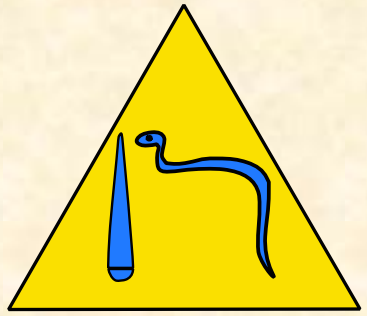
Digitalising the Pyramid Texts

grew meets 

LORIA, Nancy — 16 May 2025

Roberto Antonio Díaz Hernández

Bruno Guillaume

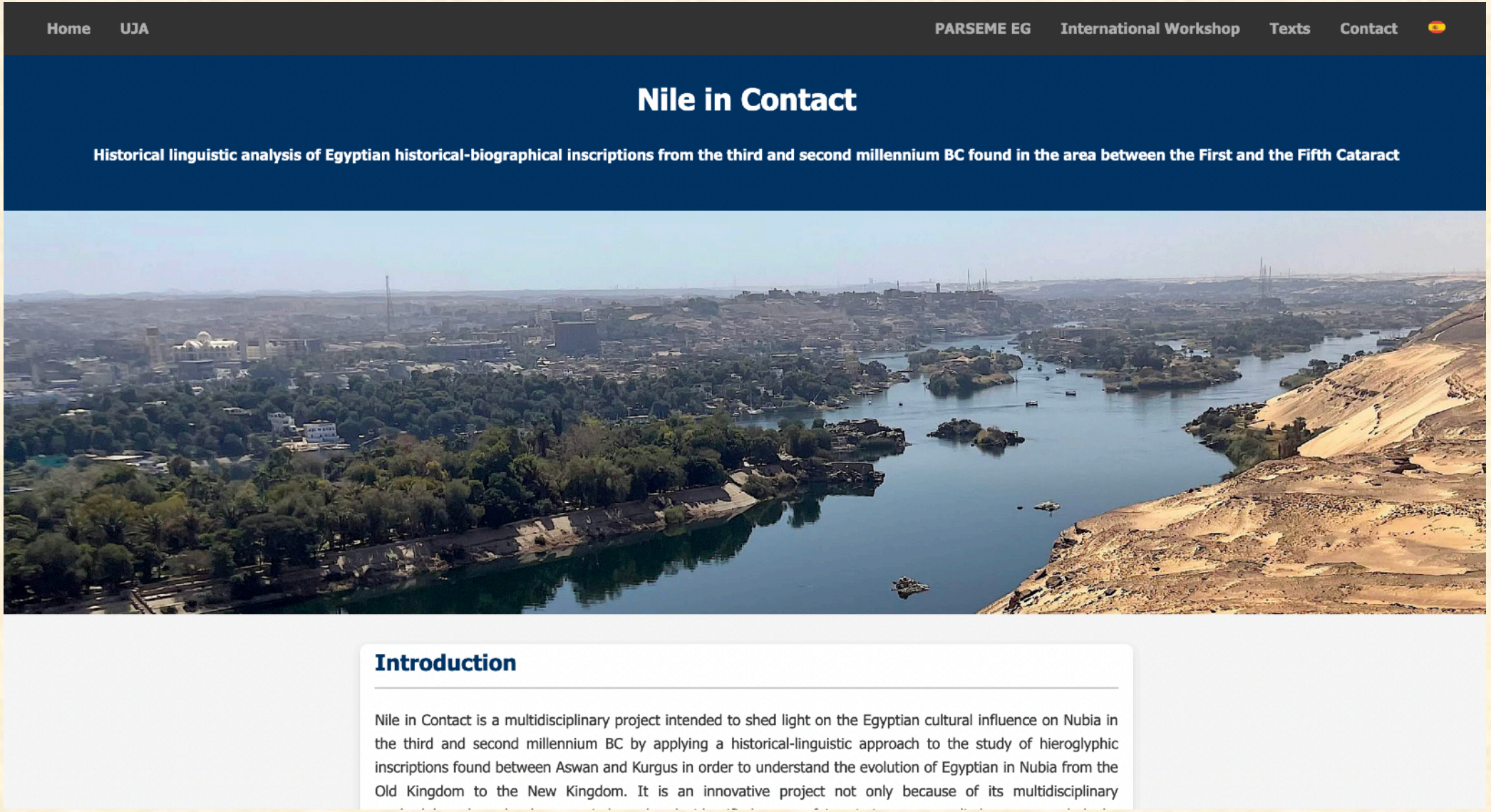
Universidad de Jaén

UniDive

Inria

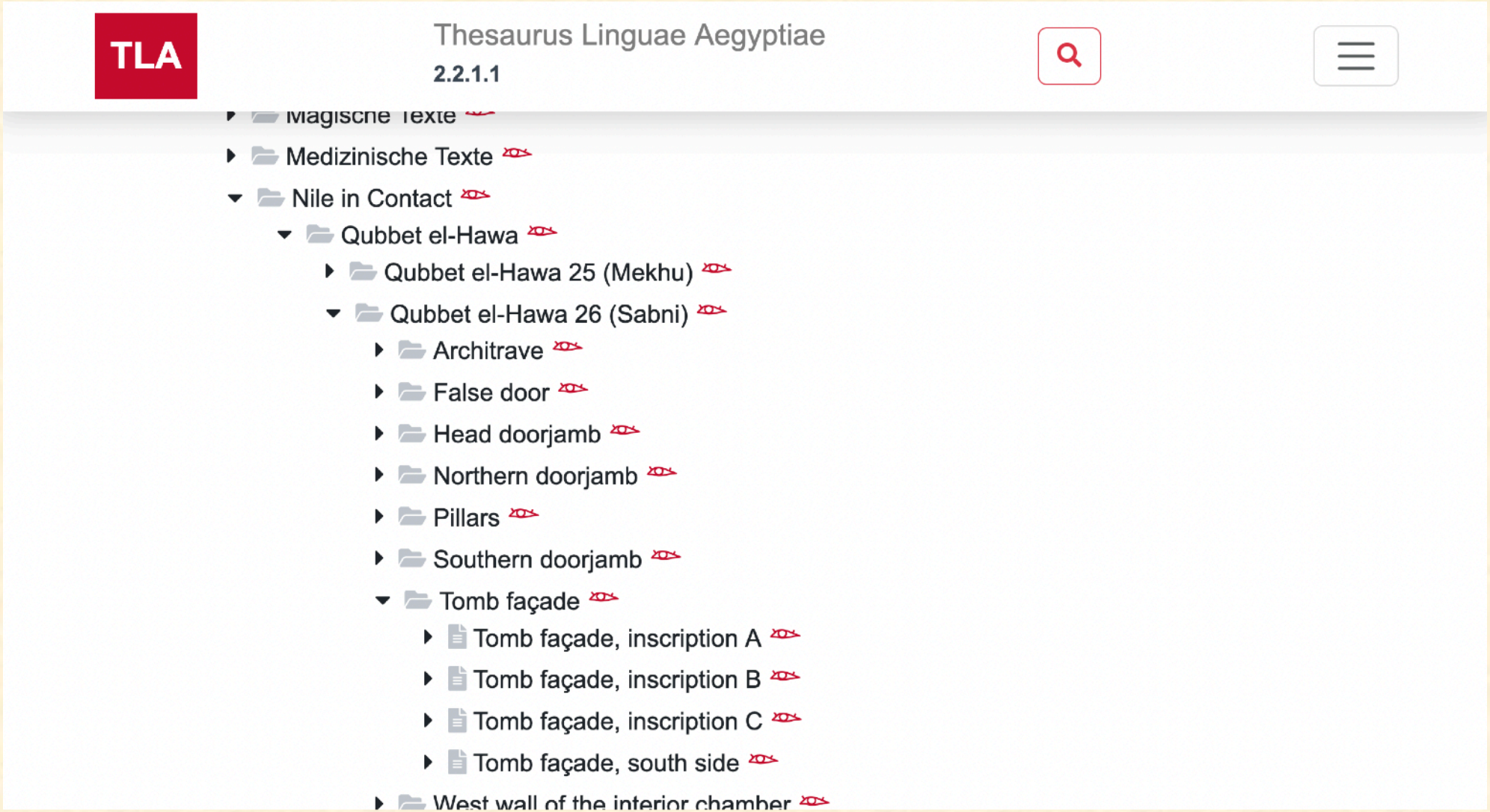
Loria

Beginning

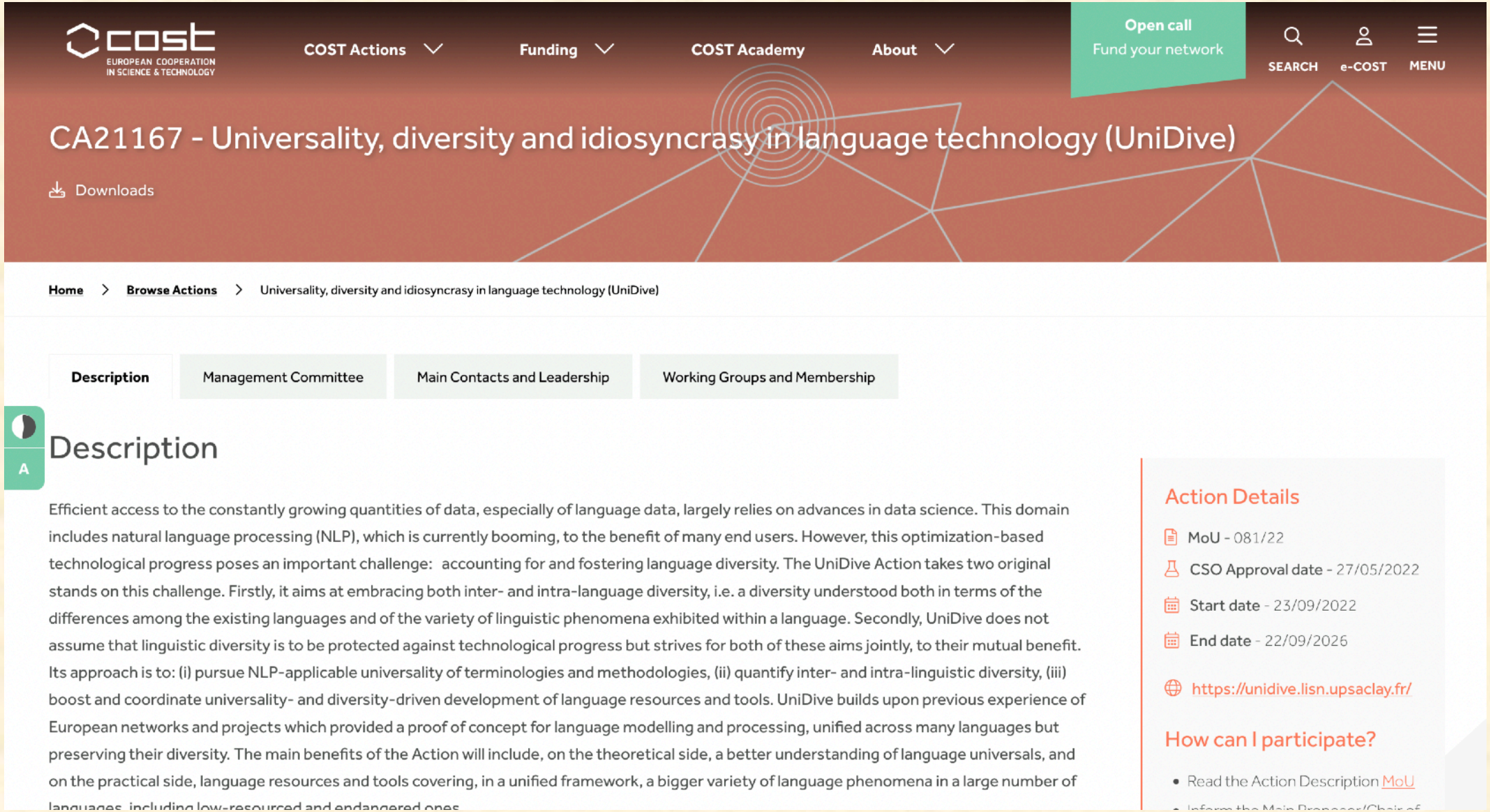


https://web.ujaen.es/investiga/nile-in-contact/


Alexandria
الإسكندرية
Cairo
القاهرة
Jerusalem
Israel
Jordan
Egypt
Madinah
المدينة المنورة
Jeddah
جدة
Makkah
مكة
Red Sea
Abha
أبها
Sudan
Khartoum
الخرطوم
Eritrea
Al Fashir
الفاشر
Google



https://thesaurus-linguae-aegyptiae.de/home



https://www.cost.eu/actions/CA21167/

UniDive Universality, diversity and idiosyncrasy in language technology
CA21167 COST Action

Log In
Search
Recent Changes Media Manager Sitemap

Trace: • start

About UniDive
- Home
- How to join us
- How to contribute
- How to acknowledge UniDive
- Related initiatives
- Contact
- News
- Logo

Organization
- Working Groups
  - WG1: Corpus annotation
  - WG2: Lexicon-corpus interface
  - WG3: Multilingual and cross-lingual language technology
  - WG4: Quantifying and promoting diversity
- Management Committee
- Core Group
- Grant Holder
- Grant periods
- Mailing lists
- Telegram and social networks

Events
- General meetings
  - Paris-Saclay
  - Naples
  - Budapest
  - Bucharest
- Workshops
  - Paris-Saclay
  - Istanbul
  - Naples
  - Turin
  - Bologna
  - Ljubljana
- Training events
  - Webinar
  - Chișinău
  - Yerevan
    - call for trainers
- Shared tasks
  - AdMIRe Vienna
  - MSP Ljubljana
  - PARSEME 2.0

start

UniDive COST Action CA21167

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY
Funded by the European Union

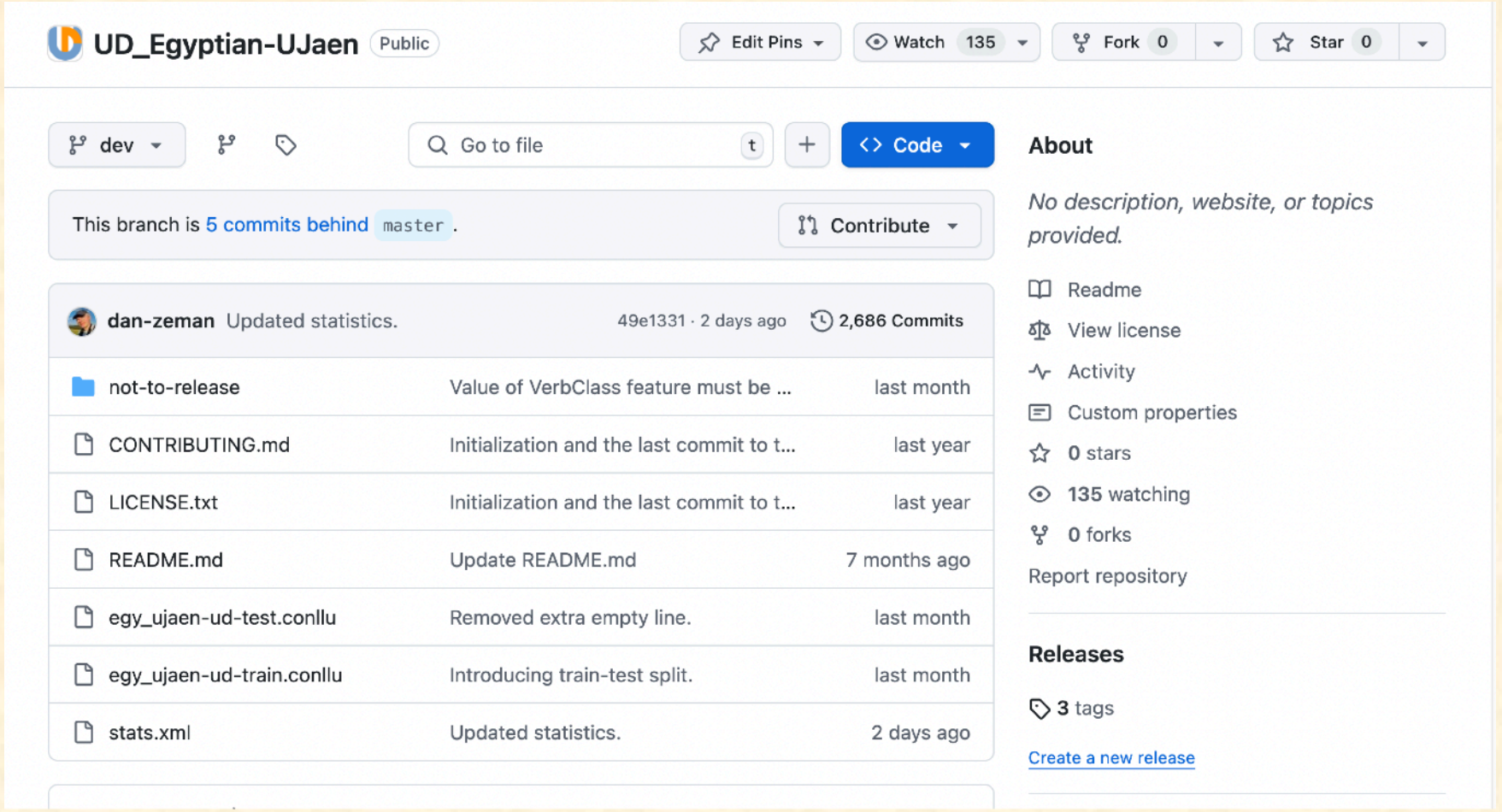
Table of Contents
- UniDive COST Action CA21167
  - Latest news
  - About UniDive
  - Basic data
  - Social Media UniDive Profiles

Latest news

- **[16 Dec 2024]**: A call for local organizers of UniDive 2026 events (the 4th general meeting and 2nd training school) has been published. Proposal submission deadline: **7 February 2025**.
- **[16 Dec 2024]**: The UniDive mailing lists hosted at the Paris-Saclay University have been **restored**.
- **[5 Nov 2024]**: The UniDive **2nd Progress Report** was submitted to e-cost today.
- **[28 Oct 2024]**: The **3rd UniDive general meeting**, co-located with a WG2 meeting, is to take place on **28-30 January 2024** in **Budapest**. If you are interested in attending, please, fill in the expression of interest form, best before **11 November**.
- **[12 Sept 2024]**: The **deadline** for abstracts to submit to the 3rd UniDive workshop in **Budapest** has been extended till the **30th of September 2024**. The **submission link** is now published, see the event's page.
- **[26 August 2024]**: Paris-Saclay University, which hosts several services for UniDive, was a victim of a major **cyber attack** on 11 August 2024. Since then, the UniDive mailing lists are not operational. The services are being progressively

https://unidive.lisn.upsaclay.fr/

Creation of the first morphosyntactic treebank for pre-Coptic Egyptian in UD (March 2024)



https://github.com/UniversalDependencies/UD_Egyptian-UJaen/tree/dev

Egyptian

Earlier Egyptian

Old Egyptian
(ca. 2700–2000 BC)

Middle Egyptian*
(ca. 2000–1550BC)

Later Egyptian

Late Egyptian
(ca. 1550–700 BC)

Demotic
(ca. 7th century BC to 5th century AD)

Coptic
(4th century to 14th century AD)

* Middle Egyptian became a standardised and classical language from 1550 BC onwards

Egyptian text corpora for the EUJA treebank

Old Egyptian

Pyramid Texts

Old Kingdom and First Intermediate Period biographical texts

Middle Egyptian

Coffin Texts

Middle Kingdom biographical texts

Literary texts

Classical Egyptian

The Book of the Dead

18th Dynasty biographical texts

Literary texts

Late Egyptian

New Kingdom biographical texts

Literary texts

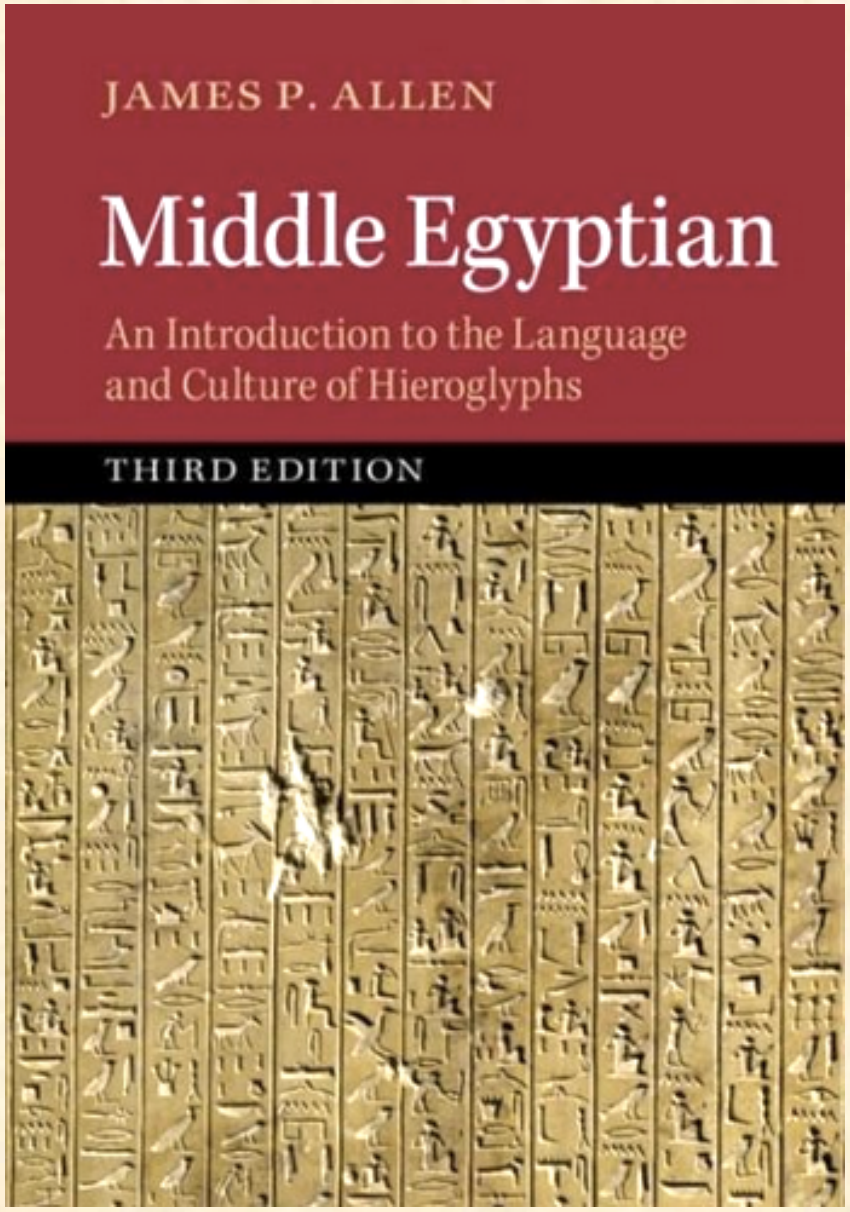
Administrative texts

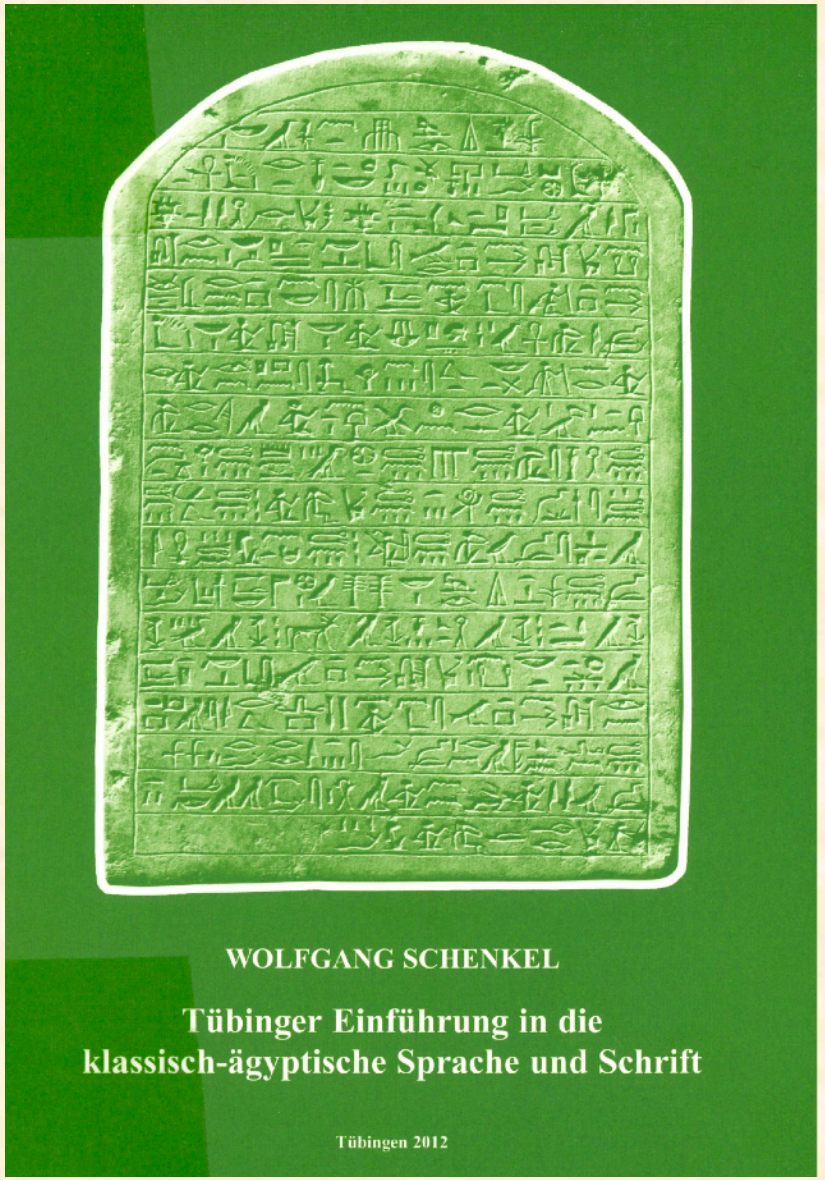
Demotic

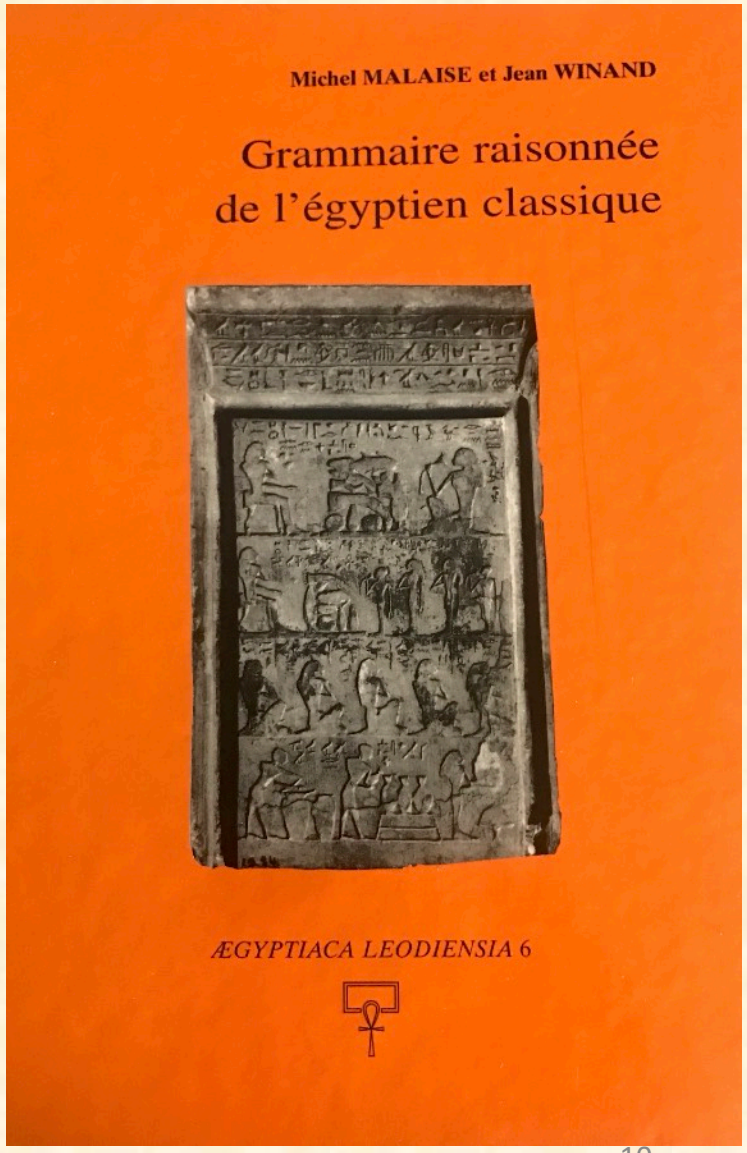
Literary texts

Administrative texts

No studies on dependency grammar in Egyptian







No digital tools for the study of pre-Coptic Egyptian grammar

Ramses Online an annotated corpus of Late Egyptian

About ▾ Corpus ▾ Searches ▾ Help **Beta Version** Register Forgotten Password login

9 A. 79; LRLC 5; P. BM 10416; P. Salt 1821/131

Text Description

Bibliography

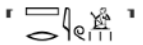
P. BM 10416 | LRLC

page 1/8

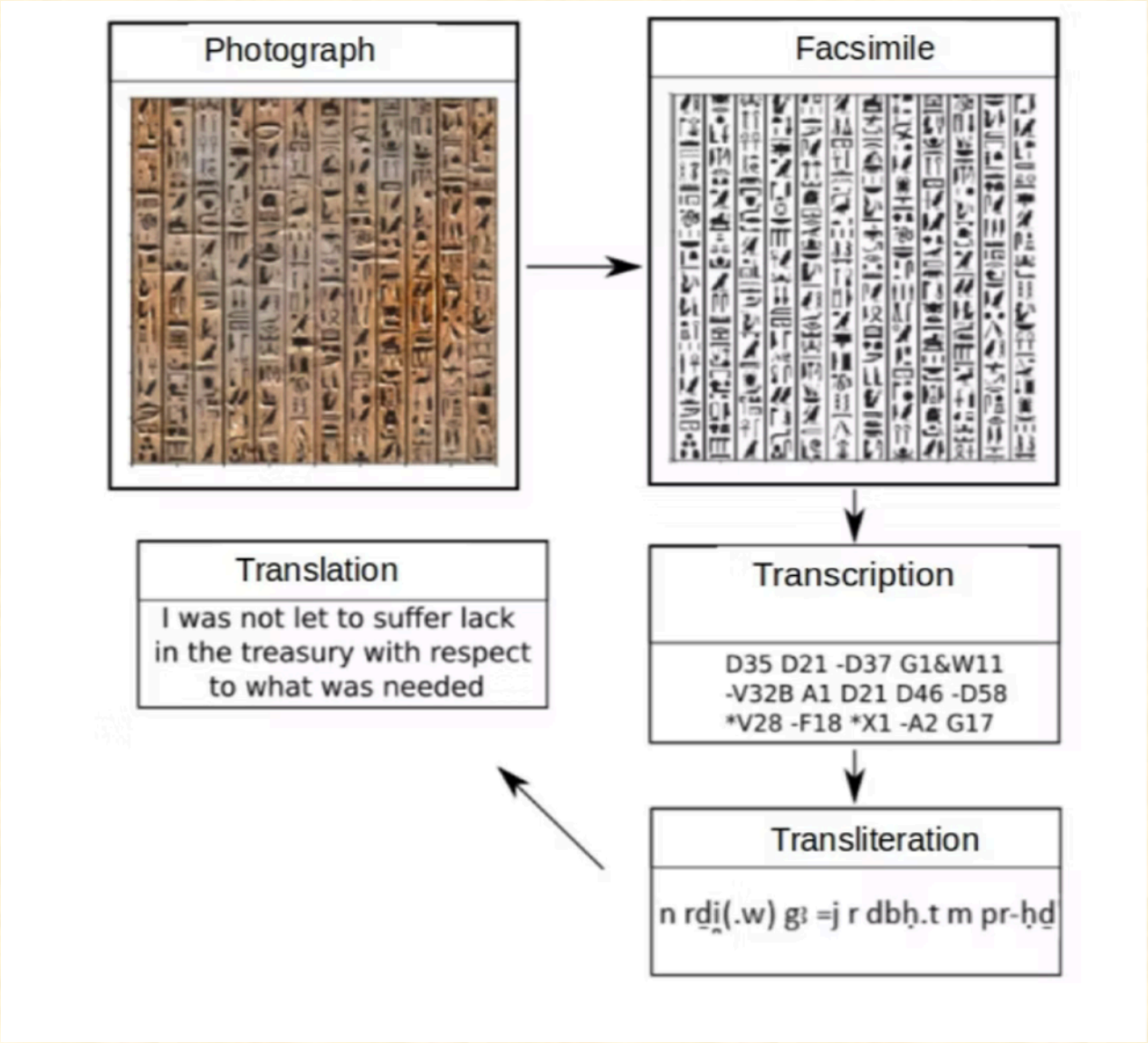
« 1 2 3 4 5 6 7 8 » 1

Cite

LRLC P. BM 10416									
		r-nty	*tfj*	*n3y.k*	*rmṯ.w*	*m*	*n3y.w*	*ᶜ3.w*	*m*
PL. 15	1	IND_INIT	VB PFV	ART.POSS	SUBST	CONNECT	ART.POSS	ADJ	CONNECT
		à savoir	*bouger, remuer, enlever*	*mes*	*homme*	*dans*	*mes*	*grand*	*dans*

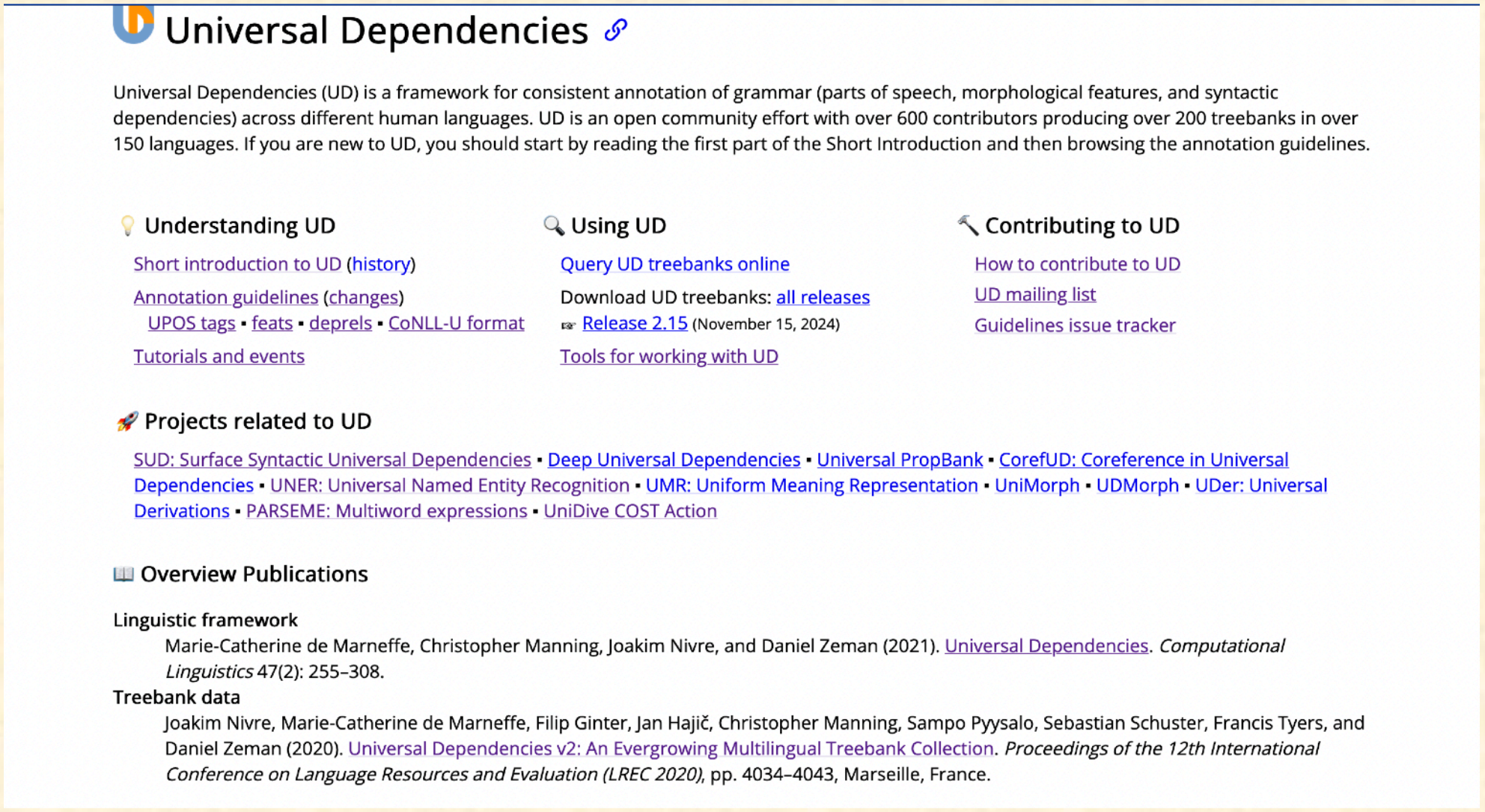
		P. BM 10416						
n3y.w	*šrj.w*		*m*	*ᶜḥ3wty.w*	*m*	*ḥm.wt*	*m*	*ḥ3wy*
ART.POSS	ADJ	2	CONNECT	SUBST	CONNECT	SUBST	CONNECT	SUBST
mes	*petit*		*dans*	*combattant, homme*	*dans*	*femme, épouse*	*dans*	*soir*

Multidisciplinary Approach to Image and Language Processing in Ancient Egyptian Texts using Transformers



https://www.mortexvar.com/ampitet

Universal Dependencies approach



Universal Dependencies

Universal Dependencies (UD) is a framework for consistent annotation of grammar (parts of speech, morphological features, and syntactic dependencies) across different human languages. UD is an open community effort with over 600 contributors producing over 200 treebanks in over 150 languages. If you are new to UD, you should start by reading the first part of the Short Introduction and then browsing the annotation guidelines.

Understanding UD

Short introduction to UD (history)

Annotation guidelines (changes)
UPOS tags ▪ feats ▪ deprels ▪ CoNLL-U format

Tutorials and events

Using UD

Query UD treebanks online

Download UD treebanks: all releases
☞ Release 2.15 (November 15, 2024)

Tools for working with UD

Contributing to UD

How to contribute to UD

UD mailing list

Guidelines issue tracker

Projects related to UD

SUD: Surface Syntactic Universal Dependencies ▪ Deep Universal Dependencies ▪ Universal PropBank ▪ CorefUD: Coreference in Universal Dependencies ▪ UNER: Universal Named Entity Recognition ▪ UMR: Uniform Meaning Representation ▪ UniMorph ▪ UDMorph ▪ UDer: Universal Derivations ▪ PARSEME: Multiword expressions ▪ UniDive COST Action

Overview Publications

Linguistic framework

Marie-Catherine de Marneffe, Christopher Manning, Joakim Nivre, and Daniel Zeman (2021). Universal Dependencies. *Computational Linguistics* 47(2): 255–308.

Treebank data

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman (2020). Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, pp. 4034–4043, Marseille, France.

https://universaldependencies.org/

1st STSM at the Catholic University of Miland (May-July of 2024)

Developing the Egyptian-UJaen Treebank

Roberto Antonio Díaz Hernández,[1] **Marco Carlo Passarotti**[2]
[1] University of Jaén (radiaz@ujaen.es)
[2] Università Cattolica del Sacro Cuore (marco.passarotti@unicatt.it)

Abstract

This paper presents preliminary results of the development of the Egyptian-UJaen treebank, the first dependency treebank created for pre-Coptic Egyptian in Universal Dependencies. It describes the current state of the treebank, explains the approach adopted for the morphosyntactic annotation and discusses some issues concerning the adoption of the CoNLL-U format for the annotation of Egyptian texts. This treebank

state of the treebank consisting of 1,573 sentences and 14,650 words (UD release 2.15 to appear on 15 November 2024).

The aim of this paper is to describe the methodology used in the development of the EUJA treebank. It provides a brief overview of Egyptian and its scripts (2) and a description of the sources selected for the treebank (3). There follows a discussion on the annotation of Egyptian texts (4) and the evaluation of an NLP model trained on the treebank (5). Finally, the next stages of the development of

https://aclanthology.org/2024.tlt-1.1/

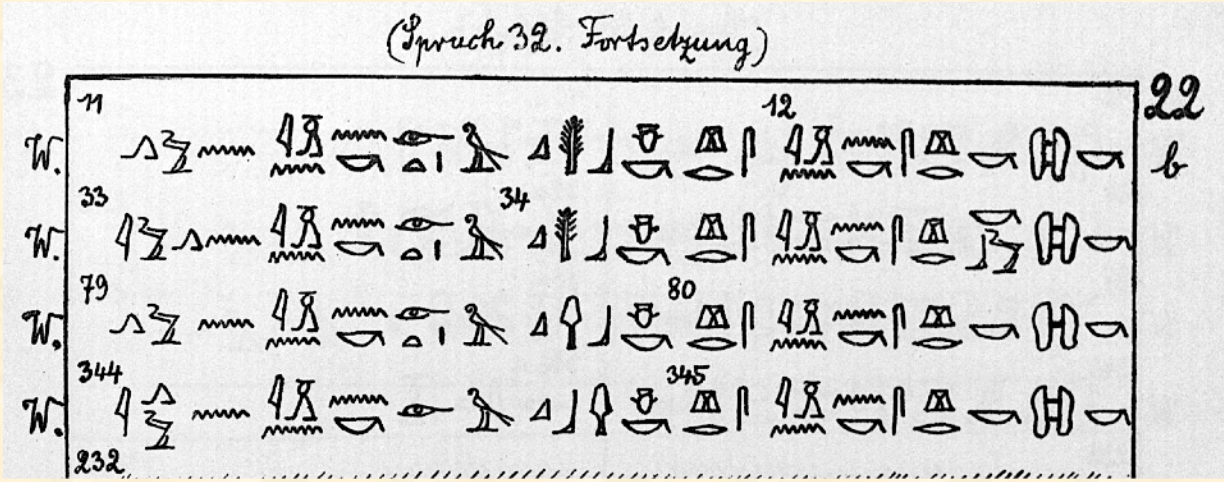
Morphosyntactic Annotation of the Pyramid Texts

What are the Pyramid Texts?

The Pyramid Texts are a collection of spells recorded on the walls of the pyramids of Old Kingdom kings. They contain more than eight hundred magic formulae recited during mortuary rituals for the king.

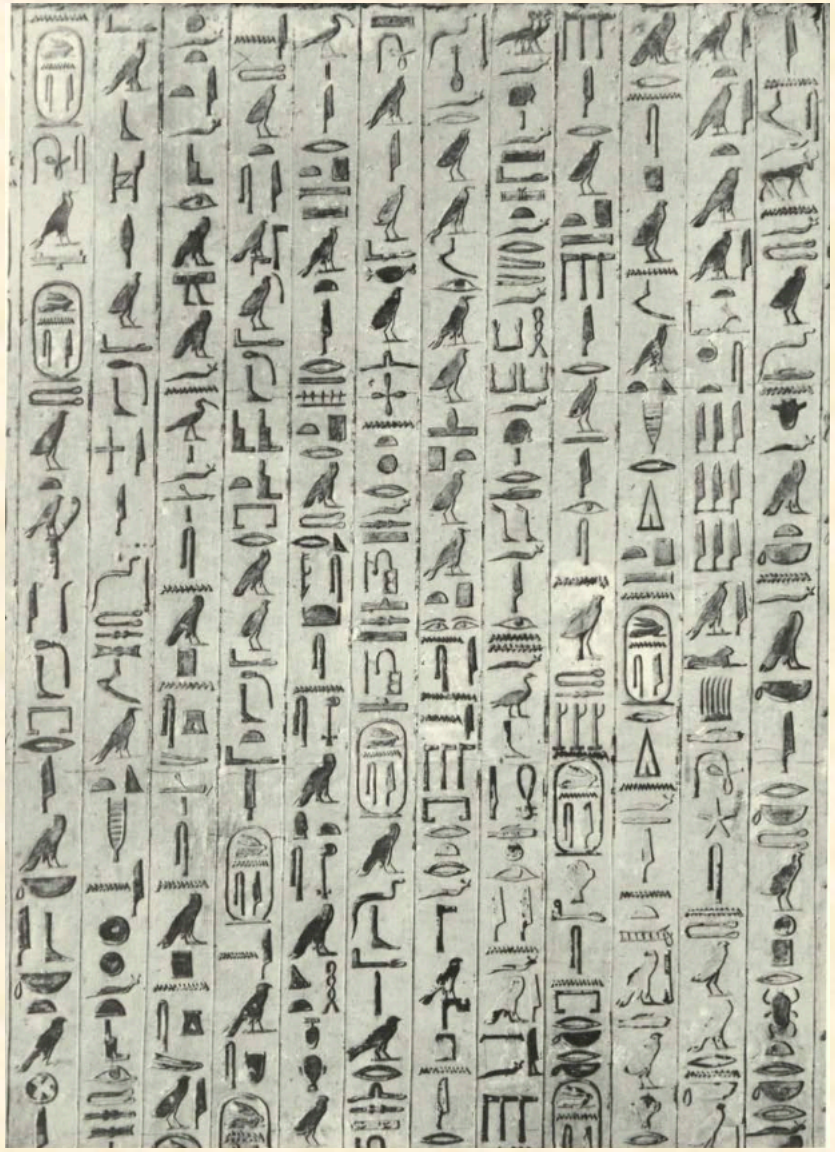
The **Pyramid of Unas** at Saqqara contains the oldest version of the Pyramid Texts, dated to 2353-2323 BC.

Unas's Pyramid Texts



Sethe's edition of the Pyramid Texts (1908-1922)

Pyramid of Unas (ca. 2300 BC)



DIE

ALTAEGYPTISCHEN PYRAMIDENTEXTE

NACH DEN

PAPIERABDRÜCKEN UND PHOTOGRAPHIEN

DES BERLINER MUSEUMS

NEU HERAUSGEGEBEN UND ERLÄUTERT

VON

KURT SETHE

ERSTER BAND

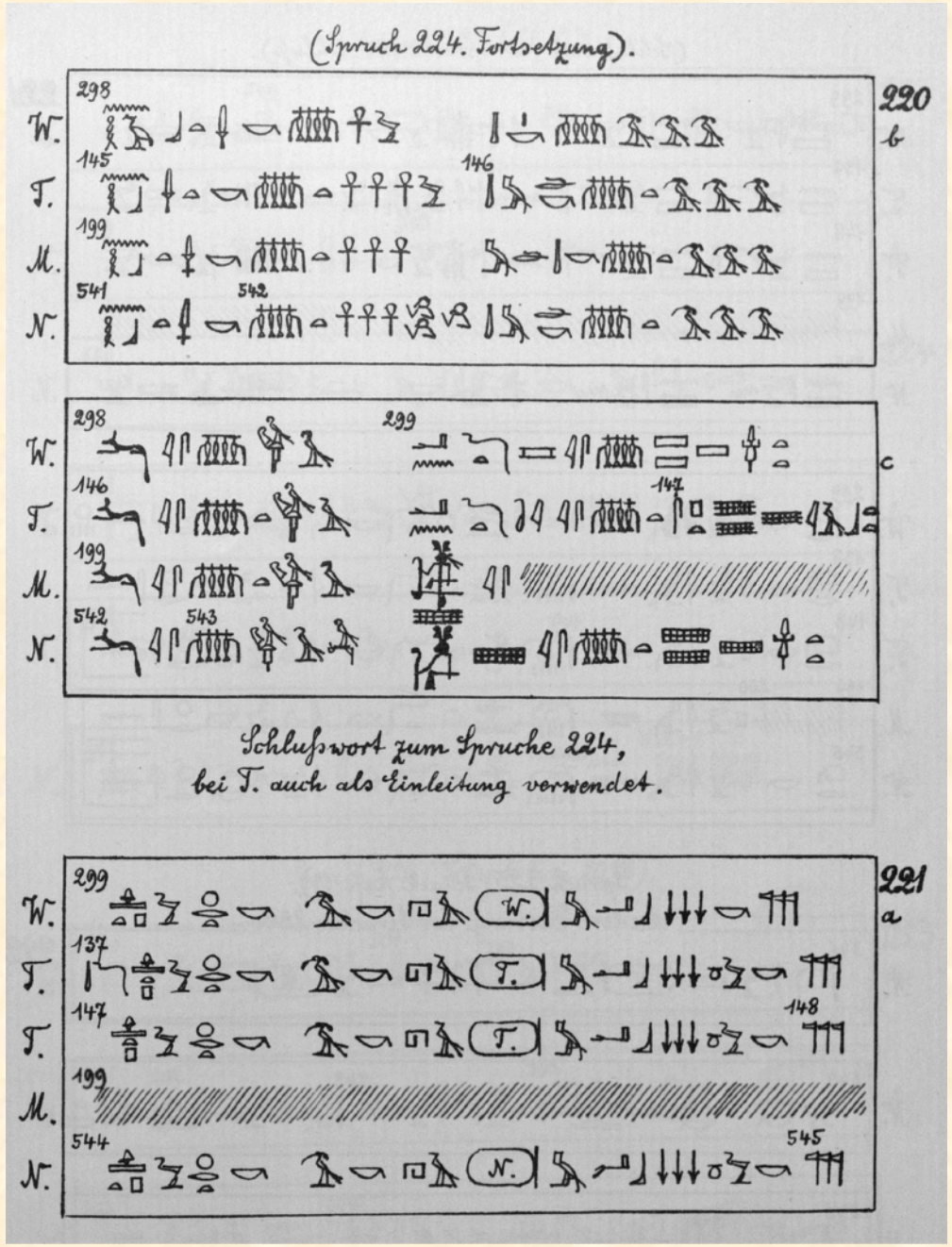
Text, erste Hälfte

Spruch 1—468 (Pyr. 1—905)

LEIPZIG

J. C. HINRICHS'SCHE BUCHHANDLUNG

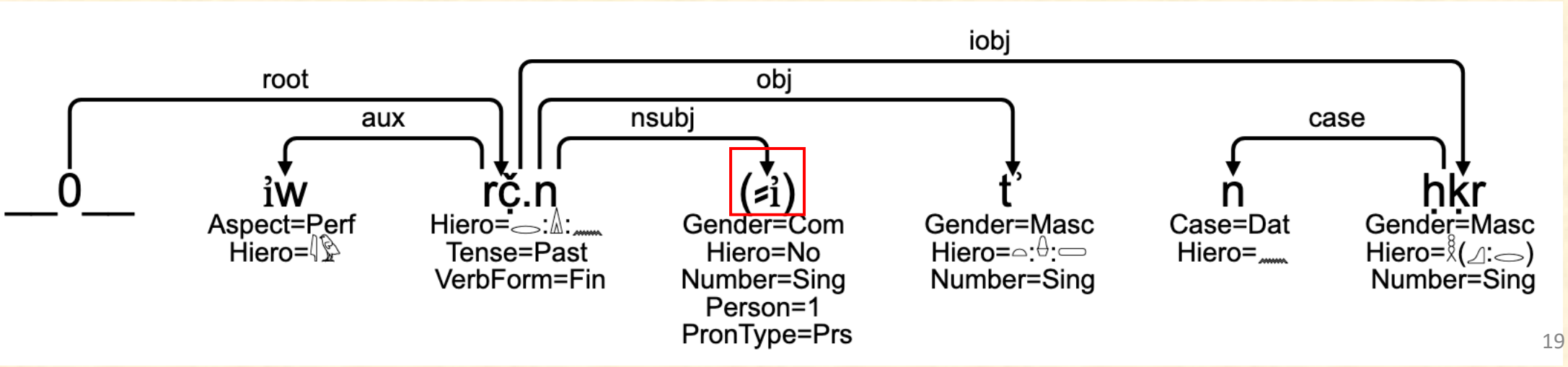
1908



Description of the columns

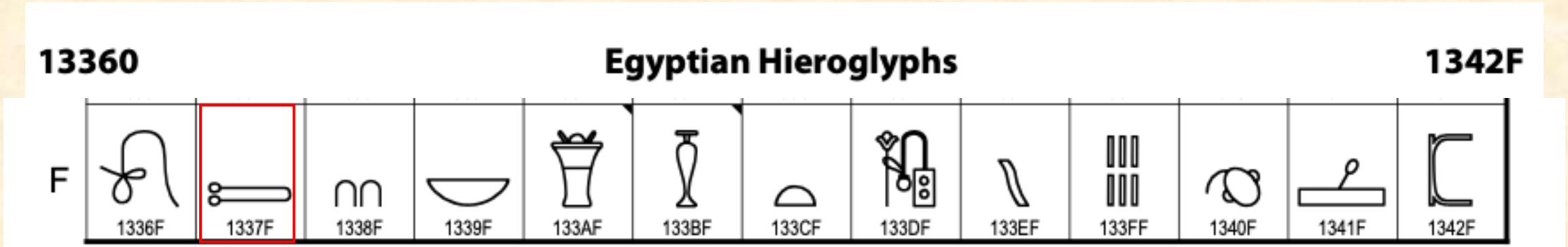
```
1    # sent_id = EUJA-1
2    # ref = Edel 2008 (eds. Seyfried/Vieler): pl. LV, line 5, Qubbet el-Hawa (QH35e), 6th Dynasty, histbio, OE
3    # text = ỉw rḳ.n (ỉ) t’ n ḥḳr
4    1   ỉw   ỉw   AUX Particle    Aspect=Perf 2   aux _   Hiero=
5    2   rḳ.n   rḳi VERB    SPC=Past-2|Type=Pred    Tense=Past|VerbForm=Fin 0   root    _   Hiero=
6    3   (ỉ)   ỉ   PRON    Pron=SFP    Gender=Com|Number=Sing|Person=1|PronType=Prs    2   nsubj   _   Hiero=No
7    4   t’   t’   NOUN    Hierocl=Yes Gender=Masc|Number=Sing 2   obj _   Hiero=
8    5   n   n   ADP Status=Cons Case=Dat    6   case    _   Hiero=
9    6   ḥḳr ḥḳr NOUN    _   Gender=Masc|Number=Sing 2   iobj    _   Hiero=
```

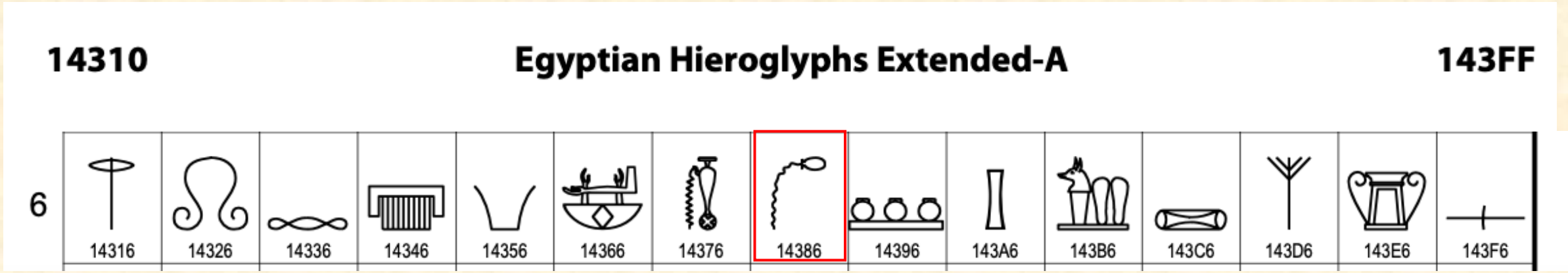
trans = (I) have given bread to the hungry



Unicode Font for Hieroglyphs

```
1085    8    s3č  s3č  VERB  _  Gender=Masc|VerbForm=Inf  1  parataxis  _  Hiero=(𓋴𓐍:𓊪)UC_14386
```





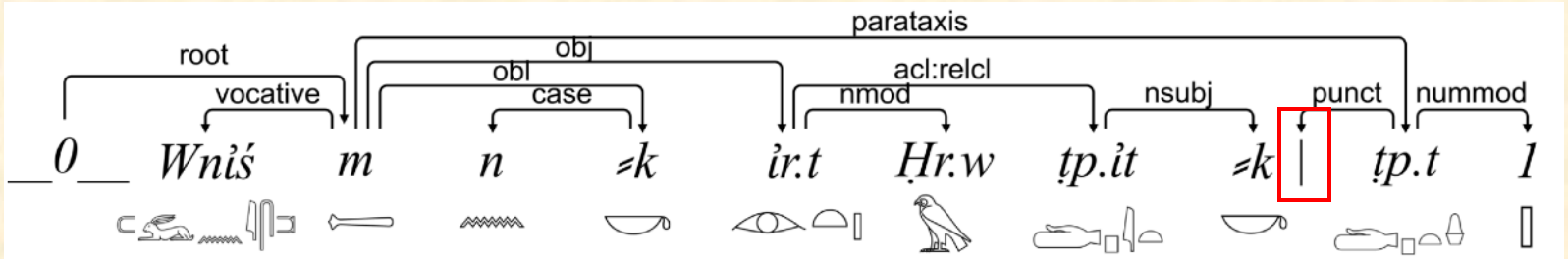
Transcription System

	LUT	Tübingen	Unicode
	ꜣ	ꜣ	A723
	i	i	A7BD
	y	y	
	ï	ï	00EF
	ꜥ	ꜥ	A725
	w	w	
	b	b	
	p	p	
	f	f	
	m	m	
	n	n	
	r	r	
	h	h	
	ḥ	ḥ	1E25
	ḫ	ḫ	1E2B
	ẖ	ẖ	1E96
	z	s	
	s	ś	015B
	š	š	0161
	q	ḳ	1E33
	k	k	
	g	g	
	t	t	
	ṯ	č	010D
	ḏ	ṭ	1E6D
	ḏ	ç̌	010D+0323

= [ʧ] = **č** voiceless postalveolar affricate like in Czech
≠ ***ṯ*** used for /θ/ (ث) in Arabic

= [t'] = **ṭ** alveolar and dental ejective stops like in Ge'ez

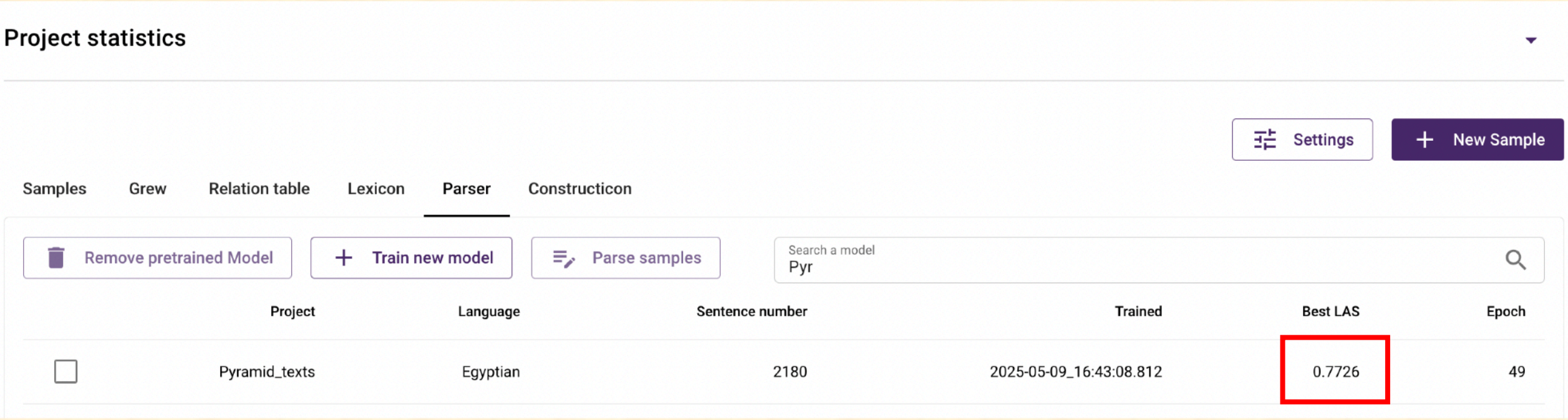
≠ ***d***



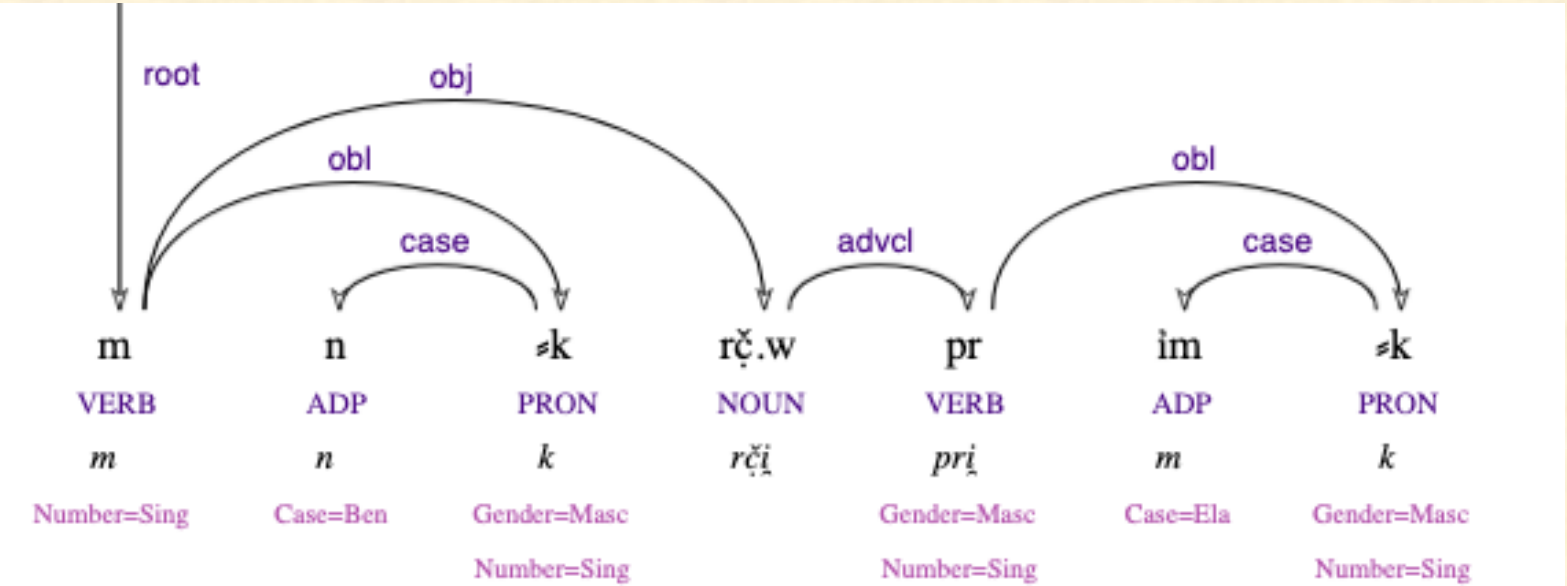
"Unas, take the Eye of Horus which you shall taste—one cake."

PT 38a (= EUJA-178), Piankoff, 1968, pl. 67

Parsing


Project statistics
Settings
New Sample
Samples
Grew
Relation table
Lexicon
Parser
Constructicon
Remove pretrained Model
Train new model
Parse samples
Search a model
Pyr
Project
Language
Sentence number
Trained
Best LAS
Epoch
Pyramid_texts
Egyptian
2180
2025-05-09_16:43:08.812
0.7726
49

Semi-automatic annotation using the trained model




```
# sent_id = New sentences_2 (not yet ready)__16
# text = m n ≠k rc̗.w pr ìm ≠k
# timestamp = 0
# user_id = parser
1	m	m	VERB	_	Mood=Imp|Number=Sing|VerbClass=Def	0	root	_	_
2	n	n	ADP	_	Case=Ben	3	case	_	_
3	≠k	k	PRON	_	Gender=Masc|Number=Sing|Person=2|PronType=Emp	1	obl	_	_
4	rc̗.w	rc̗i̗	NOUN	_	VerbClass=3aeinf|VerbForm=Inf	1	obj	_	_
5	pr	pri̗	VERB	_	Gender=Masc|Number=Sing|Tense=Past|VerbClass=3aeinf|VerbForm=Part|Voice=Act	4	advcl	_	_
6	ìm	m	ADP	_	Case=Ela	7	case	_	_
7	≠k	k	PRON	_	Gender=Masc|Number=Sing|Person=2|PronType=Prs	5	obl	_	_
```

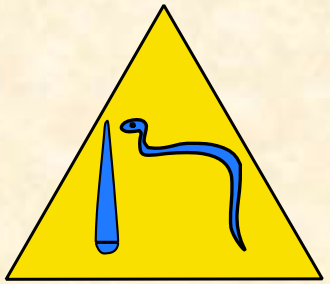
Improving the data encoding

```
# sent_id = EUJA-270
# ref = Pyramid Texts § 64d|81a W, Saqqara, 5th Dynasty, rel, OE
# text = Wśr(.w) Wniś m n ≠k św.t(i)t ir.t Ḥr.w | św.t ḥnw.t 1
# lit = "Osiris Unas, take for yourself the full equivalent (?) of the Eye of Horus – shank of meat, one bowl."
# trans = "Osiris Unas, take the full equivalent (?) of the Eye of Horus – one bowl with shank of meat."
# type = Verbal sentence
# comment = No
1 Wśr(.w) Wśr.w PROPN _ Gender=Masc 3 vocative _ Hiero=:|Name=God
2 Wniś Wniś PROPN _ Gender=Masc 1 appos _ Hiero=(:)|Name=King
3 m m VERB _ Mood=Imp|Number=Sing|VerbClass=Def 0 root _ Hiero=
4 n n ADP _ Case=Ben 5 case _ Hiero=|Status=Pron
5 ≠k k PRON _ Gender=Masc|Number=Sing|Person=2|PronType=Emp 3 obl _ Hiero=|Pron=SFP
6 św.t(i)t św.tit NOUN _ Gender=Fem|Number=Sing 3 obj _ Hiero=(:)
7 ir.t ir.t NOUN _ Gender=Fem|Number=Sing 6 nmod:poss _ Hiero=(:)
8 Ḥr.w Ḥr.w PROPN _ Gender=Masc 7 nmod:poss _ Hiero=(:)|Name=God
9 | | PUNCT _ _ 10 punct _ _|LINE
10 św.t św.t NOUN _ Gender=Fem|Number=Sing 3 parataxis _ Hiero=|Hierocl=Yes
11 ḥnw.t ḥnw.t NOUN _ Gender=Fem|Number=Sing 10 appos _ Hiero=
12 1 1 NUM _ _ 11 nummod _ NumType=Card|Hiero=
```

```
# sent_id = EUJA-270
# title = PT
# spell = 96
# section = 64d
# king = Unas
# date = 5th Dynasty
# lang = Old_Egyptian
# genre = religious
# place = Saqqara
# text = Wśr(.w) Wniś m n ≠k św.t(i)t ir.t Ḥr.w | św.t ḥnw.t 1
# lit = "Osiris Unas, take for yourself the full equivalent (?) of the Eye of Ho
# trans = "Osiris Unas, take the full equivalent (?) of the Eye of Horus – one b
# type = Verbal sentence
# comment = No
# ref = Pyramid Texts § 96/64d W => 81a W, Saqqara, 5th Dynasty, rel, OE
1 Wśr(.w) Wśr.w PROPN _ Gender=Masc 3 vocative _ Hiero=:|Name=God
2 Wniś Wniś PROPN _ Gender=Masc 1 appos _ Hiero=(:)|Name=King
3 m m VERB _ Mood=Imp|Number=Sing|VerbClass=Def 0 root _ Hiero=
4 n n ADP _ Case=Ben 5 case _ Hiero=|Status=Pron
5 ≠k k PRON _ Gender=Masc|Number=Sing|Person=2|PronType=Emp 3 obl _ Hiero=|P
6 św.t(i)t św.tit NOUN _ Gender=Fem|Number=Sing 3 obj _ Hiero=(:)
7 ir.t ir.t NOUN _ Gender=Fem|Number=Sing 6 nmod:poss _ Hiero=(:)
8 Ḥr.w Ḥr.w PROPN _ Gender=Masc 7 nmod:poss _ Hiero=(:)|Name=God
9 | | PUNCT _ _ 10 punct _ LINE=Punct
10 św.t św.t NOUN _ Gender=Fem|Number=Sing 3 parataxis _ Hiero=|Hiero
11 ḥnw.t ḥnw.t NOUN _ Gender=Fem|Number=Sing 10 appos _ Hiero=
12 1 1 NUM _ NumType=Card 11 nummod _ Hiero=
```

Building a dedicated Grew-match instance

https://pt.grew.fr



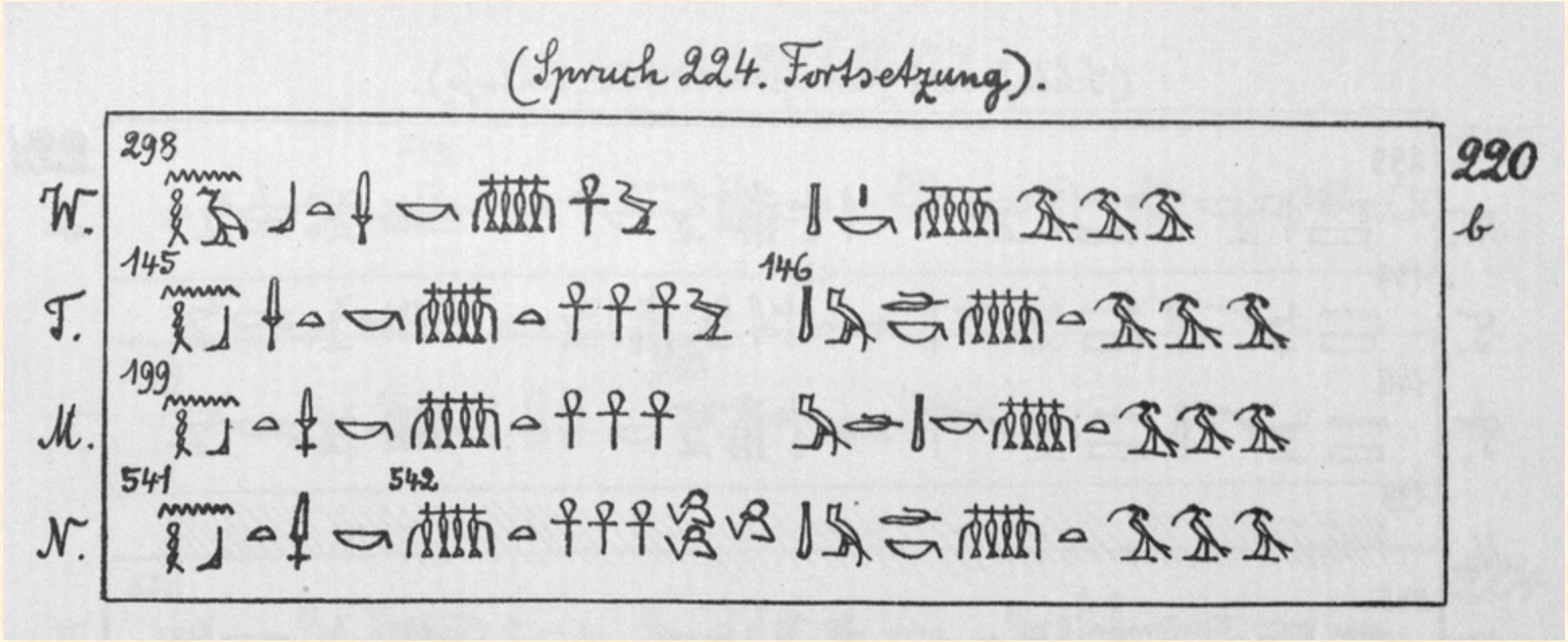
Already available:

- Images display (ongoing task)
- Multi-treebanks grew-match requests. Ex: different spelling for "Horus"
- Dedicated snippets

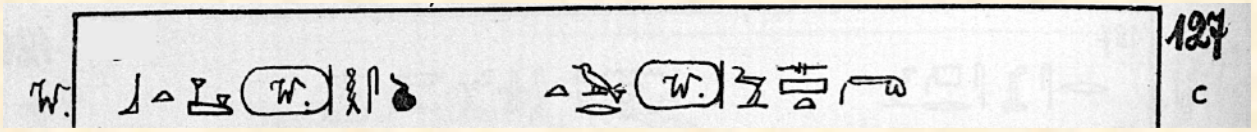
More to come:

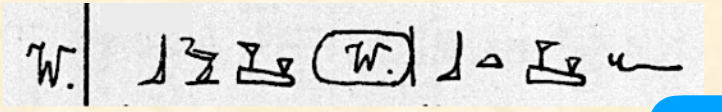
- Simultaneous display of parallel sentences

Keeping parallel data

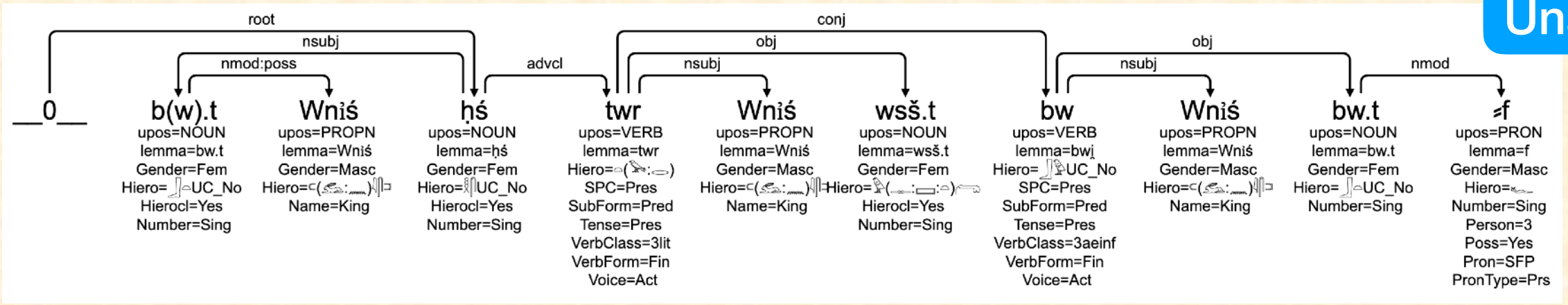


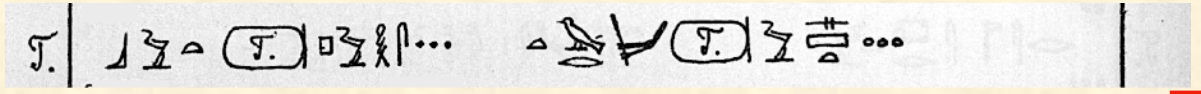
Keeping parallel data

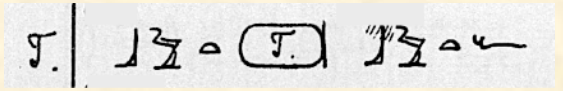




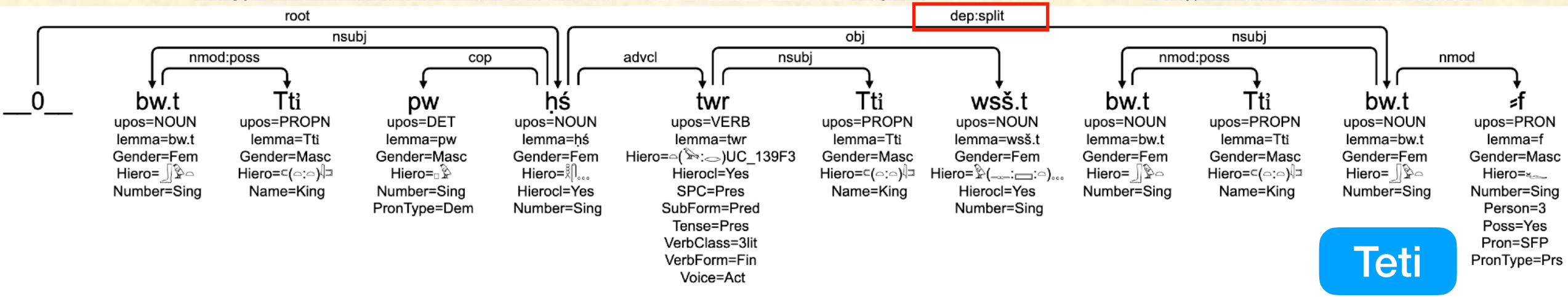
Unas







Teti

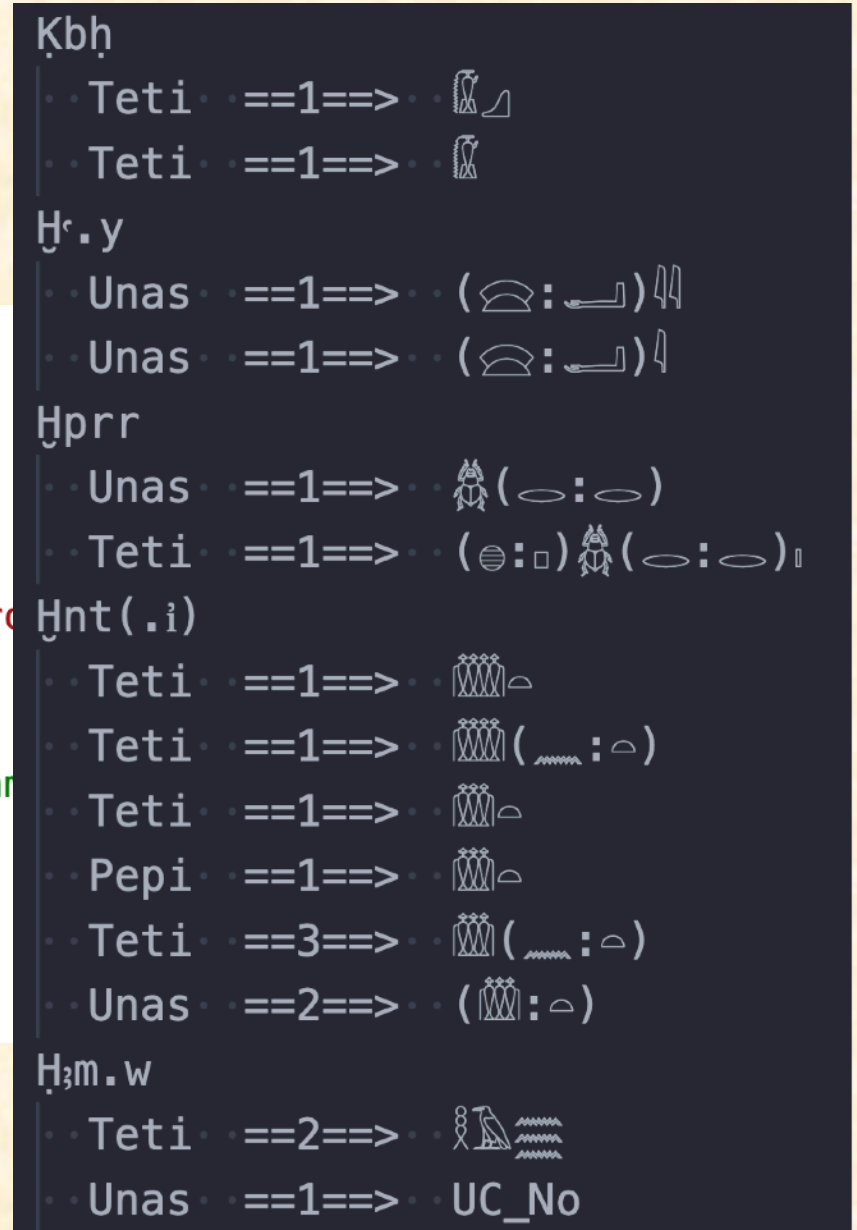


Grewpy on PT data

- Search for ambiguous spelling

```
from grewpy import Corpus, Request, set_config
set_config ("pt")

corpus = Corpus("egy_ujaen_PT.conllu")
req = Request("pattern {X [upos=PROPN, form, Hiero]}")
occ = corpus.count (req, clustering_parameter=["X.form", "X.Hiero
for form, hiero_dict in occ.items():
  if len(hiero_dict) > 1:  # More than one Hiero form for the sam
    print(f'{form}')
    for hiero, sub_dict in hiero_dict.items():
      for king, size in sub_dict.items():
        print(f'  {king}  =={size}==>  {hiero}')
```



Grammar exploration in PT • Grex

- Our **question**: what are the main (syntactic) differences between **Unas** and **Teti**?
- **Grex**: Given a scope S and a binary question Q;
  what are the patterns P that trigger Q (or $\neg Q$)

$$S \Rightarrow (P \Rightarrow Q)$$

- We need to define a "search space" for P

Santiago Herrera, Caio Corro, and Sylvain Kahane. 2024. Sparse Logistic Regression with High-order Features for Automatic Grammar Rule Extraction from Treebanks. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15114–15125, Torino, Italia. ELRA and ICCL.

Grammar exploration in PT • Grex

$$S \Rightarrow (P \Rightarrow Q)$$

Annotation differences around ADP (adposition)

- Corpus: **Unas** (11,660 tokens) + **Teti** (8616 tokens)
- S: `pattern { X [upos=ADP] }`
- Q: `king: "Teti"`
- P:
  - The features `Prefix`, `SubForm`, `VerbClass`, `Voice`, `Mood`, `Hierocl` on `X`;
  - The features `Prefix`, `SubForm`, `VerbClass`, `Voice`, `Mood`, `Hierocl` on the governor of `X`;
  - The relation between the governor of `X` and `X` ;
  - The relative position between the the governor of `X` and `X` ;

Grammar exploration in PT • Grex

Annotation difference around ADP (adposition)

Data viewer

Select data

PT_ADP

Intercepts:

[...]

Scope S: pattern { X [upos=ADP] }

pattern Q: king=Teti

S occurrences: 2592

Q occurrences: 1103

pattern	n_pattern_occurences	n_pattern_positive_occurences	n_pattern_negative_occurrences	precision	ratio	coverage	alpha	
node:X:parent:Hierocl=Yes	512	253	259	49.414062	0.977	22.937	0.013	
node:X:parent:upos=PROPN	434	217	217	50.000000	1.000	19.674	0.012	yes
node:X:own:rel_shallow=case,node:X:parent:upos=NOUN	1097	439	658	59.981768	1.499	44.191	0.011	no
node:X:own:rel_shallow=case,node:X:parent:upos=PROPN	430	215	215	50.000000	1.000	19.492	0.007	yes
node:X:parent:Hierocl=Yes,node:X:parent:upos=NOUN	436	212	224	48.623853	0.946	19.220	0.004	yes
node:X:own:rel_shallow=case	2538	1071	1467	57.801418	1.370	98.522	0.003	
node:X:own:rel_shallow=case,node:X:parent:Hierocl=Yes	504	249	255	49.404762	0.976	22.575	0.003	
node:X:parent:upos=NOUN	1109	445	664	59.873760	1.492	44.594	0.002	no
node:X:parent:Voice=Act,node:X:own:rel_shallow=case	29	5	24	82.758621	4.800	1.612	0.002	no
node:X:own:rel_shallow=fixed	16	12	4	75.000000	3.000	1.088	0.001	yes

Classifiers tend to be more used in more recent data (Teti)

PROPN are overrepresented among ADP governor in Teti

NOUN are overrepresented among ADP governor in Unas

Grammar exploration in PT • Grex

Annotation difference around ADP (adposition)

NOUN are overrepresented among ADP governor in Unas

PROPN are overrepresented among ADP governor in Teti

More general observation: POS of governors of ADP

```
pattern { P -> X; X [upos=ADP] }
```

clustered by P.upos

Corpus \ P.upos	42.79% NOUN	36.96% PRON	16.74% PROPN
100% PT_Unas	44.59%	37.54%	14.57%
100% PT_Teti	40.34%	36.17%	19.67%

Grammar exploration in PT • Contrastive patterns

Join work with **Guillaume Bonfante**

```
VERB -[parataxis]-> NOUN
```

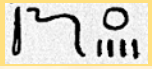
T:39, U:116

$\cos = 0.475$

```
* -[parataxis]-> VERB -[parataxis]-> NOUN
```

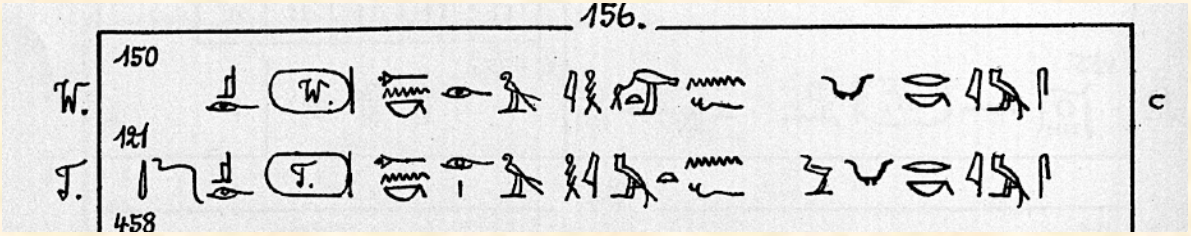
T: 29, U: 5

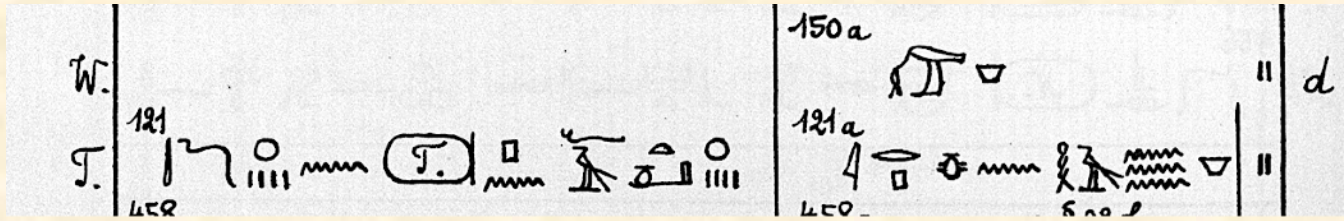
Offering texts are more frequent in Unas
Recitation text | object to offer

 **: collocation used only in Teti**

Among the 5 Unas occurrences:

- **one metadata error, it should be Teti**
- **one annotation to revise**

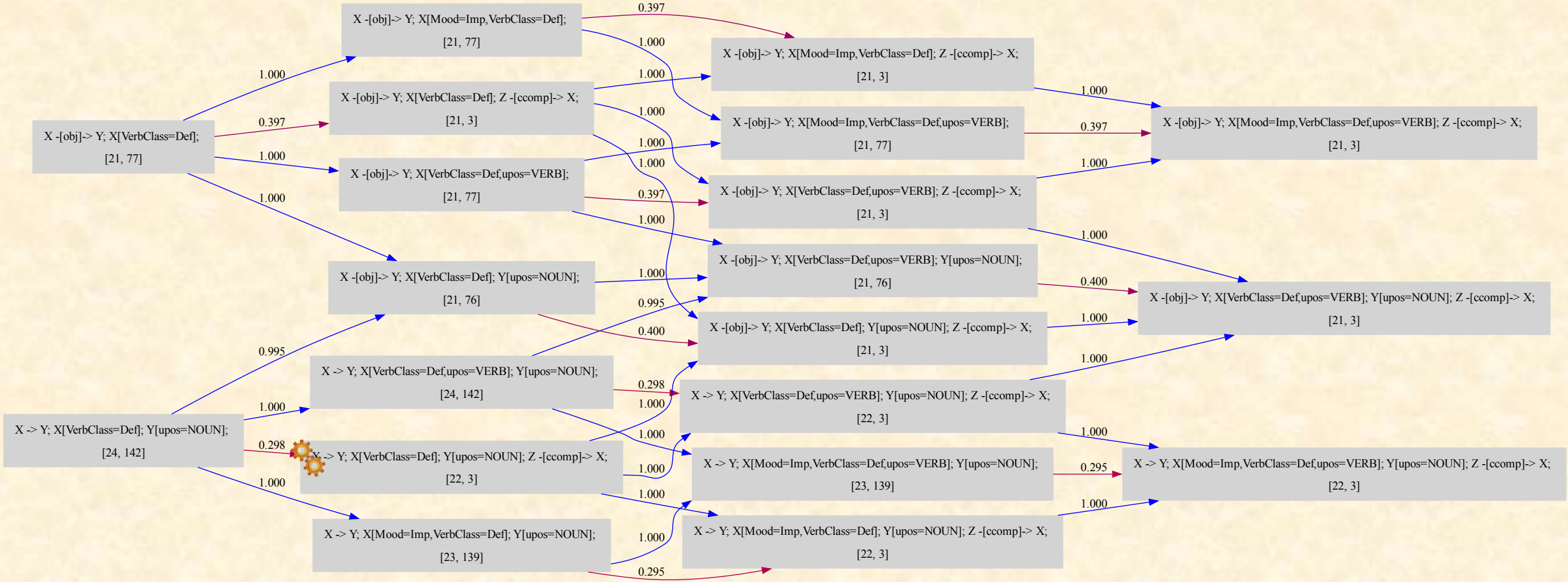




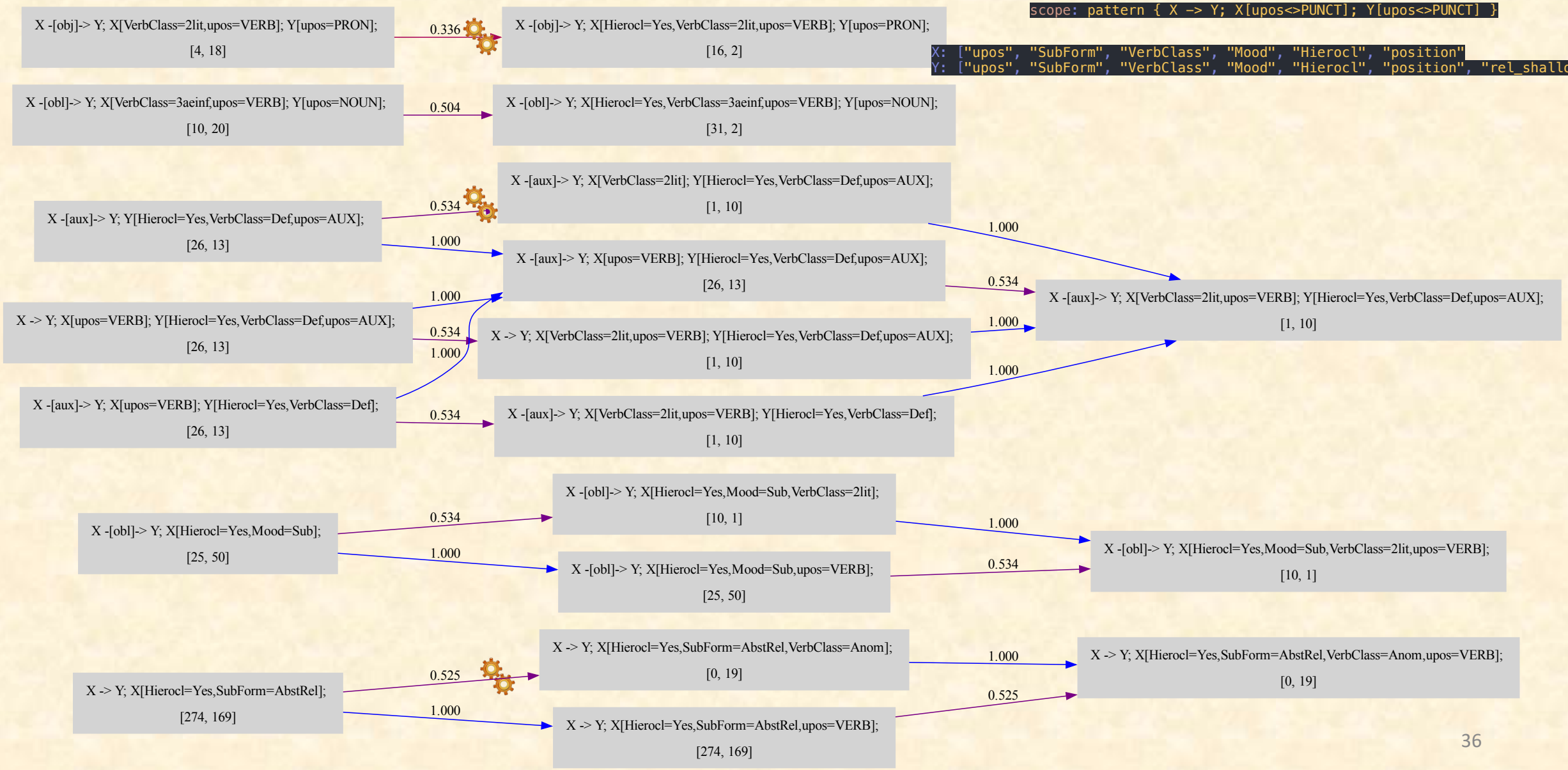
Grammar exploration in PT • Contrastive patterns

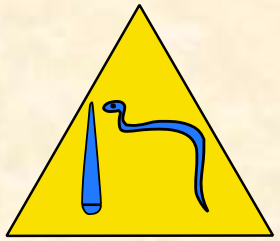
```
scope: pattern { X -> Y; X[upos] }
```

```
    X: ["upos", "VerbClass", "Mood", "Hierocl", "rel_shallow"]
    Y: ["upos", "VerbClass", "Mood", "Hierocl", "rel_shallow"]
```



Grammar exploration in PT • Contrastive patterns



.grew meets 

- What has been done?
  - **Grew-PT**
  - Automatic Grammar exploration: **Grex** + **Contrastive patterns**
  - PT in **ArboratorGrew** / **parser**
- What's next?
  - **more Kings** will be annotated
  - **Promote** the work for Egyptian philologists
  - **New features** in Grew-PT (parallel data)
- Challenges?
  - Annotate **all** Old Egyptian / Middle Egyptian corpus
  - Develop **new tools** for Egyptology

Thank you!

Roberto Antonio Díaz Hernández

Bruno Guillaume