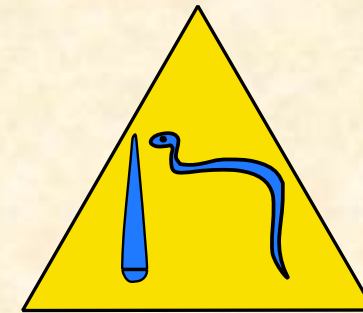


Digitalising the Pyramid Texts

grew meets



LORIA, Nancy — 16 May 2025

Roberto Antonio Díaz Hernández

Bruno Guillaume



Universidad
de Jaén



Beginning

[Home](#) [UJA](#) [PARSEME EG](#) [International Workshop](#) [Texts](#) [Contact](#) 

Nile in Contact

Historical linguistic analysis of Egyptian historical-biographical inscriptions from the third and second millennium BC found in the area between the First and the Fifth Cataract



Introduction

Nile in Contact is a multidisciplinary project intended to shed light on the Egyptian cultural influence on Nubia in the third and second millennium BC by applying a historical-linguistic approach to the study of hieroglyphic inscriptions found between Aswan and Kurgus in order to understand the evolution of Egyptian in Nubia from the Old Kingdom to the New Kingdom. It is an innovative project not only because of its multidisciplinary

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



- ▶  magische Texte 
- ▶  Medizinische Texte 
- ▼  Nile in Contact 
 - ▼  Qubbet el-Hawa 
 - ▶  Qubbet el-Hawa 25 (Mekhu) 
 - ▼  Qubbet el-Hawa 26 (Sabni) 
 - ▶  Architrave 
 - ▶  False door 
 - ▶  Head doorjamb 
 - ▶  Northern doorjamb 
 - ▶  Pillars 
 - ▶  Southern doorjamb 
 - ▼  Tomb façade 
 - ▶  Tomb façade, inscription A 
 - ▶  Tomb façade, inscription B 
 - ▶  Tomb façade, inscription C 
 - ▶  Tomb façade, south side 
 - ▶  West wall of the interior chamber 

CA21167 - Universality, diversity and idiosyncrasy in language technology (UniDive)

📄 Downloads

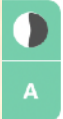
[Home](#) > [Browse Actions](#) > Universality, diversity and idiosyncrasy in language technology (UniDive)

Description

Management Committee

Main Contacts and Leadership

Working Groups and Membership



Description

Efficient access to the constantly growing quantities of data, especially of language data, largely relies on advances in data science. This domain includes natural language processing (NLP), which is currently booming, to the benefit of many end users. However, this optimization-based technological progress poses an important challenge: accounting for and fostering language diversity. The UniDive Action takes two original stands on this challenge. Firstly, it aims at embracing both inter- and intra-language diversity, i.e. a diversity understood both in terms of the differences among the existing languages and of the variety of linguistic phenomena exhibited within a language. Secondly, UniDive does not assume that linguistic diversity is to be protected against technological progress but strives for both of these aims jointly, to their mutual benefit. Its approach is to: (i) pursue NLP-applicable universality of terminologies and methodologies, (ii) quantify inter- and intra-linguistic diversity, (iii) boost and coordinate universality- and diversity-driven development of language resources and tools. UniDive builds upon previous experience of European networks and projects which provided a proof of concept for language modelling and processing, unified across many languages but preserving their diversity. The main benefits of the Action will include, on the theoretical side, a better understanding of language universals, and on the practical side, language resources and tools covering, in a unified framework, a bigger variety of language phenomena in a large number of languages, including low-resourced and endangered ones.

Action Details

- 📄 MoU - 081/22
- 🏛️ CSO Approval date - 27/05/2022
- 📅 Start date - 23/09/2022
- 📅 End date - 22/09/2026
- 🌐 <https://unidive.lisn.upsaclay.fr/>

How can I participate?

- Read the Action Description [MoU](#)
- Inform the Main Proposer/Chair of

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

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](#)
 - [Chişinău](#)
 - [Yerevan](#)
 - [call for trainers](#)
- [Shared tasks](#)
 - [AdMiRe Vienna](#)
 - [MSP Ljubljana](#)
 - [PARSEME 2.0](#)

UniDive COST Action CA21167



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

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

The screenshot shows the GitHub interface for the repository 'UD_Egyptian-UJaen'. At the top, the repository name is followed by a 'Public' badge. To the right are buttons for 'Edit Pins', 'Watch' (135), 'Fork' (0), and 'Star' (0). Below this is a navigation bar with 'dev' selected, a search bar 'Go to file', and a 'Code' button. A status bar indicates 'This branch is 5 commits behind master'. The main content area shows a commit by 'dan-zeman' titled 'Updated statistics.' with a commit hash '49e1331' and '2,686 Commits'. Below the commit is a table of files:

not-to-release	Value of VerbClass feature must be ...	last month
CONTRIBUTING.md	Initialization and the last commit to t...	last year
LICENSE.txt	Initialization and the last commit to t...	last year
README.md	Update README.md	7 months ago
egy_ujaen-ud-test.conllu	Removed extra empty line.	last month
egy_ujaen-ud-train.conllu	Introducing train-test split.	last month
stats.xml	Updated statistics.	2 days ago

On the right side, the 'About' section states 'No description, website, or topics provided.' and lists links for 'Readme', 'View license', 'Activity', and 'Custom properties'. It also shows '0 stars', '135 watching', and '0 forks'. Below this is the 'Releases' section with '3 tags' and a link to 'Create a new release'.

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

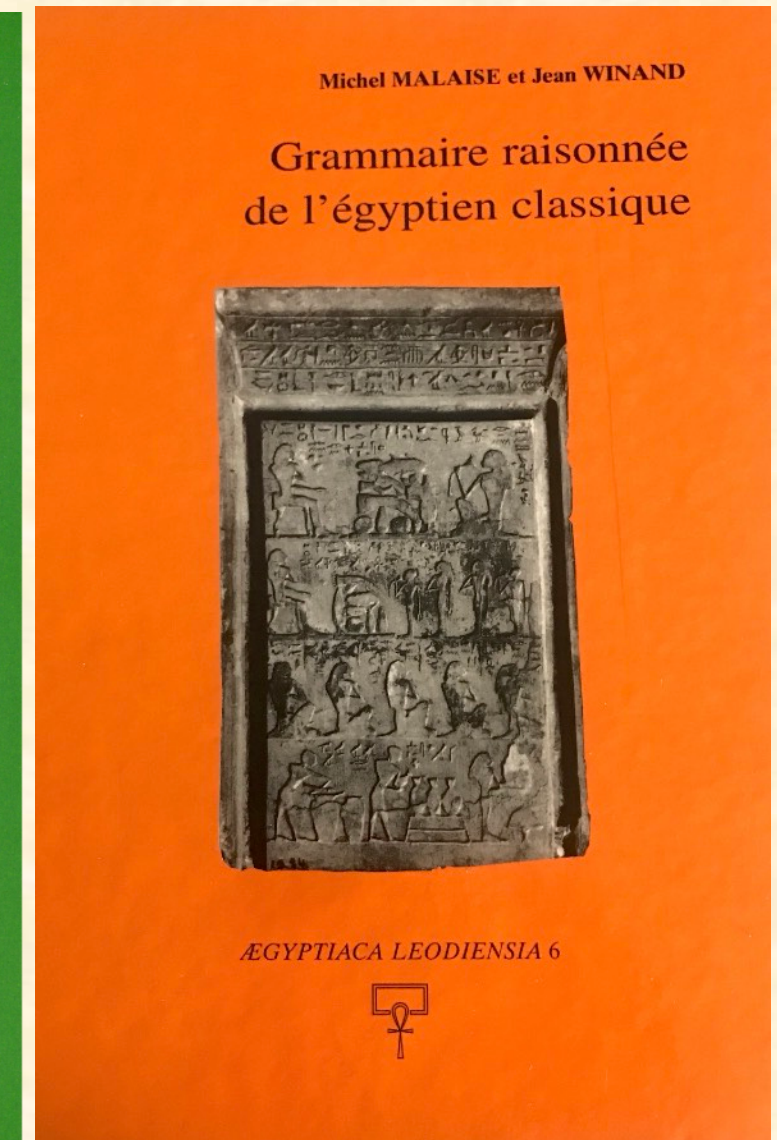
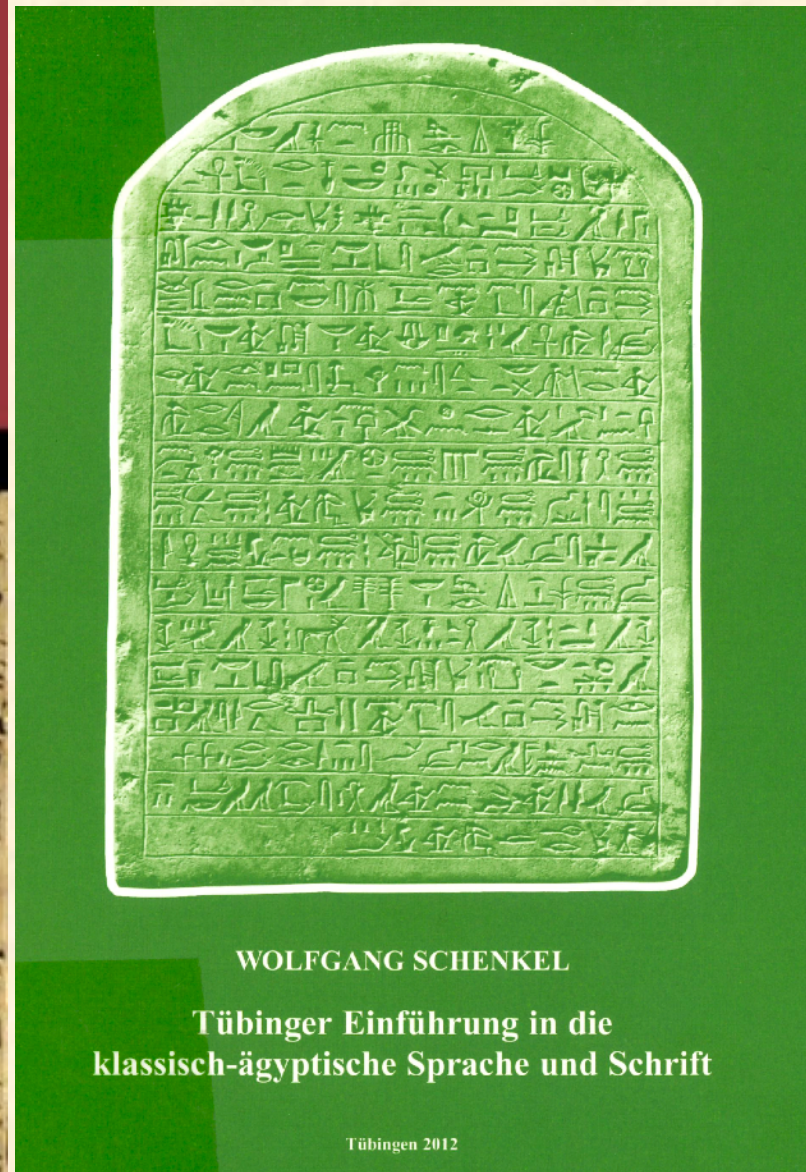
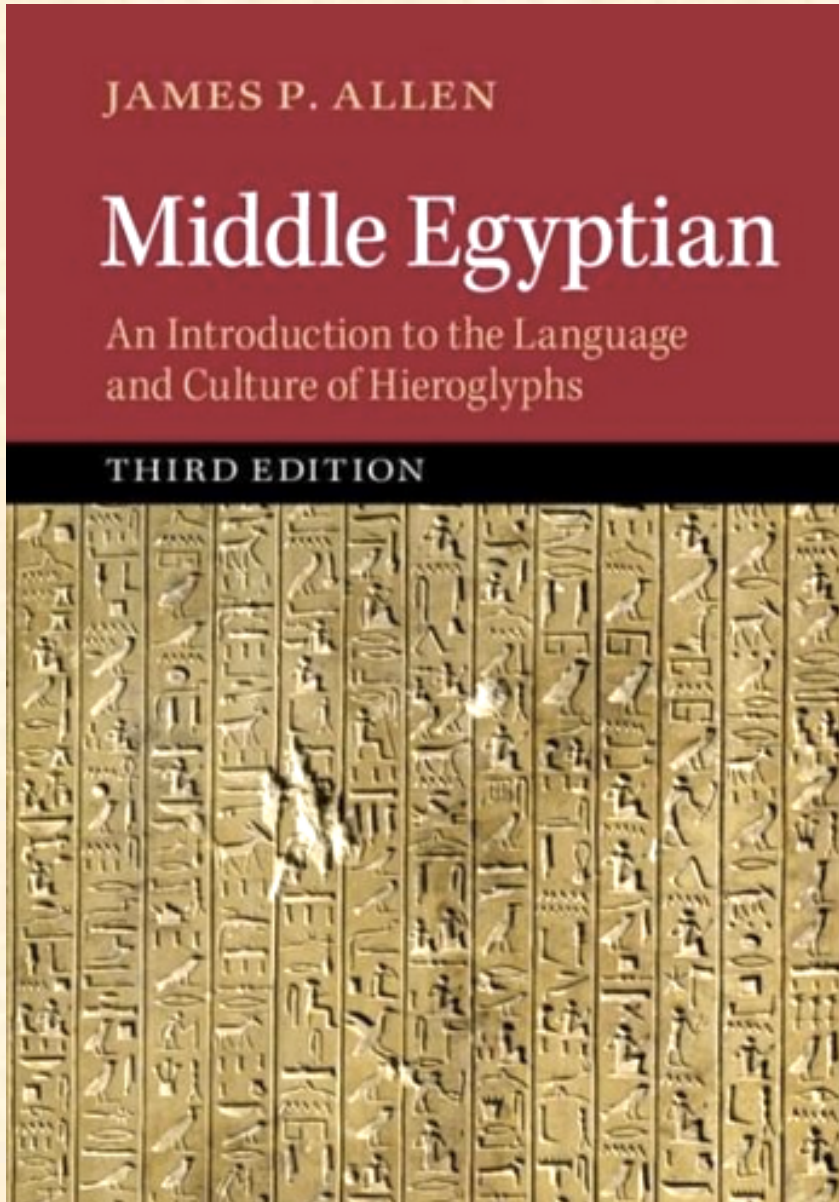
Administrative texts

Demotic

Literary texts

Administrative texts

No studies on dependency grammar in Egyptian



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

TLA Thesaurus Linguae Aegyptiae
2.2.1.1

Home / Text / Sentences

Sentences of Text
ZB4VUANVXJHAZESPZWSXSUEJUY

Share

Info

1 2 3 4

(762b)

(762c)

(31)

Glyphs artificially arranged

(762b) [P/F/W
sup 19 = 19]
mdwḯ =k d.t =k
šzp.n =k jr.w nṯr
(762c) ʕi =k jm ḥr
nṯr.PL ḥnt(.j).w ši

DE Du sollst selbst reden (?),
nachdem du die Gestalt
eines Gottes empfangen
hast, auf daß du dadurch
groß seist unter den Göttern,
die an der Spitze des Sees
sind.

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](#)[Cite](#)**LRLC P. BM 10416***r-nty*

PL. 15

1

IND_INIT

*tj*

VB

PFV

*n3y.k*

ART.POSS

*rmt.w*

SUBST

*m*

CONNECT

*n3y.w*

ART.POSS

*°3.w*

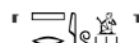
ADJ

*m*

CONNECT

à savoir bouger, remuer, enlever mes homme dans mes grand dans*n3y.w*

ART.POSS

*šrj.w*

ADJ

2

P. BM 10416*m*

CONNECT

*°h3wty.w*

SUBST

*m*

CONNECT

*hm.wt*

SUBST

*m*

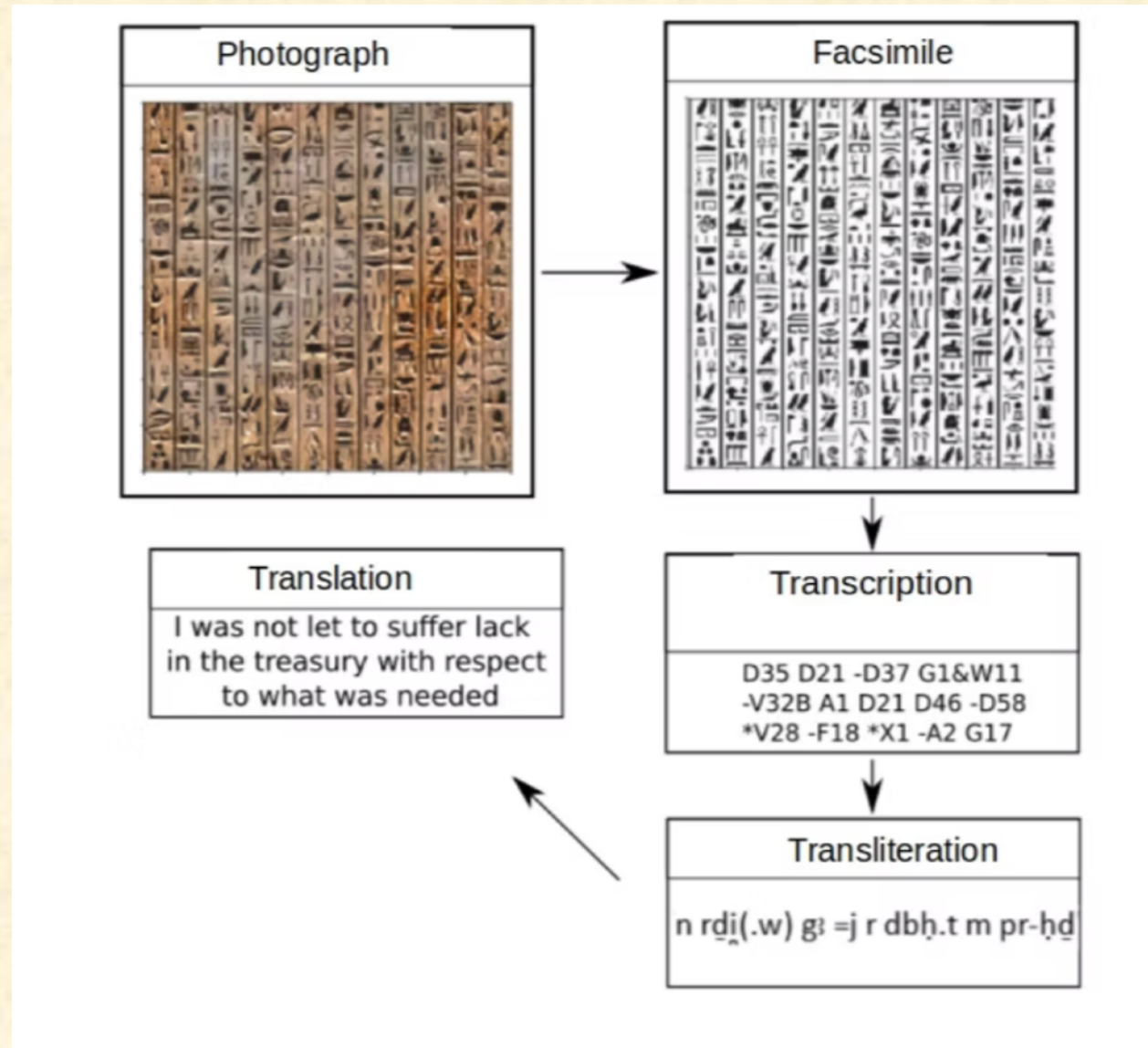
CONNECT

*h3wy*

SUBST

*mes**petit**dans**combattant, homme**dans**femme, épouse**dans**soir*

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



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.

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

Developing the Egyptian-UJaen Treebank

Roberto Antonio Díaz Hernández,¹ Marco Carlo Passarotti²

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

Morphosyntactic Annotation of the Pyramid Texts

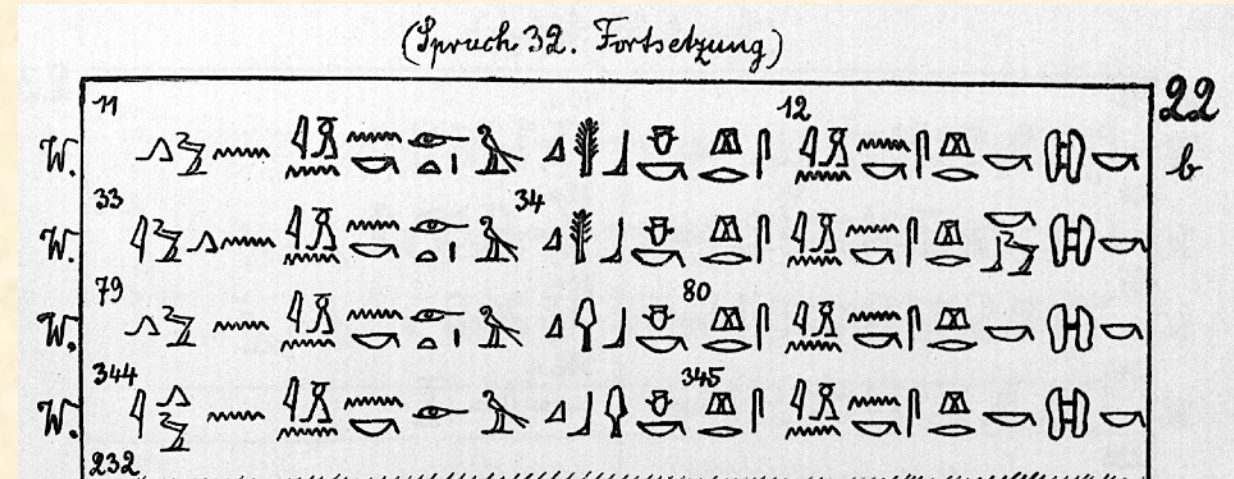
What are the Pyramid Texts?

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.



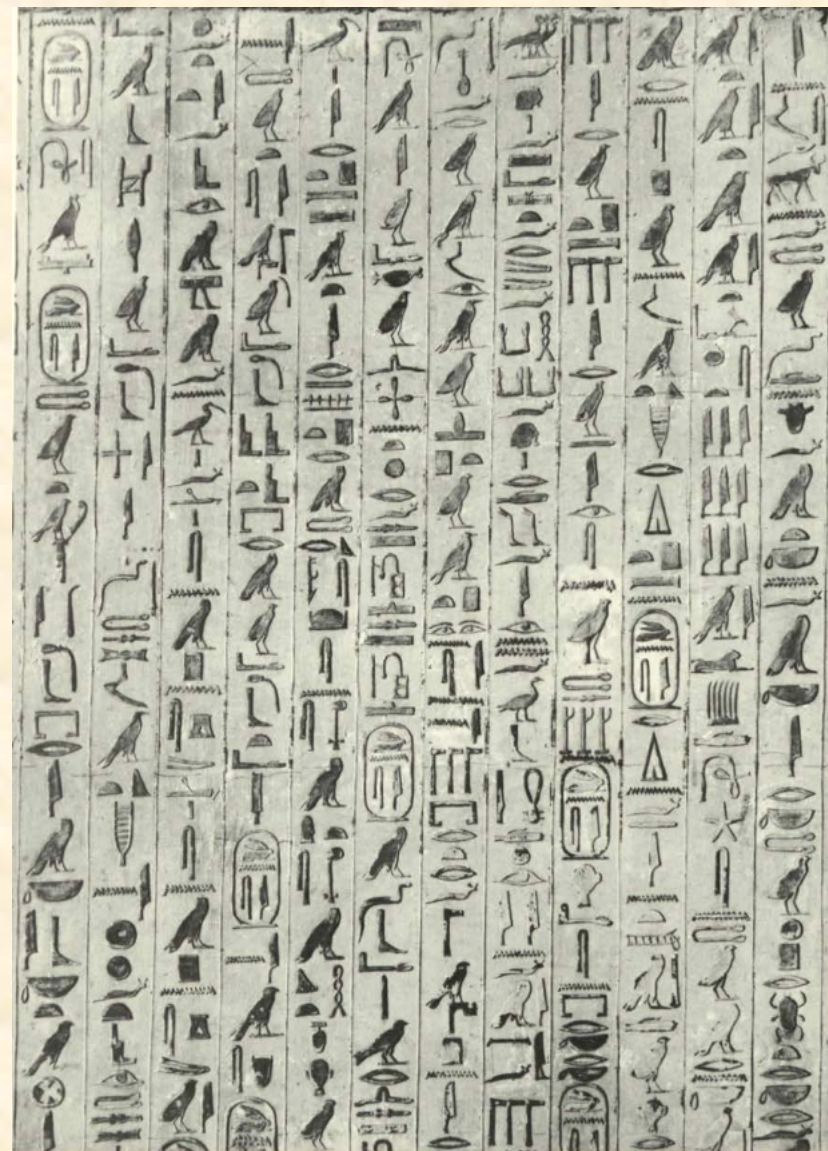
Unas's Pyramid Texts



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

(Spruch 224. Fortsetzung).

298 220
b
W. 145 146
T. 199 541 542
M. 541 542
N. 541 542

298 299
W. 146 147
T. 199 542 543
M. 542 543
N. 542 543

Schlusswort zum Spruche 224,
bei T. auch als Einleitung verwendet.

299 221
a
W. 137 147 148
T. 199 544 545
M. 544 545
N. 544 545

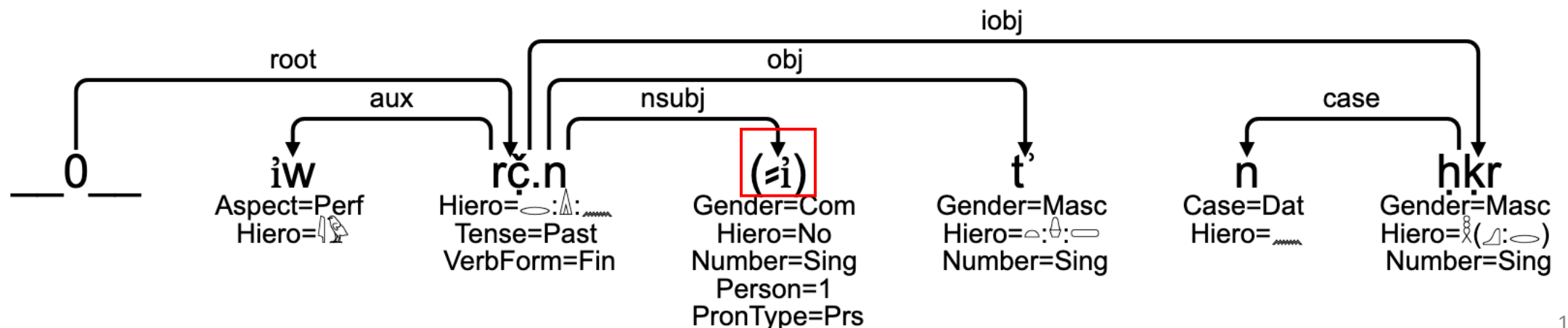
Description of the columns

```

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 = iw rç.n (≠i) t' n ḥkr
4 1 iw iw AUX Particle Aspect=Perf 2 aux _ Hiero=ⲓ
5 2 rç.n rçἰ VERB SPC=Past-2|Type=Pred Tense=Past|VerbFprm=Fin 0 root _ Hiero=ⲓ:ⲁ:ⲙ
6 3 (≠i) i 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=ⲓ:ⲁ:ⲙ
8 5 n n ADP Status=Cons Case=Dat 6 case _ Hiero=ⲓ
9 6 ḥkr ḥkr NOUN _ Gender=Masc|Number=Sing 2 iobj _ Hiero=ⲓ(ⲁ:ⲙ)

```

trans = (I) have given bread to the hungry



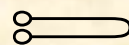
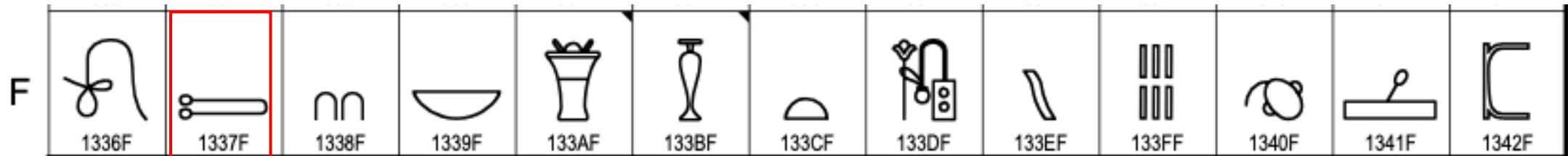
Unicode Font for Hieroglyphs

1085 8 s₃Č s₃Č VERB _ Gender=Masc|VerbForm=Inf 1 parataxis _ Hiero=(:≡)UC_14386

13360

Egyptian Hieroglyphs

1342F

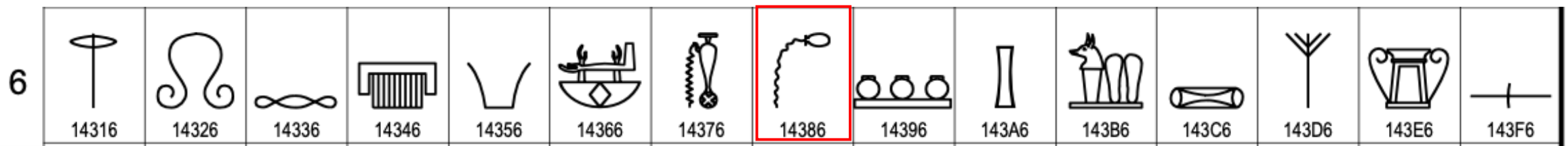


Hiero=(:≡)UC_14386

14310

Egyptian Hieroglyphs Extended-A

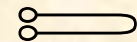
143FF



Hiero=(:≡)UC_14386

Transcription System

	LUT	Tübingen	Unicode
	<i>ʒ</i>	<i>ʒ</i>	A723
	<i>l</i>	<i>l</i>	A7BD
	<i>y</i>	<i>y</i>	
	<i>ī</i>	<i>ī</i>	00EF
	<i>‘</i>	<i>‘</i>	A725
	<i>w</i>	<i>w</i>	
	<i>b</i>	<i>b</i>	
	<i>p</i>	<i>p</i>	
	<i>f</i>	<i>f</i>	
	<i>m</i>	<i>m</i>	
	<i>n</i>	<i>n</i>	
	<i>r</i>	<i>r</i>	
	<i>h</i>	<i>h</i>	
	<i>ḥ</i>	<i>ḥ</i>	1E25
	<i>ḥ</i>	<i>ḥ</i>	1E2B
	<i>ḥ</i>	<i>ḥ</i>	1E96
	<i>z</i>	<i>s</i>	
	<i>s</i>	<i>ś</i>	015B
	<i>š</i>	<i>š</i>	0161
	<i>q</i>	<i>ḳ</i>	1E33
	<i>k</i>	<i>k</i>	
	<i>g</i>	<i>g</i>	
	<i>t</i>	<i>t</i>	
	<i>ṭ</i>	<i>č</i>	010D
	<i>d</i>	<i>ṭ</i>	1E6D
	<i>ḏ</i>	<i>č</i>	010D+0323



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

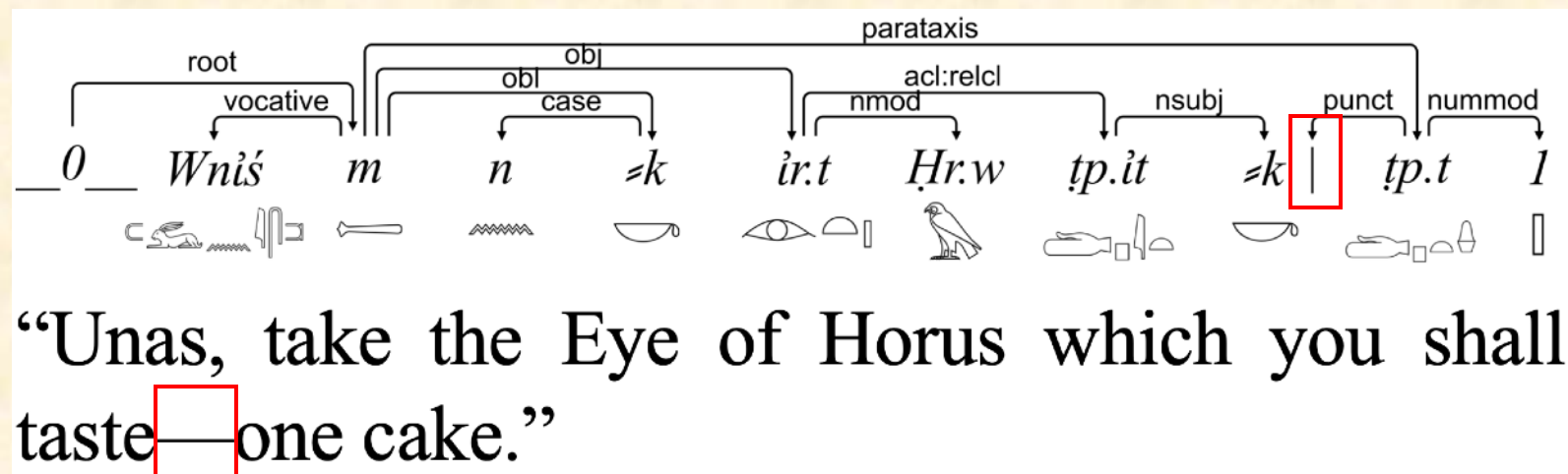


= [tʰ] = **ṭ** alveolar and dental ejective stops like in Ge'ez
 ≠ **d**

Recitation text

Line

Ritual remark



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

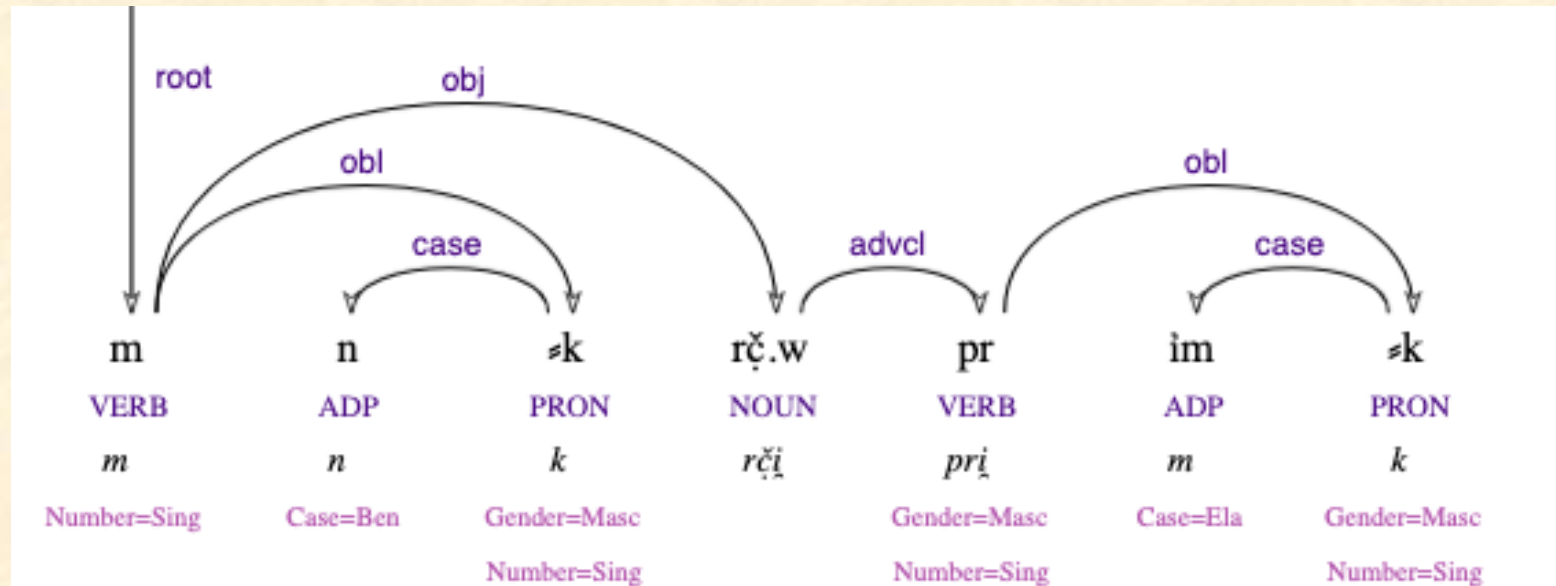
2180

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 rç.w pr im k
# timestamp = 0
# user_id = parser
1  m  m  VERB  _  Mood=ImplNumber=SingVerbClass=Def  0  root  _  _
2  n  n  ADP  _  Case=Ben  3  case  _  _
3  k  k  PRON  _  Gender=MascNumber=SingPerson=2PronType=Emp  1  obl  _  _
4  rç.w  rçṯ NOUN  _  VerbClass=3aeinflVerbForm=Inf  1  obj  _  _
5  pr  prṯ VERB  _  Gender=MascNumber=SingTense=PastVerbClass=3aeinflVerbForm=PartVoice=Act  4  advcl  _  _
6  im  m  ADP  _  Case=Ela  7  case  _  _
7  k  k  PRON  _  Gender=MascNumber=SingPerson=2PronType=Prs  5  obl  _  _
```

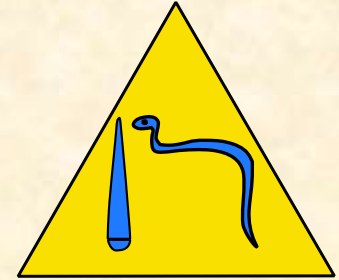
Improving the data encoding

```
# 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 Hr.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 PROPON _Gender=Masc 3 vocative _Hierogl=𓂠|Name=God
2 Wniś Wniś PROPON _Gender=Masc 1 appos _Hierogl=𓂠:𓂠|Name=King
3 m m VERB _Mood=Imp|Number=Sing|VerbClass=Def 0 root _Hierogl=𓂠
4 n n ADP _Case=Ben 5 case _Hierogl=𓂠|Status=Pron
5 =k k PRON _Gender=Masc|Number=Sing|Person=2|PronType=Emp 3 obl _Hierogl=𓂠|Pron=SFP
6 św.t(i)t św.tit NOUN _Gender=Fem|Number=Sing 3 obj _Hierogl=𓂠:𓂠
7 ir.t ir.t NOUN _Gender=Fem|Number=Sing 6 nmod:poss _Hierogl=𓂠(:𓂠)
8 Hr.w Hr.w PROPON _Gender=Masc 7 nmod:poss _Hierogl=𓂠(𓂠:𓂠)|Name=God
9 | | PUNCT _ 10 punct _ |LINE
10 św.t św.t NOUN _Gender=Fem|Number=Sing 3 parataxis _Hierogl=𓂠:𓂠|Hierogl=Yes
11 ḥnw.t ḥnw.t NOUN _Gender=Fem|Number=Sing 10 appos _Hierogl=𓂠
12 1-1 NUM _ 11 nummod _NumType=Card|Hierogl=𓂠
```

```
# 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 Hr.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
# ref = Pyramid Texts § 96/64d W => 81a W, Saqqara, 5th Dynasty, rel, OE
1 Wśr(.w) Wśr.w PROPON _Gender=Masc 3 vocative _Hierogl=𓂠|Name=God
2 Wniś Wniś PROPON _Gender=Masc 1 appos _Hierogl=𓂠:𓂠|Name=King
3 m m VERB _Mood=Imp|Number=Sing|VerbClass=Def 0 root _Hierogl=𓂠
4 n n ADP _Case=Ben 5 case _Hierogl=𓂠|Status=Pron
5 =k k PRON _Gender=Masc|Number=Sing|Person=2|PronType=Emp 3 obl _Hierogl=𓂠|Pron=SFP
6 św.t(i)t św.tit NOUN _Gender=Fem|Number=Sing 3 obj _Hierogl=𓂠:𓂠
7 ir.t ir.t NOUN _Gender=Fem|Number=Sing 6 nmod:poss _Hierogl=𓂠(𓂠:𓂠)
8 Hr.w Hr.w PROPON _Gender=Masc 7 nmod:poss _Hierogl=𓂠(𓂠:𓂠)|Name=God
9 | | PUNCT _ 10 punct _ |LINE=Punct
10 św.t św.t NOUN _Gender=Fem|Number=Sing 3 parataxis _Hierogl=𓂠:𓂠|Hierogl=Yes
11 ḥnw.t ḥnw.t NOUN _Gender=Fem|Number=Sing 10 appos _Hierogl=𓂠
12 1-1 NUM _NumType=Card 11 nummod _Hierogl=𓂠
```

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:

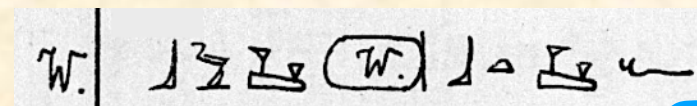
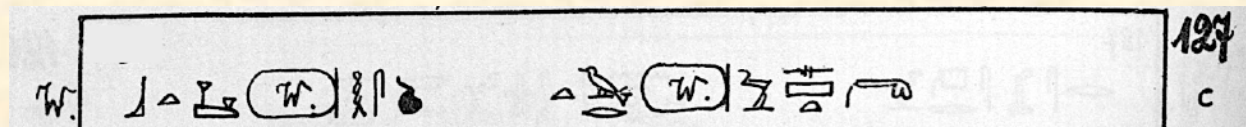
- ▶ Simultaneous display of parallel sentences

Keeping parallel data

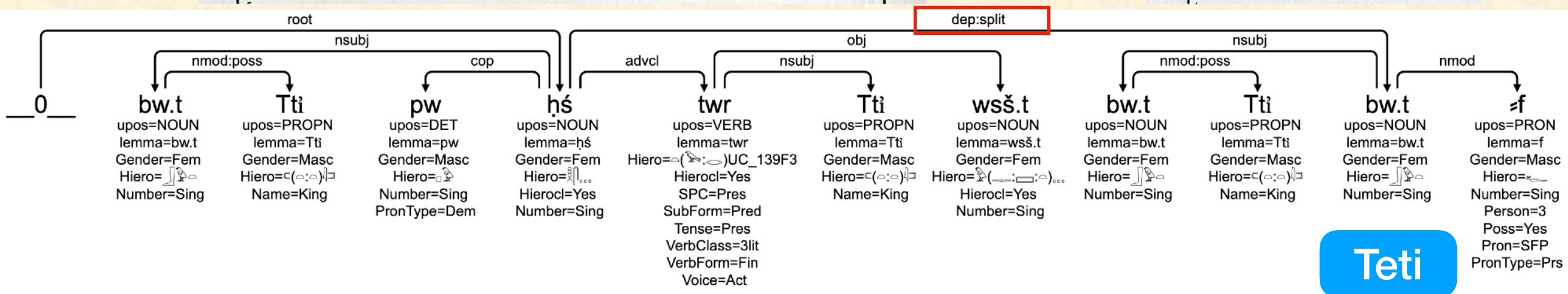
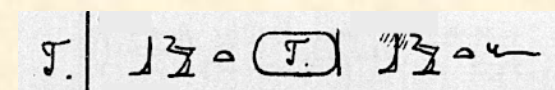
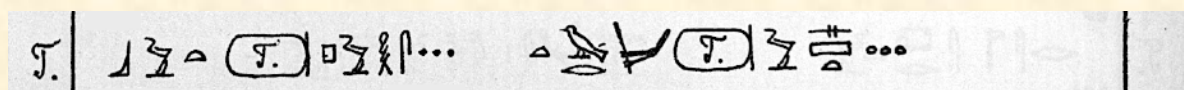
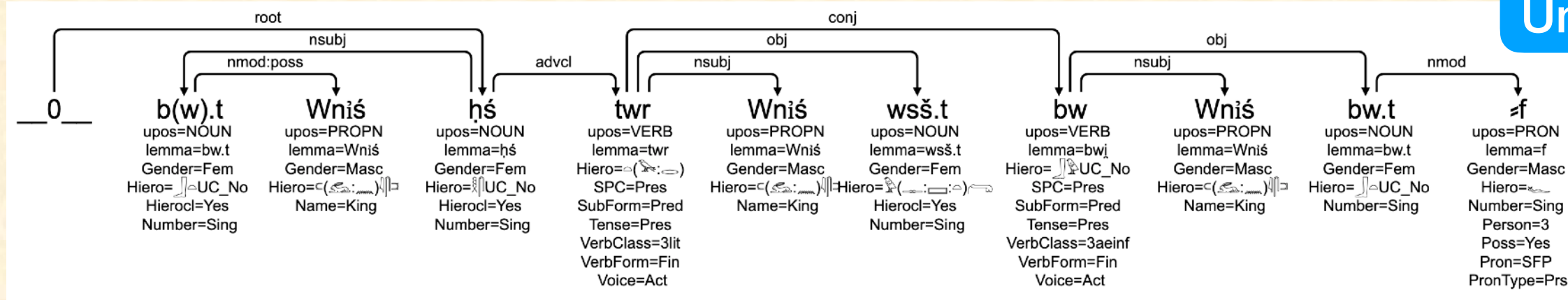
(Spruch 224. Fortsetzung).

	298		220
W.			b
J.	145	146	
M.	199		
N.	541	542	

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 same
        print(f'{form}')
        for hiero, sub_dict in hiero_dict.items():
            for king, size in sub_dict.items():
                print(f'    {king} == {size} ==> {hiero}')
```

Kbh

Teti ==1==> 
Teti ==1==> 

Hr.y

Unas ==1==> 
Unas ==1==> 

Hpr

Unas ==1==> 
Teti ==1==> 

Hnt(.i)

Teti ==1==> 
Teti ==1==> 
Teti ==1==> 
Pepi ==1==> 
Teti ==3==> 
Unas ==2==> 

H3m.w

Teti ==2==> 
Unas ==1==> UC_No

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] }

S occurrences: 2592

pattern Q : king=Teti

Q occurrences: 1103

pattern	n_pattern_occurrences	n_pattern_positive_occurrences	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	PT_Unas	44.59%	37.54%	14.57%
	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

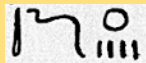
cos = 0.475



* -[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:

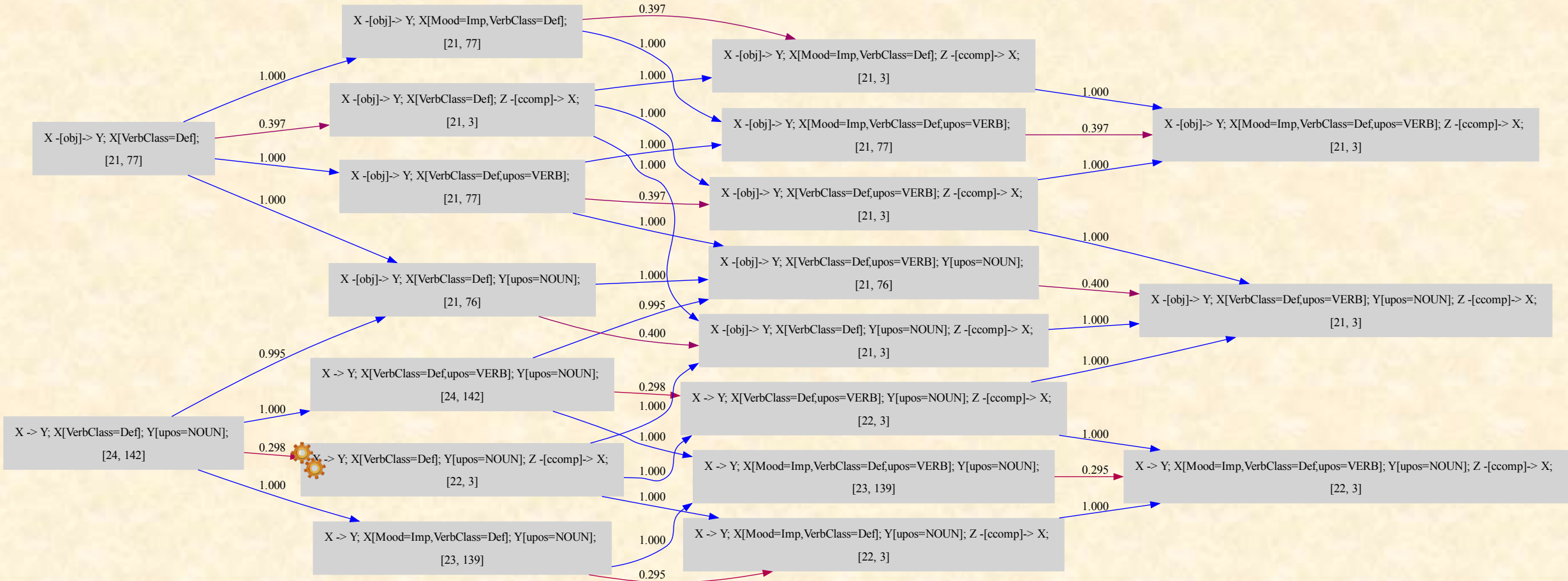
- ▶ one metadata error, it should be Teti
- ▶ one annotation to revise



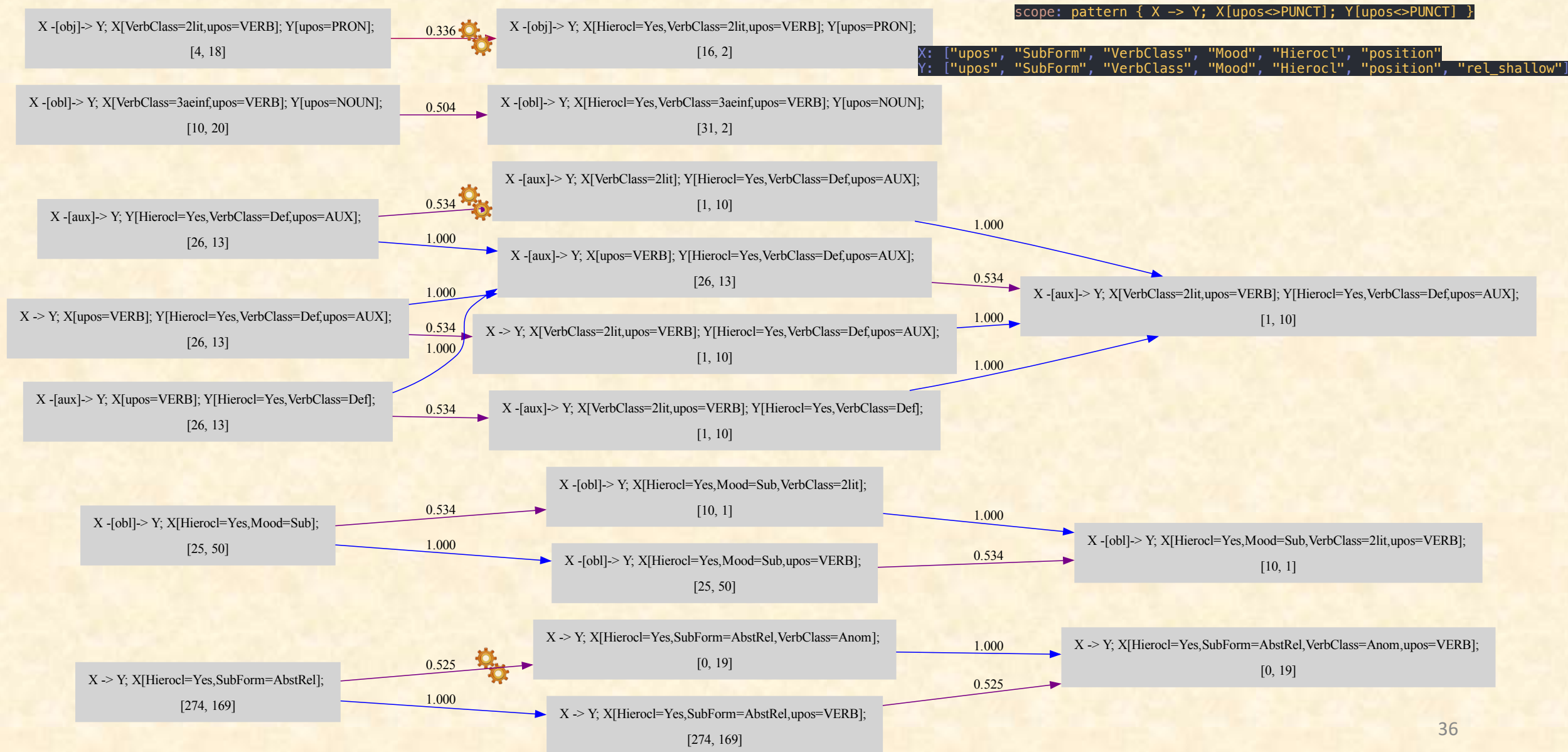
Grammar exploration in PT • Contrastive patterns

```
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

