

COMHUM 2018 – Workshop on Computational Methods in the Humanities
June 4 - 5 2018, Lausanne (Switzerland)



Modelling vagueness - A criteria-based system for the qualitative assessment of reading proposals for the deciphering of Classic Mayan hieroglyphs

Franziska Diehr³, Maximilian Brodhun³, Sven Gronemeyer^{1,2},
Christian Prager¹, Elisabeth Wagner¹, Katja Diederichs¹,
Nikolai Grube¹

¹ Rheinische Friedrich-Wilhelms-Universität, Bonn

² La Trobe University, Melbourne

³ Niedersächsische Staats- und Universitätsbibliothek, Göttingen

Text Database and Dictionary of Classic Mayan



a-ALAY-ya T1014-yi u tz'i-bi-na-ja-la yu-k'i-bi ti yu-ta-la ka-ka-wa

alay ?-V_y utz'ihbnajal yuk'ib ti yutal kakaw

Goals

- making all **known inscriptions** and **their text carriers** accessible in a machine-readable **corpus**
- compilation of a **dictionary** that maps the entire **vocabulary** and its use in the **script**

Funding

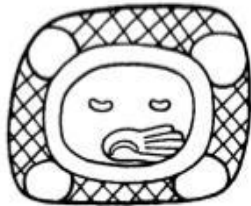
- long-term project: 2014 ~2028
- North Rhine-Westphalian Academy of Sciences, Humanities and the Arts; Union of the German Academies of Sciences and Humanities

Partners

- Department of Anthropology of the Americas at University Bonn
- Göttingen State and University Library

The Script of Classic Mayan

- script with iconic character
- logo-syllabic writing system
= logogram **PAKAL** + syllabogram /pa/ /ka/ /la/
- arrangement of signs in a hieroglyphic block $\hat{=}$ word



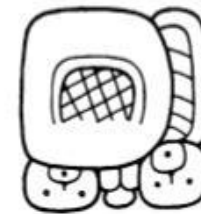
PAKAL

Logogram



PAKAL-la

Logogram and Syllabogram



pa-ka-la

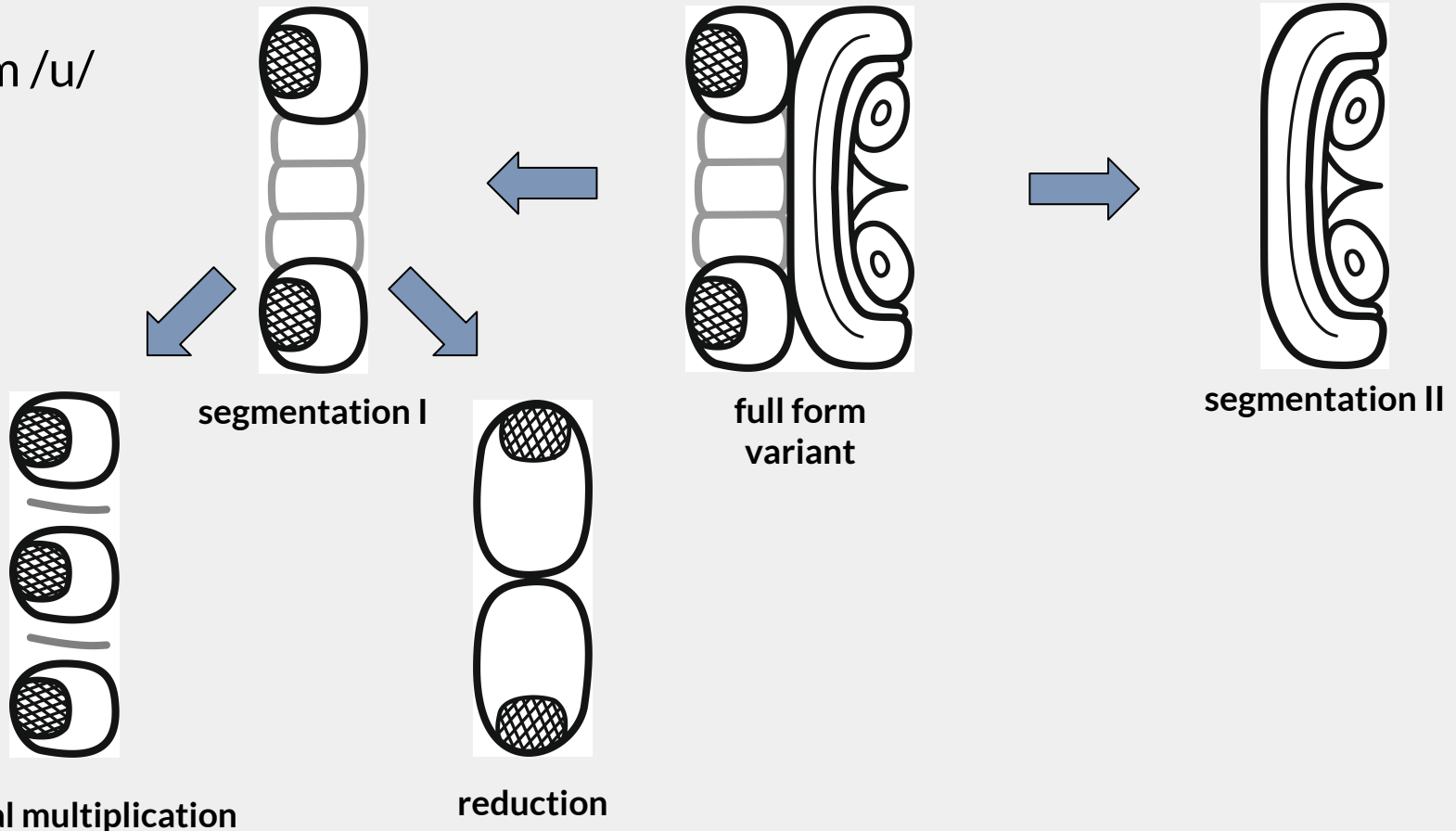
Syllabograms

Graph Variants

distinct calligraphic complexity

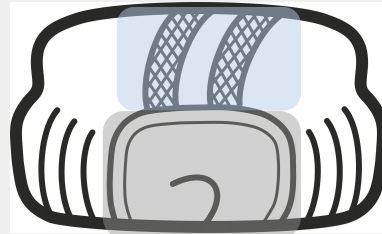
↳ one sign - many graphical expressions

syllabogram /u/

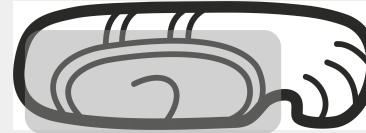


drawings: Christian Prager, CC-BY-4.0

syllabogram /yi/



full form
variant



segmentation and
compression

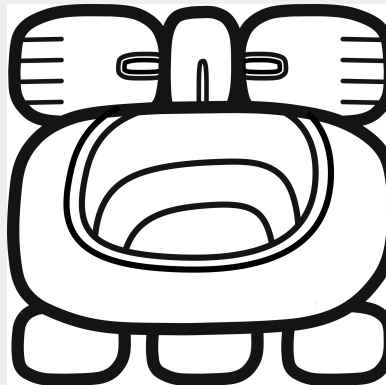


transformation

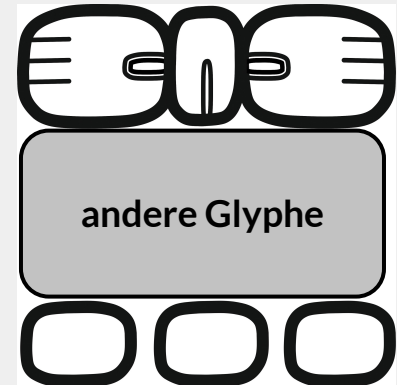
syllabogram /ma/



segmentation I



full form
variant



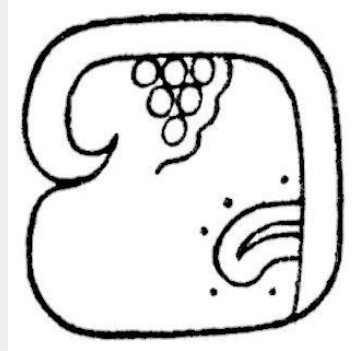
segmentation III



segmentation II



multiple Sign Functions and Readings (Polyvalence of Signs)



reading:

TUN

CHAHUK

ku

meaning:

stone

19th day name

ku

funktion:

logogram

logogram

syllable

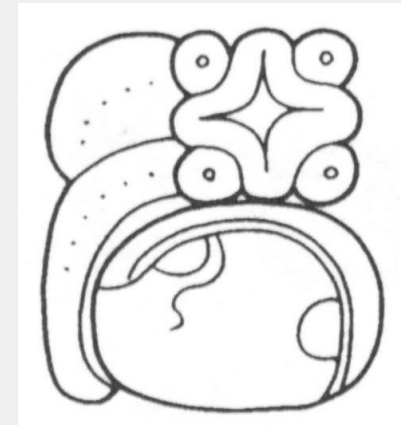
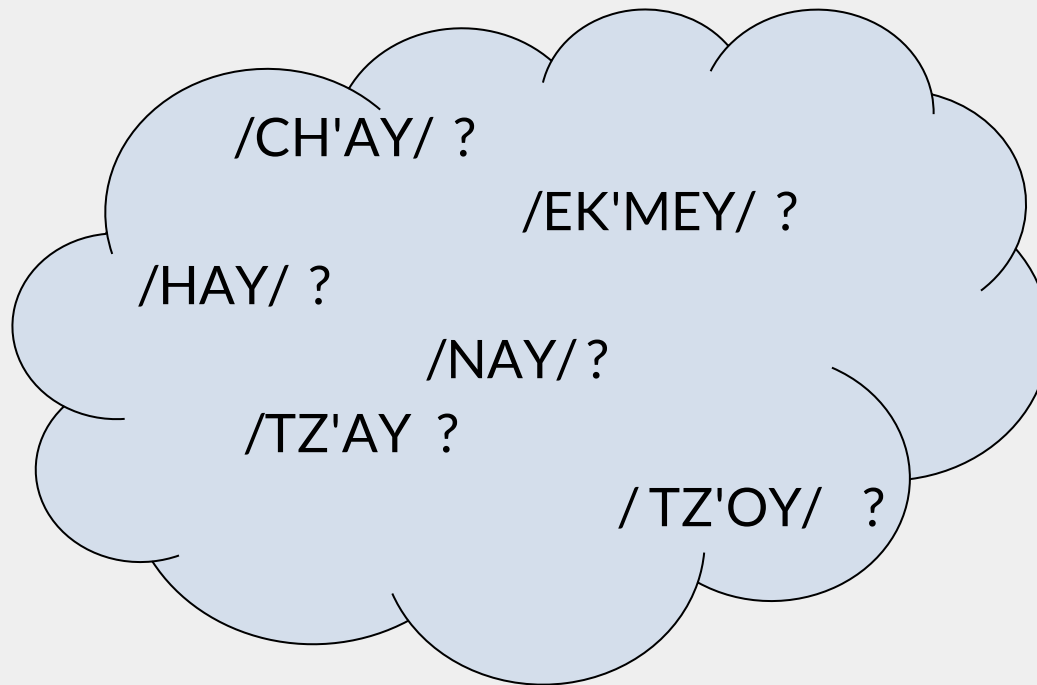
drawing: © J. Eric S. Thompson, A Catalog of Maya Hieroglyphs, 1962

Reading Hypotheses

Reading hypotheses for (undeciphered) signs

one sign → more than one reading proposal

“Star War” glyph



drawing: © Macri, Martha J., Gabrielle Vail (2009): The New Catalog of Maya Hieroglyphs: The Codical Texts. Civilization of the American Indian Series 264. University of Oklahoma Press, Norman, OK.

Why do we need a digital Sign Catalogue?

sign catalogue = inventory of all signs
→ aid for deciphering

encoding the corpus in XML using TEI P5 Guidelines

text ≠ phonemic-transliterated values

text ≠ Unicode

text = identifiable graphical representations

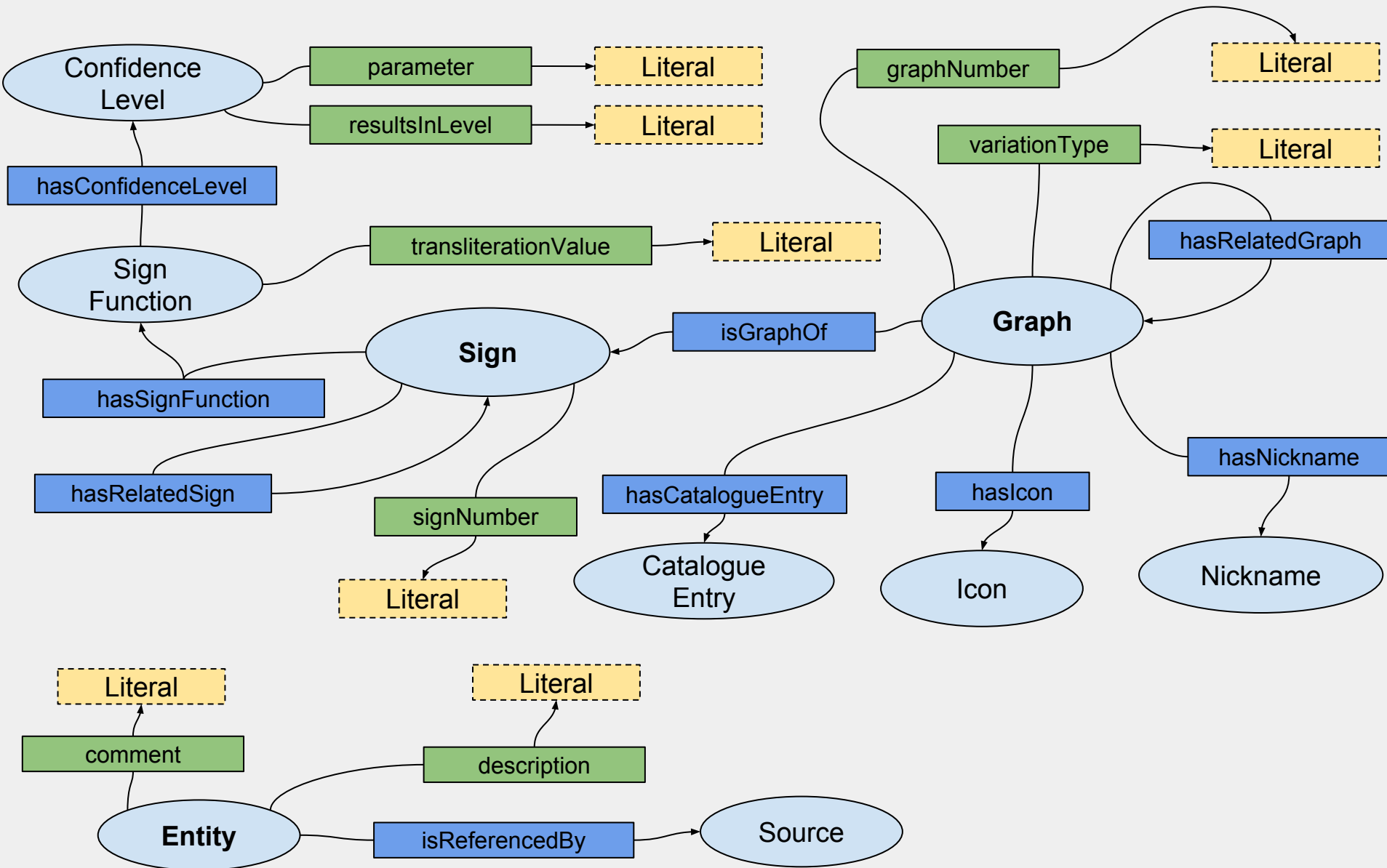
→ How do we encode the text?

```
<text>
  <body>
    <p>???
```



?

Ontological Modelling Approach



Implementation and Storing

```

2113 idiomcat:colonialSources a rdf:Property;
2114   rdfs:label "Colonial Sources"@en;
2115   rdfs:domain idiomcat:CLLR;
2116   rdfs:range xsd:boolean;
2117   tgforms:isRepeatable false;
2118   tgforms:hasInput tgforms:checkbox;
2119   tgforms:hasPriority "215";
2120   rdfs:comment "The reading is attested in Colonial sources, chi
2121
2122 idiomcat:completePhoneticSubstitution a rdf:Property;
2123   rdfs:label "Complete Phonetic Substitution"@en;
2124   rdfs:domain idiomcat:CLLR;
2125   rdfs:range xsd:boolean;
2126   tgforms:hasInput tgforms:checkbox;
2127   tgforms:isRepeatable false;
2128   tgforms:hasPriority "214";
2129   rdfs:comment "The reading
2130
2131 idiomcat:partialPhoneticSubstitution a rdf:Property;
2132   rdfs:label "Partial Phonetic Substitution"@en;
2133   rdfs:domain idiomcat:CLLR;
2134   rdfs:range xsd:boolean;
2135   tgforms:isRepeatable false;
2136   tgforms:hasInput tgforms:checkbox;
2137   tgforms:hasPriority "215";
  
```



Fuseki
Triplestore

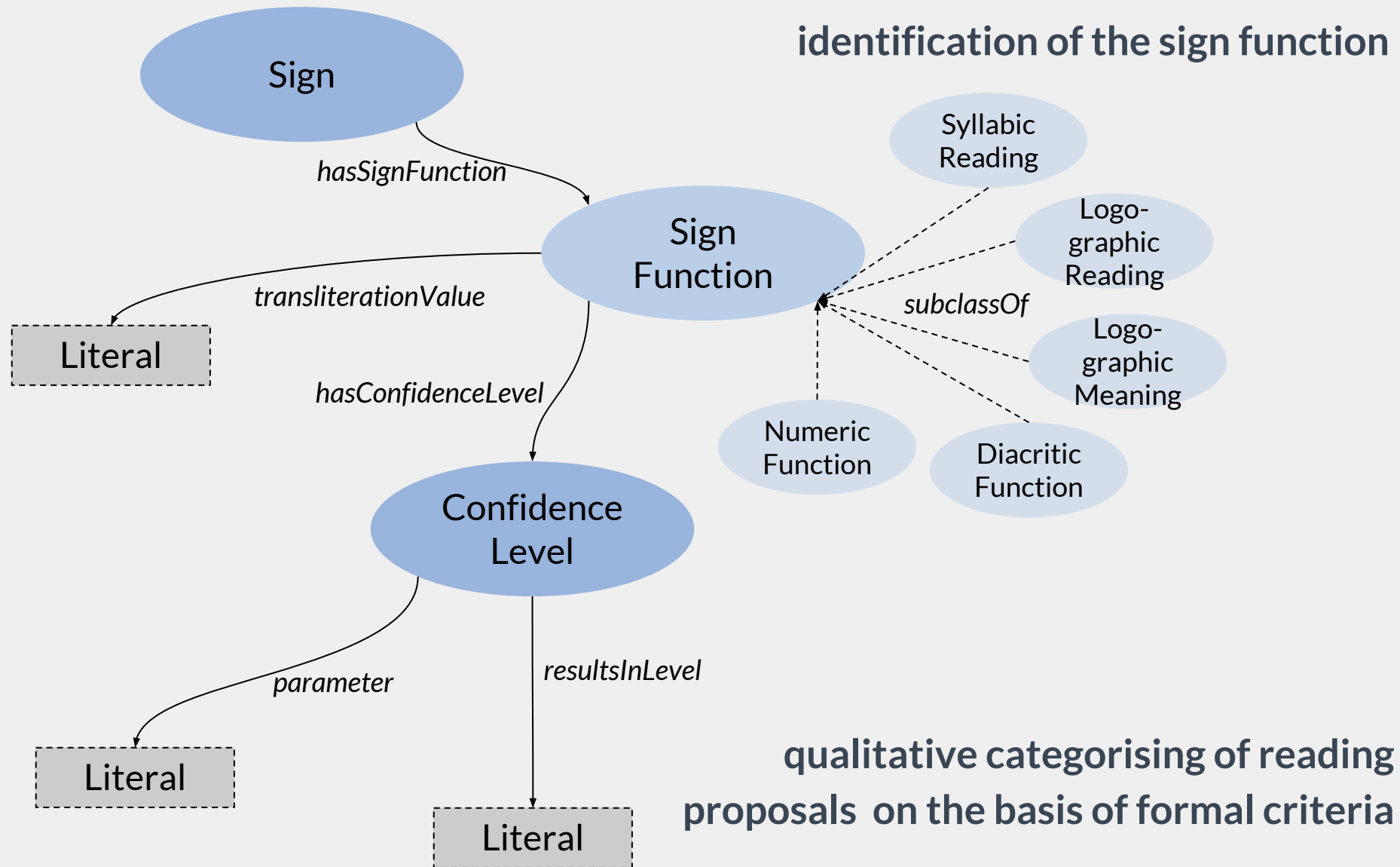
The screenshot shows the TextGridLab web interface. On the left is a 'Navigator' pane showing a project structure with folders like 'FranziTestProjekt', 'Freiburg-Geoworkshop', and 'idiom'. Under 'idiom', there are several categories like 'Modern Administrative Units', 'Archaeological Sites', etc. The 'Sign Catalogue' is expanded, showing 'Sign 0001-0100' and '0001'. The main area displays 'Sign 0001' with a 'textgrid:2skhj : Sign' header. Below this, there are fields for 'Status' (Work in Progress), 'Sign Nr.' (1), and three sections for 'has logographic reading', 'has logographic meaning', and 'has syllabic reading', each with 'Choose' and 'Remove' buttons.

RDF-Data Model
implemented in VRE
TextGrid



TextGrid
Virtuelle Forschungsumgebung
für die Geisteswissenschaften

Modelling of Reading Hypotheses



Qualitative Assessment on Basis of Formal Criteria

- one set of criteria for each sign function
- criteria are combined by means of propositional logic

criteria for logograms

d = Colonial sources

k = complete phonetic substitution

h = homophonic substitution

p = expected part of speech

t = partial phonetic substitution

o = preposed complementation [-prefix(es)]

c = postponed complementation [-suffix(es)]

u = preconsonantal ergative pronoun

y = prevocalic ergative pronoun

l = attested in GL languages

m = attested in other Mayan languages

n = attested in non-Mayan languages

i = semantic correspondance with graph icon

s = semantic correspondence with context

g = problematic postponed suffixation

→ depending on the combination, a *Confidence Level* is inferred

rules for reasoning

$$1 = d \vee h \vee (k \wedge p)$$

$$2 = ((o \vee y) \wedge c) \vee t$$

$$3 = u \wedge c \wedge (l \vee m \vee n) \wedge s \wedge i$$

$$4 = (o \vee c) \wedge (l \vee m \vee n) \wedge s$$

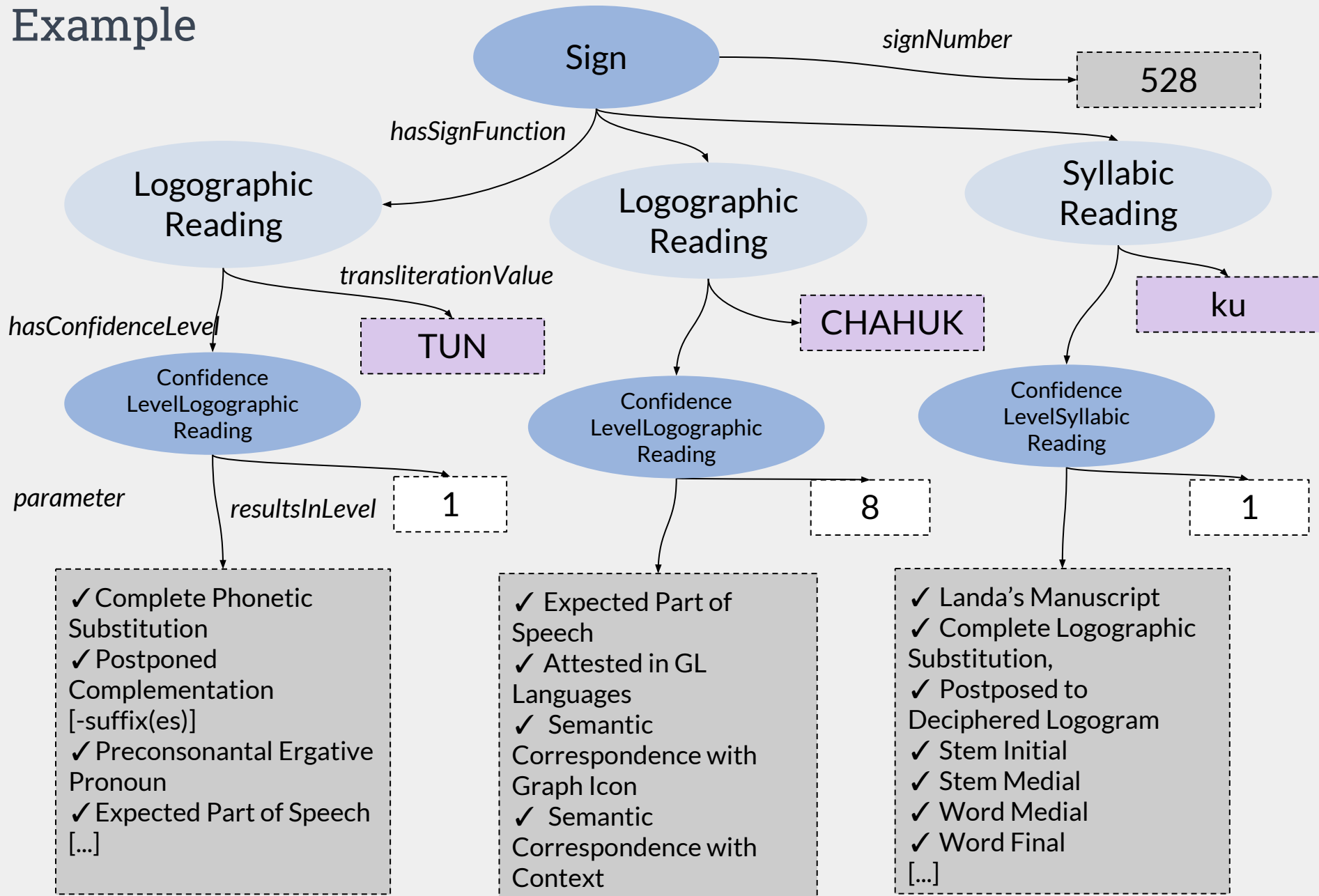
$$5 = g \wedge (l \vee m \vee n) \wedge p \wedge s$$

$$6 = g \wedge s \wedge p$$

$$7 = i \wedge s \wedge (u \vee y)$$

$$8 = i \wedge s$$

Example





sign

528

with reading

TUN

selected parameter:

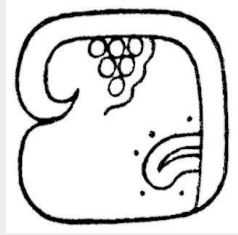
rules for reasoning:

- ✓ **k** Complete Phonetic Substitution
- ✓ c Postponed Complementation [-suffix(es)]
- ✓ u Preconsonantal Ergative Pronoun
- ✓ **p** Expected Part of Speech
- ✓ l Attested in GL Languages
- ✓ m Attested in other Mayan Languages
- ✓ i Semantic Correspondence with Graph Icon
- ✓ s Semantic Correspondence with Context

$$\mathbf{k} + \mathbf{c} + \mathbf{u} + \mathbf{p} + \mathbf{l} + \mathbf{m} + \mathbf{i} + \mathbf{s}$$

- $$\begin{aligned}
 1 &= d \vee h \vee (k \wedge p) \\
 2 &= ((o \vee y) \wedge c) \vee t \\
 3 &= u \wedge c \wedge (l \vee m \vee n) \wedge s \\
 &\quad \wedge i \\
 4 &= (o \vee c) \wedge (l \vee m \vee n) \wedge s \\
 5 &= g \wedge (l \vee m \vee n) \wedge p \wedge s \\
 6 &= g \wedge s \wedge p \\
 7 &= i \wedge s \wedge (u \vee y) \\
 8 &= i \wedge s
 \end{aligned}$$

$$d \vee h \vee (k \wedge p) = \text{Level 1}$$



sign

528

with reading

CHAHUK

selected parameter:

rules for reasoning:

- ✓ p Expected Part of Speech,
- ✓ l Attested in GL Languages,
- ✓ m Attested in other Mayan Languages,
- ✓ i Semantic Correspondence with Graph Icon
- ✓ s Semantic Correspondence with Context

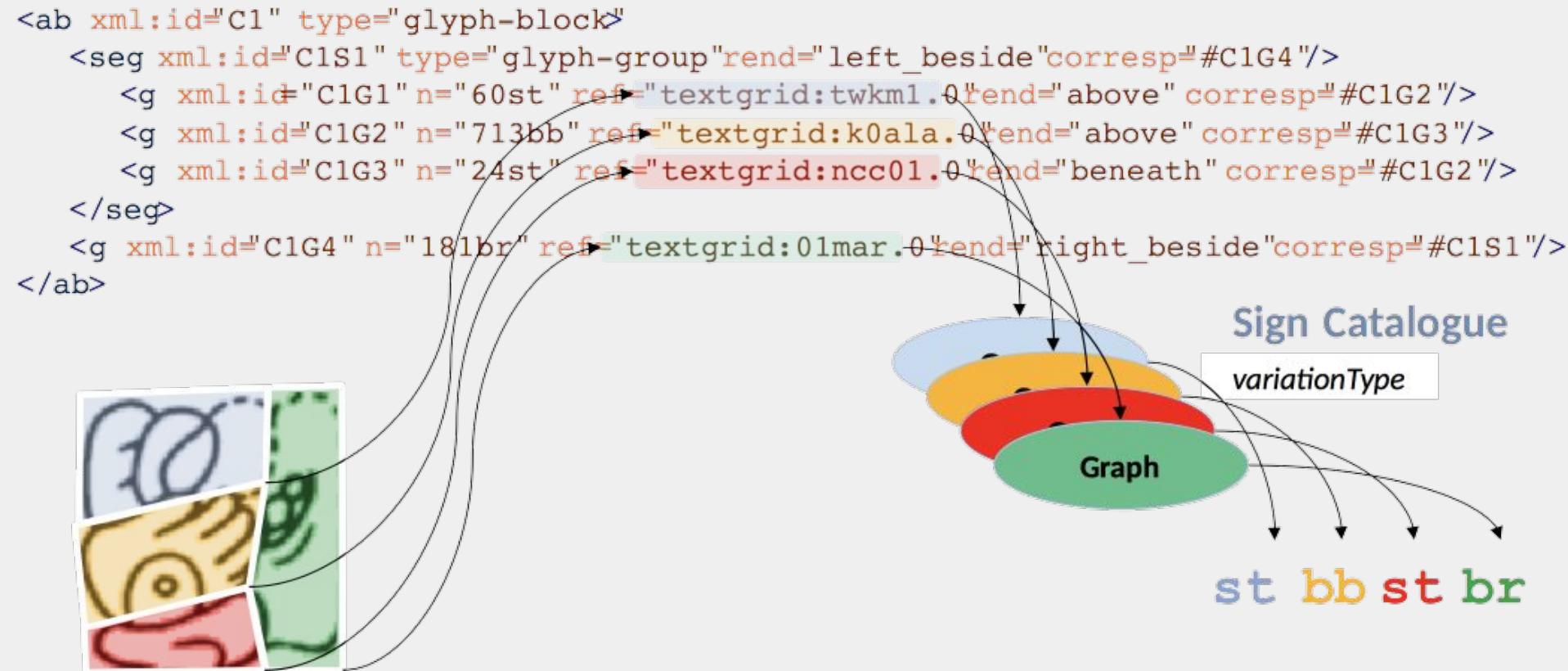
$$p + l + m + \mathbf{i} + \mathbf{s}$$

- 1 = $d \vee h \vee (k \wedge p)$
- 2 = $((o \vee y) \wedge c) \vee t$
- 3 = $u \wedge c \wedge (l \vee m \vee n) \wedge s$
 $\wedge i$
- 4 = $(o \vee c) \wedge (l \vee m \vee n) \wedge s$
- 5 = $\mathbf{g} \wedge (l \vee m \vee n) \wedge \mathbf{p} \wedge \mathbf{s}$
- 6 = $g \wedge s \wedge p$
- 7 = $i \wedge s \wedge (u \vee y)$
- 8 = $\mathbf{i} \wedge \mathbf{s}$

$$\mathbf{i} \wedge \mathbf{s} = \text{Level 8}$$

Basis for the Encoding of the Corpus

- encoding each glyph with URI of *Graph* in Sign Catalogue



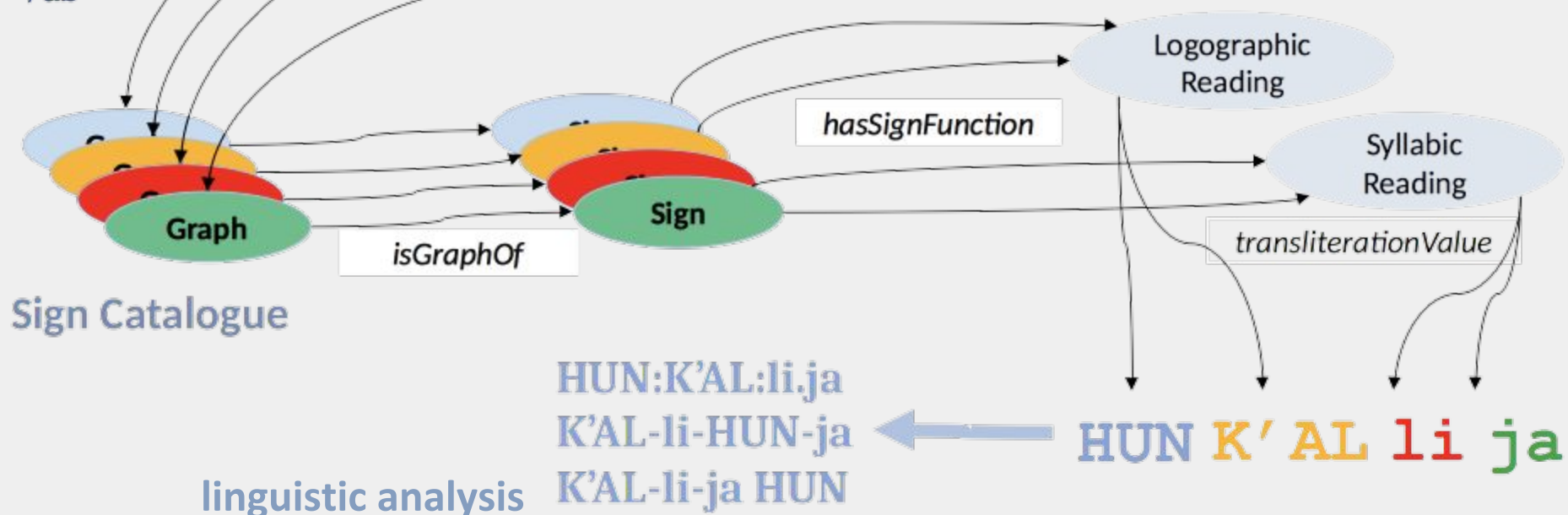
- the text becomes an ontological-interlinked entity

Generating an Analysable Text

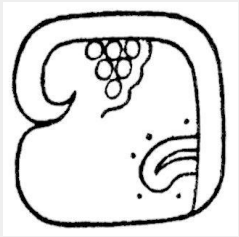
Further step of processing: querying the URIs from the Sign Catalogue
→ assignment of transliteration values

encoded corpus

```
<ab xml:id="C1" type="glyph-block">
  <seg xml:id="C1S1" type="glyph-group" rend="left_beside" corresp="#C1G4"/>
    <g xml:id="C1G1" n="60st" ref="textgrid:twkml.0" rend="above" corresp="#C1G2"/>
    <g xml:id="C1G2" n="713bb" ref="textgrid:k0ala.0" rend="above" corresp="#C1G3"/>
    <g xml:id="C1G3" n="24st" ref="textgrid:ncc01.0" rend="beneath" corresp="#C1G2"/>
  </seg>
  <g xml:id="C1G4" n="181br" ref="textgrid:01mar.0" rend="right_beside" corresp="#C1S1"/>
</ab>
```



Examination of the Readings Proposals in the Corpus



Sign
528

TUN

stone

CHAHUK

19th day name

ku

ku

references in the corpus

text A) He took the 528 from the ground.

text B) It happened on 528 when...

text C) This workshop is 528 ol!

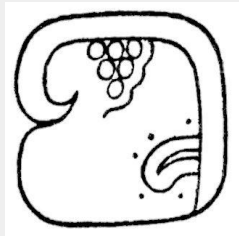
linguistic analysis

→ term for thing

→ temporal context

→ syllable

Examination of the Readings Proposals in the Corpus



Sign
528

TUN

stone

CHAHUK

19th day name

ku

ku

references in the corpus

plausibility

text A) He took the **stone** from the ground.



text B) It happened on **19th day** when...



text C) This workshop is **co**ol!



Use of sign 528 with value /CHAHUK/ only in calendar contexts!

→ **secure decipherment of glyphs and new insights on the reading of signs**

Synopsis

- **ontological model** and **RDF-implementation** for precise documentation of different states of knowledge
- **criteria-based evaluation system** supporting linguistic analysis by determining a **confidence level** that allows to compare reading hypotheses with **different plausibilities**
- **stable encoding of corpus**: text can be reprocessed when new insights are gained
- By incorporating known facts and adapting new results, our model is specifically designed to deal with **vague** and **uncertain** information in ongoing **research processes**

Further interest:

- abstracting the criteria-based approach for other cases and investigating to what extent it offers an effective method for dealing with vagueness and uncertainty

Further Information and Contact

<https://projects.gwdg.de/projects/documentations/repository>

<http://idiom-projekt.de/catalogue> Documentations

<http://mayadictionary.de/> Project Website



@idiom_project

Twitter



@idiom.project

Facebook

<https://goo.gl/FTXVZA>

Presentation (CC-BY-4.0)

Franziska Diehr

Metadata und Data Conversion Group

Göttingen State and University Library

Platz der Göttinger Sieben 1

37073 Göttingen

+49 551 39- 14070

diehr@sub.uni-goettingen.de



@FranziDiehr

References

- Diehr, Franziska, Brodhun, Maximilian, Gronemeyer, Sven, Diederichs, Katja, Prager, Christian M., Wagner, Elisabeth, and Nikolai Grube (2018). Ein digitaler Zeichenkatalog als Organisationssystem für die noch nicht entzifferte Schrift der Klassischen Maya . In C. Wartena, M. Franke-Maier, and E. De Luca, editors, Knowledge Organization for Digital Humanities - Proceedings of the 15th Conference on Knowledge Organization WissOrg'17 of the German Chapter of the International Society for Knowledge Organization (ISKO) . pages 37–43. <https://doi.org/10.17169/FUDOCs document 000000028863> .
- Diehr, Franziska, Brodhun, Maximilian, Gronemeyer, Sven, Diederichs, Katja, Prager, Christian M., Wagner, Elisabeth, and Nikolai Grube (2017) Modellierung eines digitalen Zeichenkatalogs für die Hieroglyphen des Klassischen Maya, In INFORMATIK 2017, edited by Maximilian Eibl and Martin Gaedke, pp. 1185–1196. Gesellschaft für Informatik, Bonn. doi:10.18420/in2017_120
- Gronemeyer, Sven (2014) The Orthographic Conventions of Maya Hieroglyphic Writing: Being a Contribution to the Phonemic Reconstruction of Classic Mayan. Ph.D. Dissertation, Department of Archaeology, La Trobe University, Melbourne. <http://hdl.handle.net/1959.9/505160>
- Houston, Stephen (2001). The decipherment of ancient maya writing pages 3–19.
- Kelley, David H. (1962). Review of a catalog of maya hieroglyphs by J. Eric S. Thompson. American Journal of Archaeology 66:436–438.
- Landa, Diego de (1566) Relación de las cosas de Yucatán / Fray Di[eg]o de Landa: / MDLXVI. Unpublished Manuscript ([24-B-68] 9-27-2/5153). Madrid. Biblioteca de la Real Academia de Historia.
- Macri, Martha J., and Matthew G. Looper (2003) The New Catalog of Maya Hieroglyphs: The Classic Period Inscriptions. Civilization of the American Indian Series 247. University of Oklahoma Press, Norman, OK.
- Neuroth, Heike, Rapp, Andrea and Söring, Sybille (2015). TextGrid: Von der Community für die Community. <https://doi.org/https://doi.org/10.3249/webdoc-3947> .
- Sowa, John F. (2000) Knowledge Representation: logical, philosophical, and computational foundations . Brooks/Cole.
- Thompson, J. Eric S. (1962) A Catalog of Maya Hieroglyphs. The Civilization of the American Indian Series 62. University of Oklahoma Press, Norman, OK.