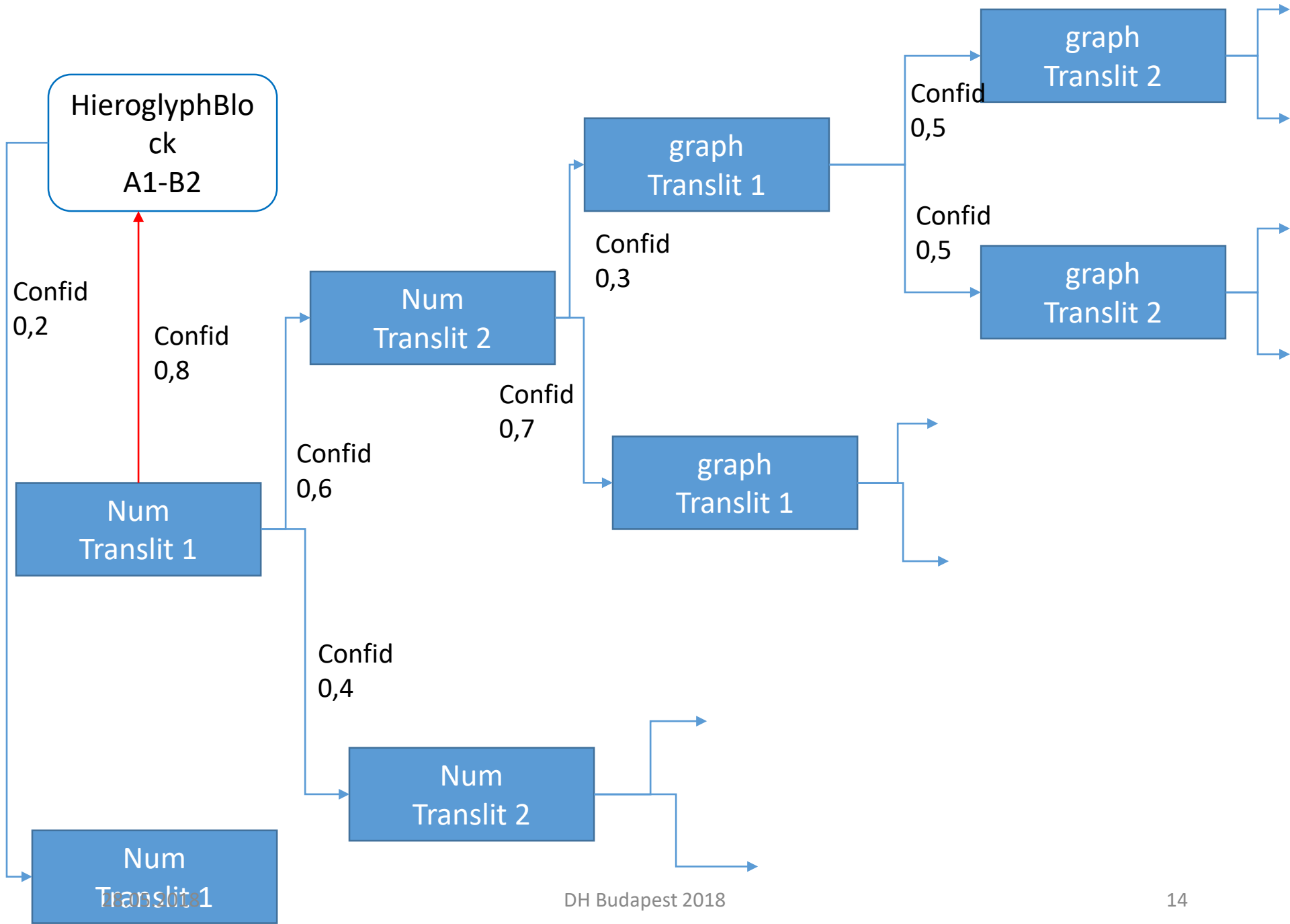


Alphanumeric  
Transliteration  
[60st:713bb:24st].181br  
Numeric  
Transliteration  
[60:713:24].181

Graphematic  
Transliteration  
Graphemic  
Transliteration  
Phonemic  
Transliteration  
Morphological  
Transcription  
Morphophonemic  
Transcription  
Morphosyntactic  
Glossing  
Final Transcription  
Translation

[HUN:K'AL:li].ja  
K'AL-li-HUN-ja  
k'al<sup>li</sup>=hu'un=ja  
*k'al-∅+hu'un-[a]j-∅*  
*k'al-∅+hu'un-aj-∅*  
V.TR:hold-NMLZ+N:paper-INCH-3s.ABS  
*k'alhu'unaj*  
he was presented the crown

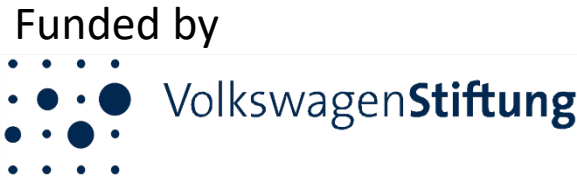
Textbeispiel						Anmerkungen
						Zeichenkatalog abfragen und muss Relation Graph → Zeichen herausuchen
Alternative 2	124:[28°738]:548:74	-	-	-	[5.201.23].683	
Alternative 3	124:[28°738]:548:74	-	-	-		
<b>Analyseebene</b>	<b>graphemische Transliteration 1</b>					semi-automatische Einsetzung der Lautwerte entsprechend der Werte im Zeichenkatalog, Auswahl entsprechend der hinterlegten Zeichenfunktionen, sofern größer 1
Alternative 1	tzi:ka°XOK:HAB:ma	9.PIK	SIH:ya.ja	SAK.ma:su	5:TZ'AM:na.ja	- mit Information zur Sicherheit (Konfidenz-Angabe)
Alternative 2	tzi:ka°XOK:HAB:ma	-	-	-	-	
Alternative 3	tzi:ka°XOK:HAB:ma	-	-	-	-	
<b>Analyseebene</b>	<b>graphemische Transliteration 2</b>					Traditionelle Transliteration in der Fachwissenschaft und korrekte Lesereihenfolge.
Alternative 1	tzi-ka XOK HAB-ma	9 PIK	SIH-ya-ja	SAK ma-su	5-TZ'AM-na-ja	Transliterationen werden in Fettschrift angegeben.  (Diese Analyseebene ist relevant für Visual Library. Export in METS/MODS ist nicht nötig. Muss per OAI PMH ausgeliefert werden können.)
Alternative 2	TZIK XOK HAB	-	-	-	5-201-na-ja	TZIK ist eine angenommene Lesung, weil wir davon ausgehen, dass 124 und 28 ein komplexes





# HerCoRe – Hermeneutic and Computer based Analysis of Reliability, Consistency and Vagueness in historical texts

- Illustrated through two main works of Dimitrie Cantemir-



April 2017 – March 2020

„Mixed Methods in Humanities“

Combine hermeneutic approaches and methods from computer science for investigating reliability and consistency of original text from 18th century as well as their translations

H

Compare for the first time “original” with translations done in the 18th- 19th century

(In)Validate assumptions about source quotations in original text

CS

Demonstrate how to include vagueness and imprecision in annotations and interpretations engines

Progress work in automatic recognition of vague expressions

## Dimitrie Cantemir (1673 -1723)



- Prince of Moldavia (historical province) as well as „universal“ humanist (linguist, ethnographer, musicologist, historian, writer)
- As member of the Royal Academy in Berlin and at the request of this institution wrote two works :
  - Description of his own country („Descriptio Moldaviae“)
  - History of ottoman empire (History of Growth and Decay of Ottoman Empire)
- Original material written in Latin; Both originals were lost already by the end of 18th century
- Several copies were used as basis for translations into German, English, French, Russian and later in Romanian
- Sometimes the translation relies on other translation (e.g. first Romanian translation of “Descriptio Moldaviae” was done after the German version from 1774.

**These translations used as reference information about the Ottoman Empire and Romanian provinces until the middle of 19th century, i.e. they give an idea about the reception about this part of the world in Western Europe.**

# Directions of investigation

- **Reliability:**

- Of the original: are the quotations made by Cantemir grounded? Is there a concordance between his degree of trust in these sources and the current knowledge about them (e.g. is there any evidence that a person which Cantemir claims to have spoken to, really lived in that time?)
- Of the translation against the original; Here an important role have the inserted editorial annotations.

- **Consistency:**

- Within the original: keeps Cantemir a constant opinion about persons, events, facts across the text? (see his own annex with annotations vs. the text)
- Across the 2 “originals”: Are common persons and events described similarly?
- Between original and translation: does the translation preserve the degree of vagueness /certainty stated by Cantemir?

- **Vagueness**

- Political or tactical reasons for imprecise expressions

# Directions of investigation

- **Reliability:**

- Of the original: are the quotations made by Cantemir grounded? Is there a concordance between his degree of trust in these sources and the current knowledge about them (e.g. is there any evidence that a person which Cantemir claims to have spoken to, really lived in that time?)
- Of the translation against the original; Here an important role have the inserted editorial annotations.

- **Consistency:**

- Within the original: keeps Cantemir a constant opinion about persons, events, facts across the text? (see his own annex with annotations vs. the text)
- Across the 2 “originals”: Are common persons and events described similarly?
- Between original and translation: does the translation preserve the degree of vagueness /certainty stated by Cantemir?

- **Vagueness**

- Political or tactical reasons for imprecise expressions

Der erste demnach, der nach Batia Einfall (\*) der Moldau ihren vorigen Glanz wieder verschafft hat, war

1. Dragosch. Obgleich unsre Jahrbücher sein Geschlecht **Legend** es

# HerCoRe – Multiple-faceted fuzzy annotation

Dragosch	<b>most_probable</b> belongs	moldavian kings
Dragosch	discovered	Moldau
Dragosch	has_acitivity	hunting
Dragosch	has_activity	development
Development	takes_place	after Batia invasion

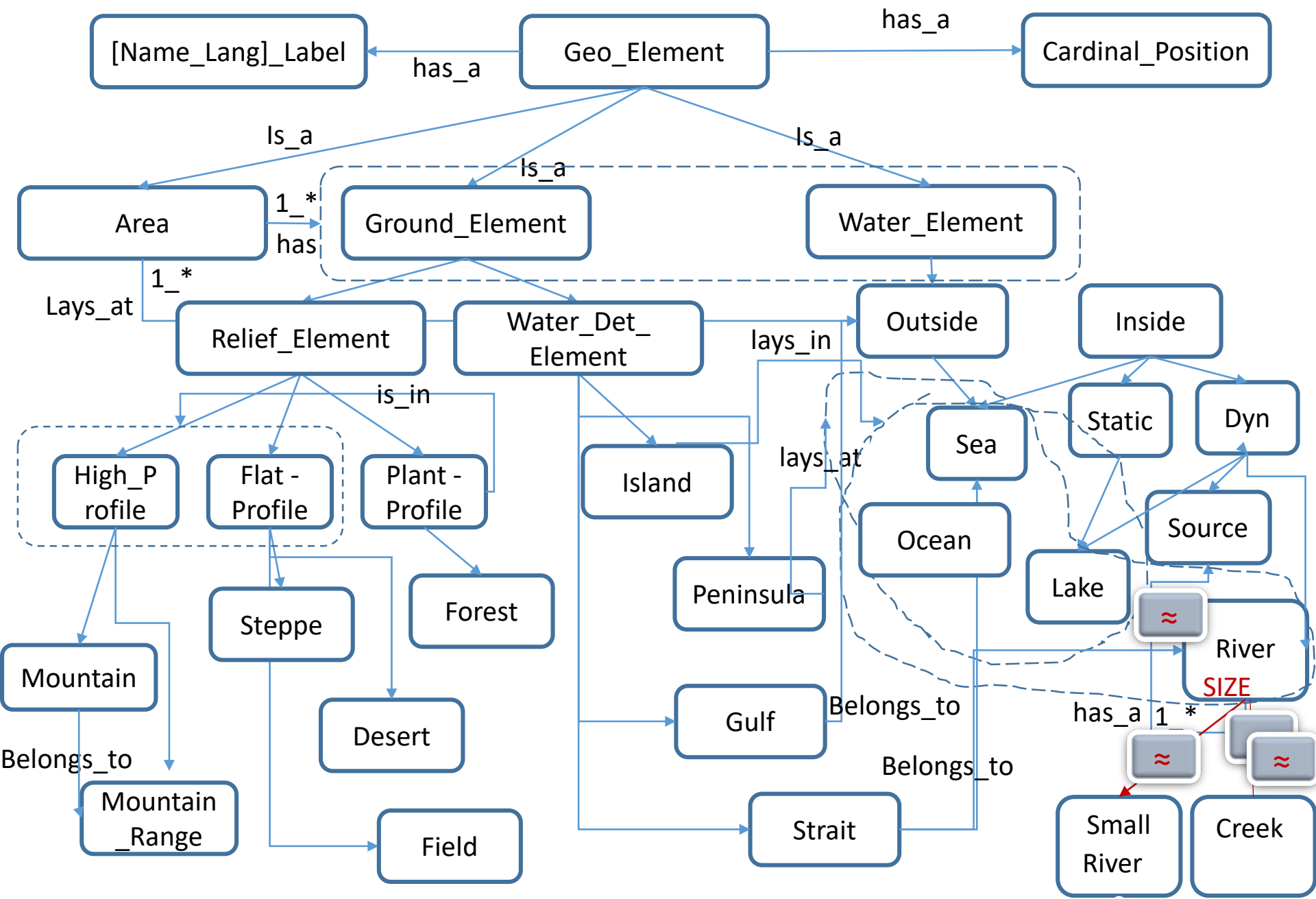
desto mehr Glauben bezumessen, weil man schwerlich glauben kann, daß einer von gemeiner Herkunft mit einem so großen Ge auf die Jagd (welche die Moldau zu entdeck Gelegenheit gegeben,) habe ausgehen, und seine ü

**Cantemir**

Dragosch	≈ son_of	Bogdan
Bogdan	≈ son_of	Johannis
Dragosch	≈ has_additional_name	Johannis
Bogdan	≈ has_additional_name	Johannis

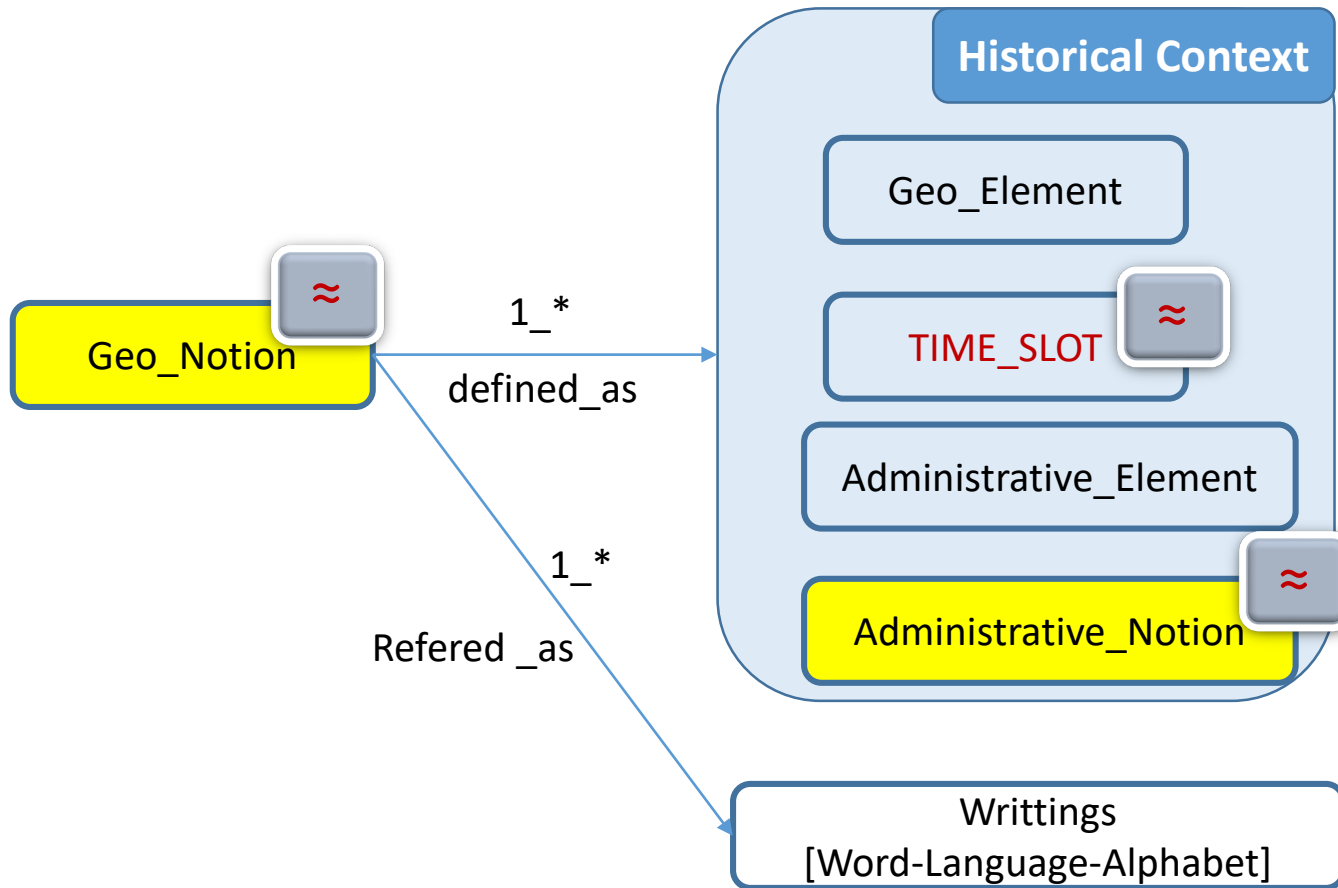
zu fo **Translator**

Dragosch	has_activity	moved
Movement	takes_place	> ≈ 1150; < ≈ 1200
Movement	takes_place	> ≈ 1200; < ≈ 1250
Bathy invasion	takes_place	≈ 1250
Bathy	has_alternative_name	Batu
Bathy	is_a	Tatar





# Ontological modelling





Orchan having in his Father's Life-time (as it is said) taken Prusa (2), and subdued the Territory of that City to his dominion, spends the first year of his Reign in settling the affairs of Afia, and establishing his new Empire

green = linguistic annotation ( N., V, Prep, ...)  
yellow = from the ontology  
orange = vagueness marker.

(2) [Having taken Prusa] The Christian Prusa to the time of Othman, who they tell us, died the following year. This mistake seems to arise from the loss of Prusa (which was a very great calamity) being known to Greece before the news of Othman's death could arrive there .

History of Growth and Decay Ottoman Empire, English Translation, pag. 24

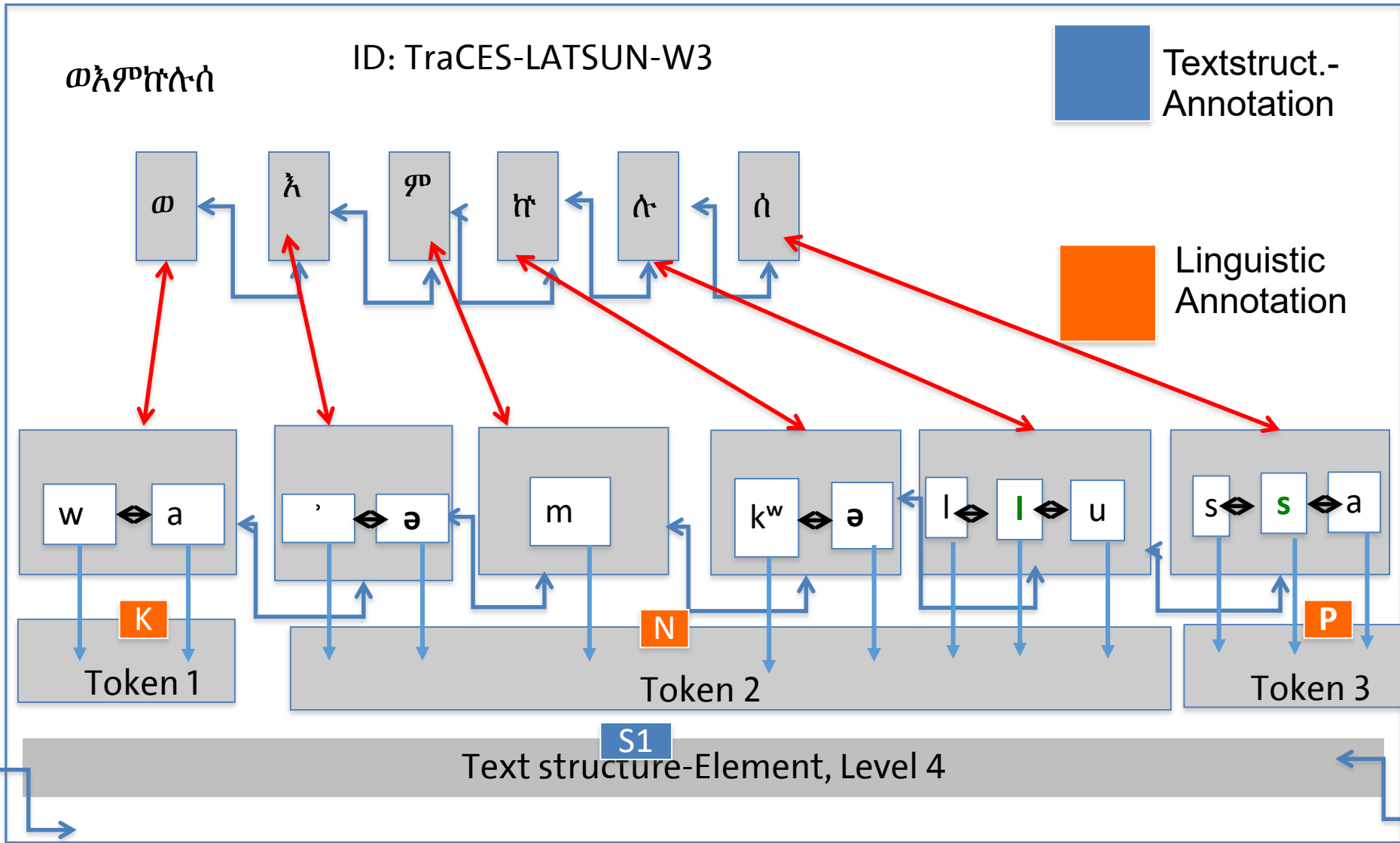
# Workflow for model design

„Words“ – are just labels for different objects

- Find the smallest unit which can be annotated -Token
- Define a GraphicalUnit (GU) - the maximal unit which can be selected with just one mouse click: (the cell of the grid.
- Define the links and eventual intermediate blocks between graphical unit and token. At the end of this step each GU will have associated a tree like structure
- Each GU receives an id; all subcomponents are identified by te GU\_Id>subcomponAbbrev\_NO
- Define spans over GUs or their subcomp. Spans do not have to contain consecutive elements
- Each span records the ids of its element. Each element records the ids of the spans.
- Annotation=Tag + a list of attributes.
- At token level and GU level changes in the text can be done (change label

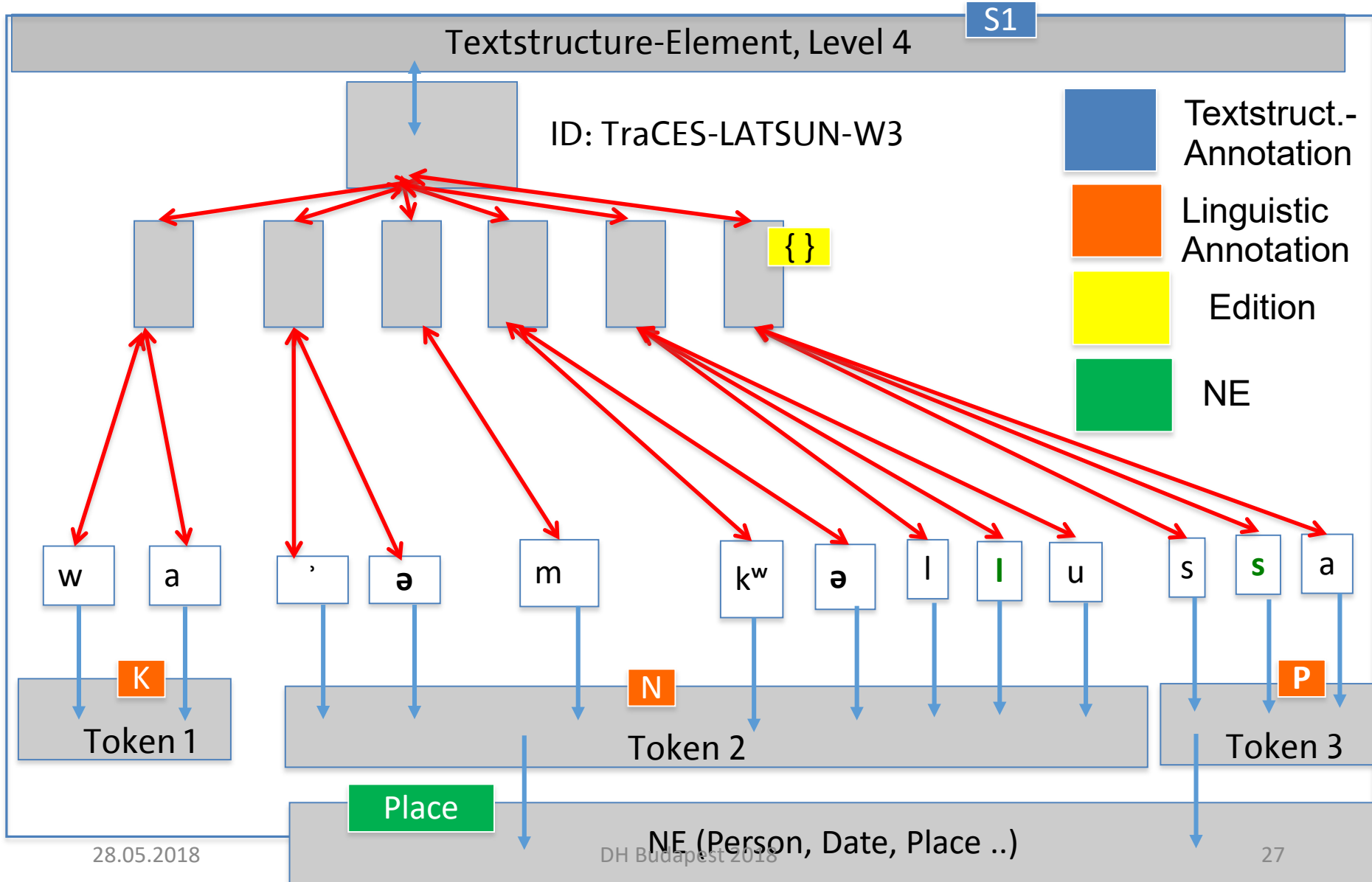
# TraCES –Classical Ethiopic

# TraCES-Annotation: Data-Model -1





# TraCES-Annotation: Data-Model -2



# GeTa Tool

- **Features**

- Easy to use GUI
- **Automatic initial transcription** (vocalized or unvocalized)
- Synchronisation between original and transcription
- Controlled changes on text while annotating
- possible
- **Controlled semi-automatic:**
  - tokenization,
  - change of the transcripts text,
  - deep linguistic annotation + link to lexicon
  - Name Entity annotation linked with the authority DB.
- **Automatic „sentence“ recognition**
- Visualisation of data model
- Visualisation of annotation progress
- Can read additionally Classical Ethiopic inscriptions written with South Arabic script

Software development

- Client-Application
- Open source; Java
- Data-encoding:JSON

# GeTa-Levels

GeTa - TraCES Annotation Tool Ver4

File History Show Window Tools View Annotations Keyboards Help Automatic Annotation

Show cursor position

Level part: part 1(Internal Nr.: 1)

Level chapter: Buch 1.1.0(Internal Nr.: 1.0)

Level sentence: S 1.1.0.2(Internal Nr.: 1.0.2)

Original AmdaSeyon.json

በስመ :	አብ :	ወወልድ :
ወመንፈስ :	ቅዱስ :	ጆአምላክ :
ንጽሕፍ :	በረድኤተ :	እግዚእነ :
ኢየሱስ :	ክርስቶስ :	ኅይለ :
ወመዊእ :	ዘገብረ :	እግዚአብሔር :
በ[እ]ደዊሁ :	ለዓምደ :	ጽዮን :
ንጉሠ :	ኢትዮጵያ :	ወስመ :
መንግሥቱ :	ገብረ :	መስቀል :
በ፲ወ፰ዓመተ :	መንግሥቱ :	እምዘ :
ነግሠ :	ወዓመተ :	ምሕረትሂ :
ጅጅ፲ወ፯	።	ንጽሕፍቤ :
ዘንተ :	መጽሐፈ :	እንዘ :
ንትአመን :	አበ :	ረዳኤ :
ወወልደ :	ናዛዜ :	ወመንፈስ :
ቅዱስ :	መራሄ :	እንዘ :
ኅኅሥሥ :	ረድኤተ :	እምቅድስት :
ሥላሴ	።	በስመ :
ይቤ :	ያዕቆብ :	ሐዋርያ :
እመቦ :	ዘኅጥአ :	ለጥቡብ :
ለይስአል :	ኅብ :	ወሃቢ :
እግዚአብሔር :	ዘይሁብ :	ለኩሉ :
በስፉሕ :	ወኢይትዒየር :	ወንሕሂ :
ንግሥሥ :	ረድኤተ :	እምኣብ :

Show current word annotation

la  
Preposition state:Nominal state lex:ለ  
kʷəll  
Pronoun of Totality Base case:Nominative lex:ኩሉ  
u  
Pronominal Suffix person:Third gender:Masculine number:Singular lex:ሁ

Transliterated AmdaSeyon.json

ba-səma	ʾab	wa-wald
wa-manfas	qəddus	l-ʾamlāk
nəṣəḥəf	ba-radʾeta	ʾəgziʾə-na
ʾiyasus	krəstos	ḥayla
wa-mawiʾa	za-gabra	ʾəgziʾabəḥer
ba-ʾədawi-hu	la-ʿāmda	ṣəyon
nəguša	ʾityopyā	wa-səma
mangəšt-u	gabra	masqal
ba-10-wa-8-ʿāmata	mangəšt-u	ʾəm-za
nagša	wa-ʿāmata	məḥratə-hi
5-100-10-wa-7	.	nəṣəḥəf-ke
zanta	maṣḥafa	ʾənza
nəʾamman	ʾaba	radāʿe
wa-walda	nāzāze	wa-manfas
qəddusa	marāhe	ʾənza
naḥāsšəs	radʾeta	ʾəm-qəddəst
šəllāse	.	ba-kama
yəbe	yāʿqob	ḥawārəyā
ʾəmma-b-o	za-ḥaʾʾa	la-ṭəbab
la-yəsʾal	ḥaba	wahābi
ʾəgziʾabəḥer	za-yəhub	la-kʷəll-u
ba-səfuḥ	wa-ʾi-yəʿeyyar	wa-nəḥna-ni
nəḥšəs	radʾeta	ʾəm-ʾab

# GeTa-Levels

GeTa - TraCES Annotation Tool Ver4

File History Show Window Tools View Annotations Keyboards Help Automatic Annotation

Show cursor position  
Level section: section 60(Internal Nr.: 60)

Show current word annotation  
wa  
Conjunction **lex:⌘**  
mawito  
Verb **tam:Gerund** **person:Third** **gender:Masc** **number:Singular** **lex:⌘**

Original Matthew\_20170615\_1931.json

Transliterated Matthew\_20170615\_1931.json

በኅዳር :	ንጉሥ :	ሐፍ	<b>həḍānāta</b>	<b>za-bcta</b>	ləḥem
ወናሁ :	ውእቲ :	ኮከ	wa-za-kʷəll-u	<b>dawal-ā</b>	za-kəlʰe
ዘአስተርአዮሙ :	በሠርቅ :	መ	<b>kəramt-u</b>	wa-za-yənoʾəs	ʾəmənne-hu
እስከ :	አብጽሐሙ :	ወፋ	<b>ba-ḥassāba</b>	taṭayyaq-omu	la-sabʾa
መልዕልተ :	ሕፃን	መ	sagal	.	kama
ወሶበ :	ርእይዎ :	ለ	yəbšāḥ	<b>qāla</b>	<b>nabiyy</b>
ኮከብ :	ተፈሥሐ :	ወፋ	za-yəbe	ʾərməyās	qāl
ትፍሥሕተ :	ጥቀ :	#	tasamʿa	ba-rāmā	<b>ḥəməām</b>
ቤተ :	ረከብዎ :	ለ	<b>wa-bəkāy</b>	<b>wa-gaʿār</b>	<b>bəzuḥ</b>
ወለማርያም :	እሙ :	ወፋ	rāḥel	təbakəyomu	la-daqiq-ā
ሎቲ :	ወአምሳዎ :	ወፋ	wa-ʾabayat	<b>nagifa</b>	<b>lāḥa</b>
ወፈትሐ :	መዛግብተሆሙ :	ለ	ʾəsmā	ʾi-konəw-ā	.
ሎቲ :	አምኝ :	ለ	<b>wa-mawito</b>	herodəs	nā-hu
ወሱኒኑ :	ወከርቤ :	ወፋ	<b>malʾaka</b>	ʾəgziʾa	bəḥer
በሕልም :	ከመ :	#	ʾastarʾay-o	<b>ba-ḥəlm</b>	<b>la-yosef</b>
ኅዳር :	ሄሮድስ :	ወፋ	<b>ba-bəḥera</b>	gəbš	wa-yəbel-o
ካልእ :	ፍኖት :	ወፋ	tanśəʾ	wa-nəśāʾ	<b>ḥəḍāna</b>
		ወፋ	wa-ʾəm-o	wa-ḥor	<b>bəḥera</b>
		ወፋ	ʾəsrāʾel	ʾəsmā	motu
		ኢ	ʾəlla	yaḥāsśəsəw-o	la-nafsa
		ወፋ	<b>ḥəḍān</b>	.	wa-tanśəʾa
		ወፋ	wa-naśʾa	<b>ḥəḍāna</b>	wa-ʾəm-o
		አ	wa-ḥora	<b>bəḥera</b>	ʾəsrāʾel

# Annotation **Common Noun** in GeTa

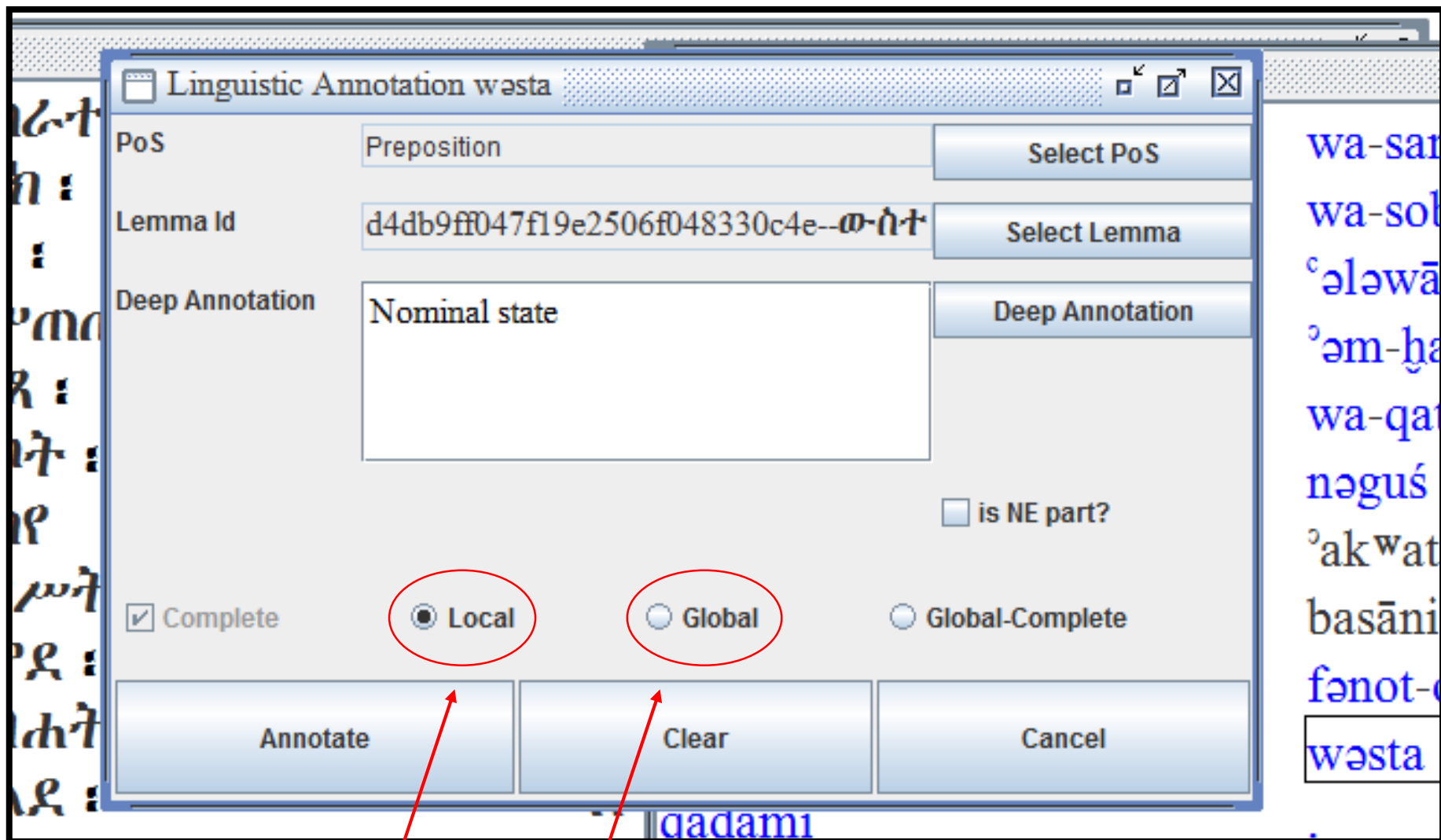
The screenshot displays the GeTa linguistic annotation interface. The main window, titled "Linguistic Annotation - masqal", shows the following information:

- PoS:** Common Noun (circled in red)
- Lemma Id:** fc8947488f414e51e89fe0f2--መስቀል
- Deep Annotation:** unmarked SingularP UnmarkedS Nominative Absolute state

Below the main window, a "Deep Annotation" dialog box is open, showing detailed settings for the annotation:

- Gender:**  masculine,  feminine,  unmarked.  Nature,  Pattern,  Syntax. A green arrow points to the  Syntax option.
- Number:**  Singular,  Plural extern,  Plural intern,  Plural of the plural,  Unmarked.  Singular,  Plural,  Unmarked. A green arrow points to the  Syntax option.
- Case:**  Nominative,  Accusative,  Vocative.
- State:**  Absolute State,  Construct State,  Pronominal State.

The background text in the interface includes "ba-ʾədawi-hu", "ʾityopyā", "masqal", and "nagśa".





# GeTa NE-Annotation

The screenshot displays the GeTa - TraCES Annotation Tool Ver4 interface. The main window shows a list of Amharic words on the left, with the word "ክሪስቶስ" (Kristolos) highlighted. The right side shows a list of transliterated words, with "krastos" highlighted. A central window titled "Insert NE iyasus krastos" is open, showing options for "Person", "Place", "Organisation", "Title", "Work", and "Date". The "Person" option is selected. Below this, there are search and authority list options. A bottom window titled "Insert Global NE iyasus krastos" is also open, showing a list of text segments with checkboxes for "Before" and "After" insertion. The text segments include phrases like "wa-manzäh 'ällu k'wäll-omu habiromu rodu-ni wa-'agatu-ni wa-'ana-hi mo'kəw-omu wa-qatalkəw-omu ba-hayla 'iyasus krastos person 0" and "wəsta dabtarā marṭul ḥaba tābot wa-'ahaza 'aqrənta məšəwā' wa-'astamḥara ḥaba 'iyasus krastos person 1".

GeTa - TraCES Annotation Tool Ver4

File History Show Window Tools View Annotations Keyboards Help Automatic Annotation

Show cursor position  
Level part: part 1(Internal Nr.: 1)  
Level sentence: S 1.1(Internal Nr.: 1.1)

Show current word annotation  
'iyasus  
Proper Name lex:ክሪስቶስ gender:male number:SingularP UnmarkedS case:Nominative is NE part

Transliterated AmdaSeyon.json

ba-səma	'ab	wa-wald
wa-manfas	qəddus	l-'amlāk
nəṣəḥəf	ba-rad'eta	'əgzi'ə-na
'iyasus	krastos	ḥayla
wa-mawi'a	za-gabra	'əgzi'abəḥer
ba-'ədawi-hu	la-'āmda	ṣəyon
nəguša	'ityopyā	wa-səma
mangəšt-u	gabra	masqal
	mangəšt-u	'əm-za
	wa-'āmata	məḥratə-hi
		nəṣəḥəf-ke
	maṣḥafa	'ənza
	'aba	radā'e
	nāzāze	wa-manfas
	marāhe	'ənza
	rad'eta	'əm-qəddəst
		ba-kama
	yā'qob	ḥawārəyā
	za-ḥaṭ'a	la-ṭəbab
	ḥaba	wahābi
	za-yəhub	la-k'wəll-u
	wa-'i-yəf'eyyar	wa-nəḥna-ni
	rad'eta	'əm-'ab
	wa-manfas	qəddus
	vərdə'anna	la-'ālama

Original AmdaSe  
ክሪስቶስ :  
ወመንፊስ :  
ንጽሕፍ :  
ኢየሱስ :  
ወመዊእ :  
በ[እ]ደዊሁ :  
ንጉሠ :  
መንግሥቱ :  
በ፲ወጃዎመተ :  
ነግሠ :  
ጴጥሮስ :  
ዘንተ :  
ንተአመን :  
ወወልደ :  
ቅዱስ :  
ነገሥሥ :  
ሥላሴ :  
ይቤ :  
እመቤ :  
ለይስአል :  
እግዚአብሔር :  
በስፋሕ :  
ንገሥሥ :  
ወወልድ :  
ከመ : 28.05.2018

ክሪስቶስ :  
ሥላሴ :  
ይቤ :  
እመቤ :  
ለይስአል :  
እግዚአብሔር :  
በስፋሕ :  
ንገሥሥ :  
ወወልድ :  
ከመ : 28.05.2018

ክሪስቶስ :  
ሥላሴ :  
ይቤ :  
እመቤ :  
ለይስአል :  
እግዚአብሔር :  
በስፋሕ :  
ንገሥሥ :  
ወወልድ :  
ከመ : 28.05.2018



# GeTa-operations

GeTa - TraCES Annotation Tool Ver4

File History Show Window Tools View Annotations Keyboards Help Automatic Annotation

Show cursor position  
Level part: part 1(Internal Nr.: 1)  
Level chapter: Buch 1.1.0(Internal Nr.: 1.0)  
Level sentence: S 1.1.0.2(Internal Nr.: 1.0.2)

Show current word annotation  
la  
Preposition state:Nominal state lex:ለ  
kʷəll  
Pronoun of Totality Base case:Nominative lex:ኩሉ  
u  
Pronominal Suffix person:Third gender:Masculine number:Singular lex:ሁ

Original AmdaSeyon.json

Transliterated AmdaSeyon.json

በሰመ :	አብ :	ወወልድ :	ba-səma	ግላ	wa-wald
ወመንፈስ :	ቅዱስ :	ጄአምላክ :	wa-manfas	ባዕደድ	1-ʼamlāk
ንጽሕፍ :	በረድኤተ :	እግዚእነ :	nəṣəḥəf	ባራደግጅታ	ʼəgzɪʼə-na
ኢየሱስ :	ክርስቶስ :	ኅይለ :	ʼiyasus	ክሬስቶስ	ḥayla
ወመዊአ :	ዘገብረ :	እግዚአብሔር :	wa-mawiʼa	ገራገራ	ʼəgzɪʼabəḥer
በ[እ]ደዊሁ :	ለዓምደ :	ጽዮን :	ba-ʼədawi-hu	ለ-አሙድ	ṣəyon
ንጉሠ :	ኢትዮጵያ :	ወሰመ :	nəguša	ግዮግዮ	wa-səma
መንግሥቱ :	ገብረ :	መስቀል :	mangəšt-u	ገራ	masqal
በገደግመተ :	መንግሥቱ :	እምዞ :	ba-10-wa-8-ʼāmata	ገራ	mangəšt-u
ኅግሠ :	ወዓመተ :	ምሕረትሂ :	nagša	ገራ	wa-ʼāmata
ጄገገወገ	#	ንጽሕፍኬ :	5-100-10-wa-7	.	nəṣəḥəf-ke
ዘንተ :	መጽሐፈ :	እንዘ :	zanta	ማሻፋ	ʼənza
ንትአመን :	ኔባ :	ረይኔ :	nəʼamman	ግላ	radāʼe
ወወልድ :			wa-walda	ገራ	nāzāze
ቅዱስ :			qəddusa		wa-manfas
ኅኅሥሥ :			naḥəsəsəs		ʼənza
ሥላሴ :			šəllāse		ʼəm-qəddəst
ይቤ :			yəbe		ba-kama
እመቤ :			ʼəmma-b-o		ḥawārəyā
ለይስአል :			la-yəsʼal		la-ṭəbab
እግዚአብሔር :	ዘይሁብ :	ለኩሉ :	ʼəgzɪʼabəḥer	ገራ	wahābi
በስፋሕ :	ወኢይትግየር :	ወንሕሂ :	ba-səfuḥ	ገራ	la-kʷəll-u
ንኅሥሥ :	ረድኤተ :	እምአብ :	nəḥsəs	ገራ	wa-nəḥna-ni
ወወልድ :	ወመንፈስ :	ቅዱስ :	wa-wald	ገራ	ʼəm-ʼab
ከመ :	ይርድኔ :	ለዓለመ :	kama	ገራ	qəddus

Word Structure

Modify

Token

kʷ ə l l

Insert ə Delete ə

Geminate Remove Gemination

Global Local Modify Cancel

Maya

### BlockElement

Id: String  
Label: String

### HieroglyphElement

Id: <Nr>-H-<Id\_HyeroglyphBlock>  
Label: String

Id's of Components  
in which is included



#### Reading 1

Label :String  
Konfidenz :Nummer

#### Reading 2

Label : String  
Konfidenz : Nummer

### OperatorElement

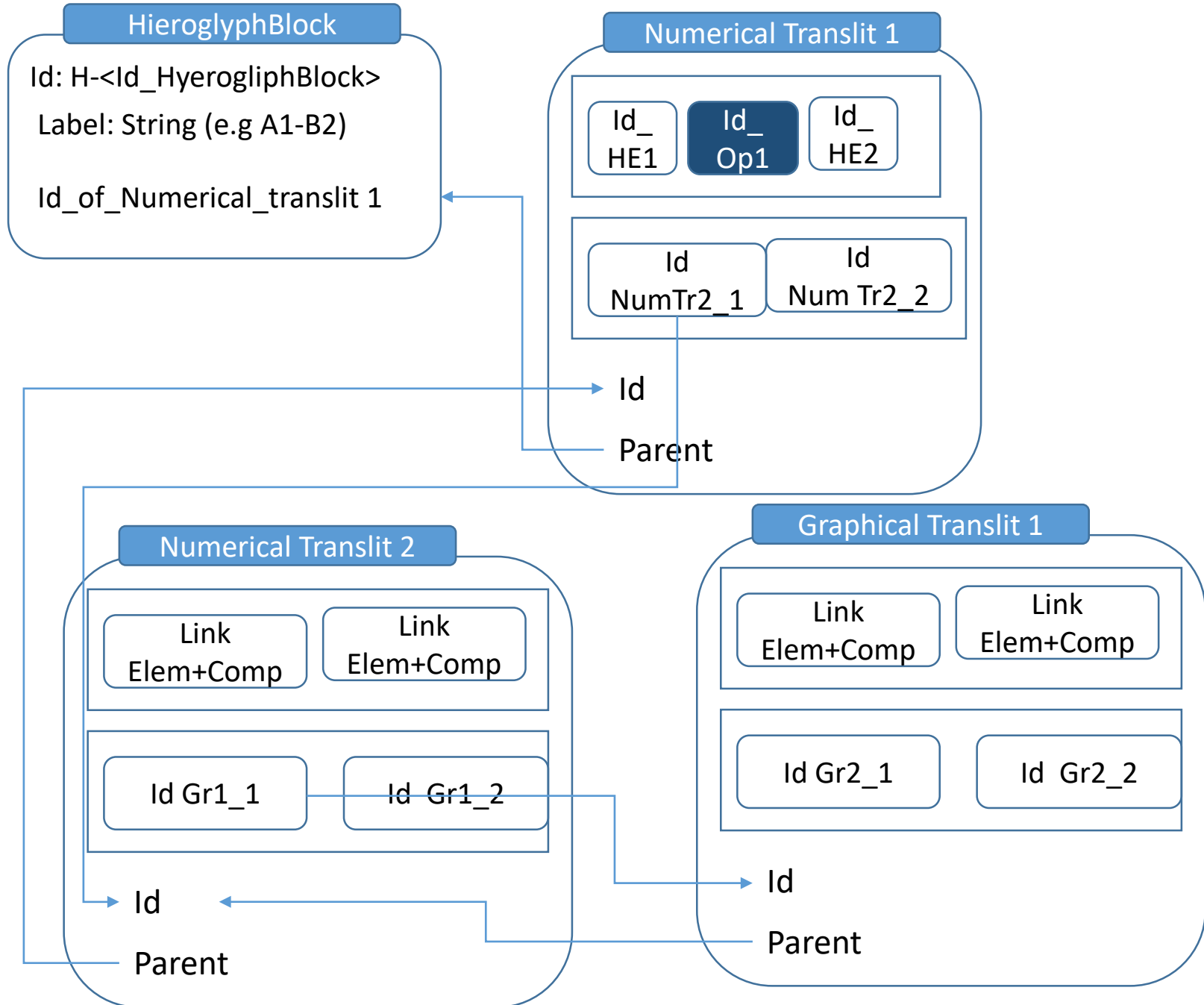
Id: O-<Id\_HyeroglyphBlock>  
Label: String  
Level: String

### ComponentElement

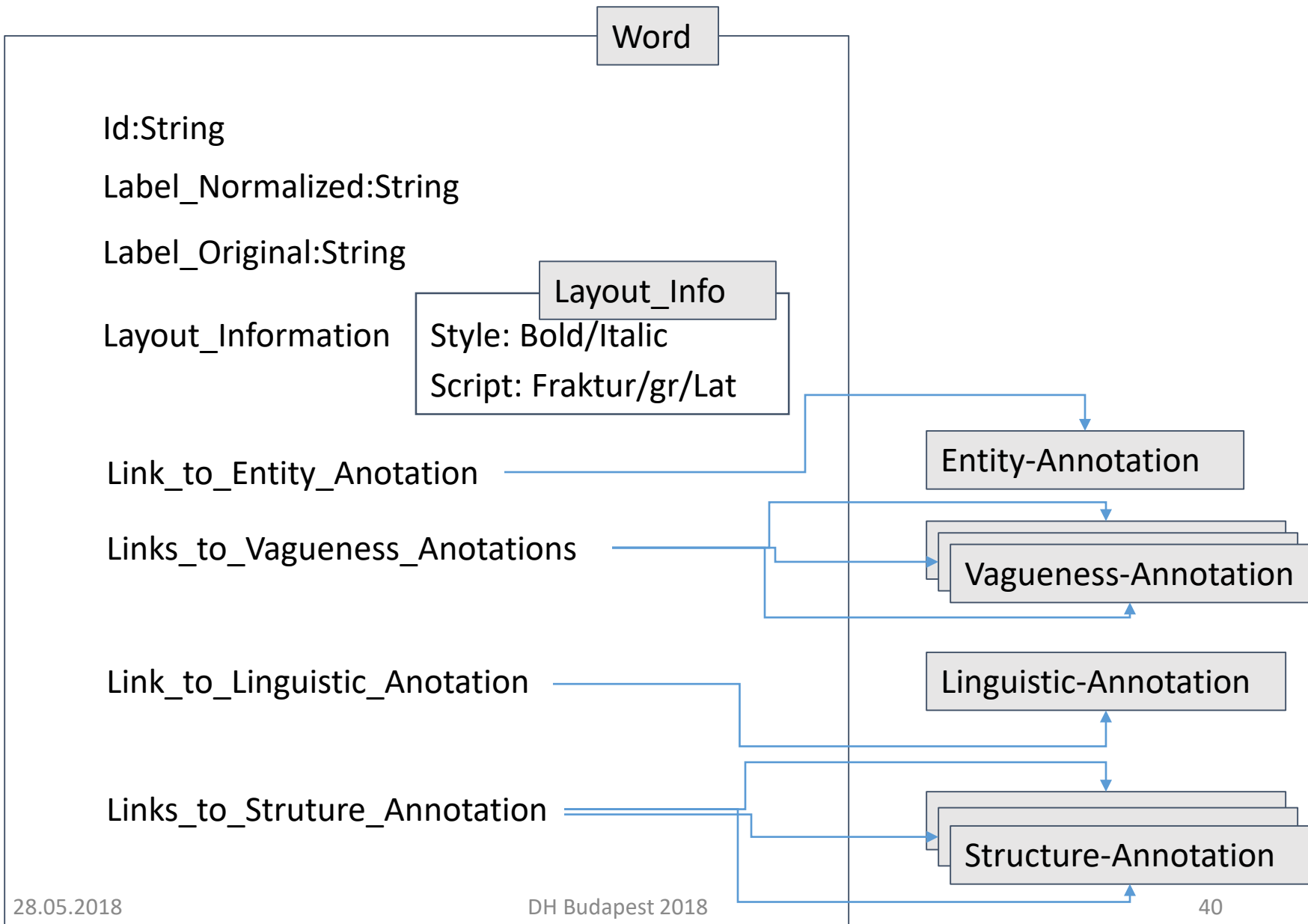
Id: <Nr>C-<Id\_HyeroglyphBlock>  
Label: String  
Color: String

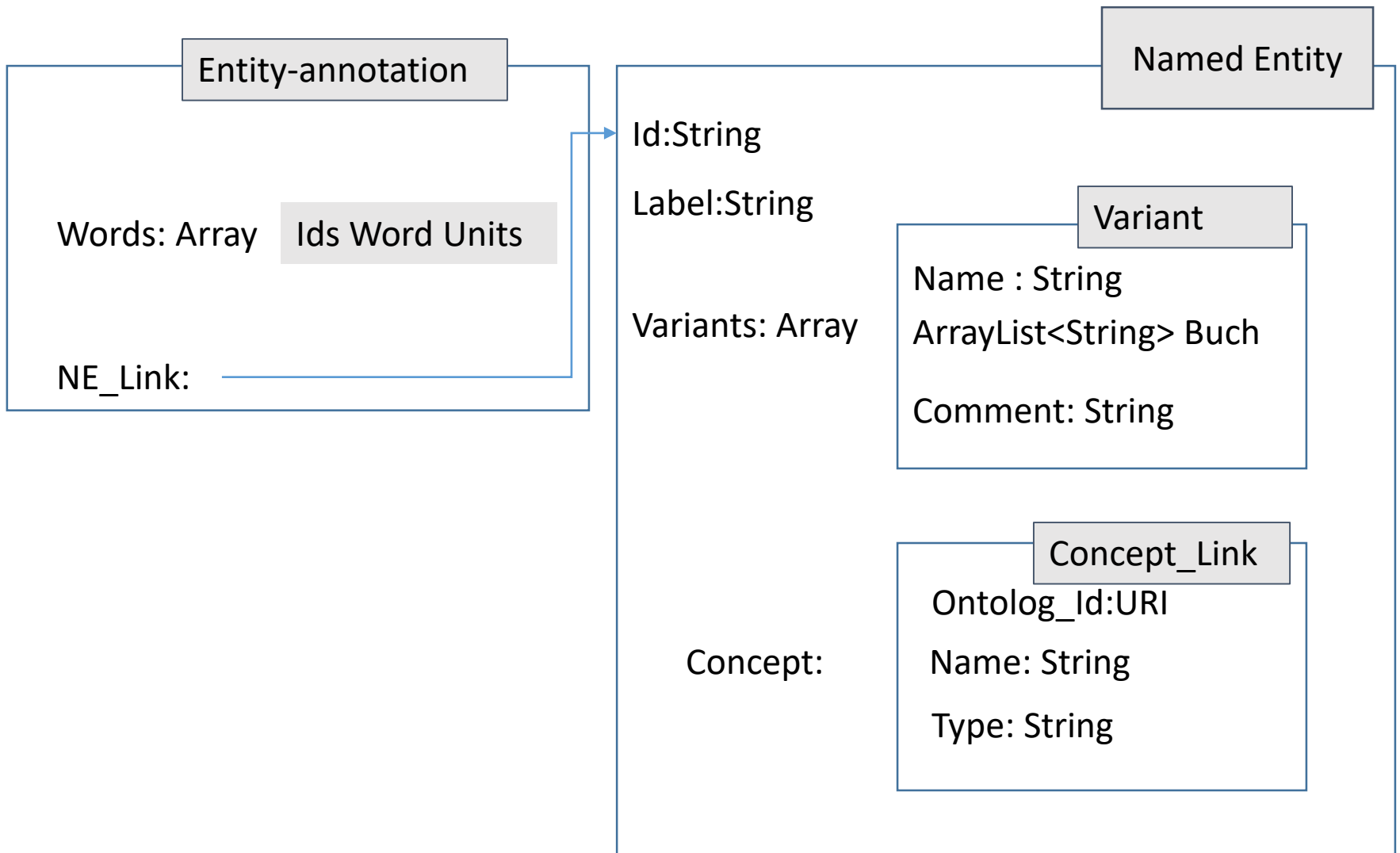
Id's of contained  
Hieroglyph elements





# HerCoRe- Multilingual Texts -18th Century





## Vagueness-annotation

Id:String

Type: String (Quotation /Linguistic / Edition /Geo/Genre)

Subtype: String (dependent on each type)

Words: Array Word\_Id

Confidence: String (low/medium/high)

## Linguistic-annotation

Id:String

Lemma: String

PoS: String

Morpho\_Features: String

Words: Array Word\_Id

## Structure-annotation

Id:String

Level: String (Chap/Paragraph/Sentence)

Type: String (Author/Editor)

Words: Array Word\_Id



S1  
Level: Sentence  
Type: Author  
Words: W1,W2,W3,W4,W5,W6,W7,W8,W9,W10,W11

W1 Die Die n <b>frak</b>	W2 Moldau Moldau b <b>frak</b>	W3 hat hat n <b>frak</b>	W4 nicht nicht n <b>frak</b>	W5 zu zu n <b>frak</b>	W6 allen allen n <b>frak</b>	W7 Zeiten Zeiten n <b>frak</b>	W8 eben eben n <b>frak</b>	W9 dieselben dieselben n <b>frak</b>	W10 Grenzen Gränzen n <b>frak</b>	W11 gehabt gehabt n <b>frak</b>
--------------------------------------	--	--------------------------------------	--	------------------------------------	--	--	--	--	---	---

Ling Ann :L2  
Lemma:Moldau  
PoS :N  
Morpho:  
Words: W2

Vagueness Ann :V1  
Type: ling  
Subtype : time expression  
Confid: low  
Words: W5,W6,W7

Ent Ann :E1  
Words: W2  
NE20

NE20  
Label: Moldova  
Var: Moldau (DDM)  
URI: https://..  
Name: placeName  
Type: country

Ling Ann :L3  
Lemma:haben  
PoS :Verb  
Morpho: 3.sg. Perf.  
Words: W3,W11

**Thank you for attention !**

**Köszönöm a figyelmet !**