

## The annotation of the Pyramid Texts in the Egyptian-UJaen treebank

Roberto Antonio Díaz Hernández, University of Jaén (Spain)

**Keywords:** Morphosyntactic annotation, Pyramid Texts, Old Egyptian

**Project phase:** Begun in March 2024.

### Introduction

Thanks to UniDive, I am acquainted with computational linguistics and NLP concepts. My previous work focused on the analysis and manual annotation of body part multiword expressions in Old Egyptian. It was a selective analysis of Old Egyptian multiword expressions found in Hannig's Old Egyptian dictionary. The present project is more ambitious, for it envisages the complete annotation of the Pyramid Texts in the Egyptian-UJaen treebank recently created in Universal Dependencies with the purpose of identifying all MWEs in such a corpus and to train an Old Egyptian parser. Attendance at the UniDive training school will be of great help to me towards a successful completion of this project.

### Description

The aim of this project is the morphosyntactic annotation of every sentence of the Pyramid Texts in the Egyptian-UJaen treebank of Universal Dependencies—the first morphosyntactic treebank created for Pre-Coptic Egyptian. The Pyramid Texts are written in Old Egyptian and they are the oldest corpus of Egyptian religious texts. They have been edited by Kurt Sethe (1908–22), *Die altägyptischen Pyramidentexte nach den Papierabdrücken und Photographien des Berliner Museums* (4 vols.), and by James P. Allen (2013), *A New Concordance of the Pyramid Texts* (6 vols.). Sethe's edition contains 2217 sections to which Allen's edition adds about 100 more. I started annotating the Pyramid Texts in the Egyptian-UJaen treebank at the end of March of 2024 and a month later I had annotated 130 sections containing 423 sentences and 3364 words. I estimate that the annotation of the Pyramid Texts will be completed in two years at a rate of 100 sections per month. The annotation of the Pyramid Texts in the Egyptian-UJaen treebank pursues the following results:

- 1) The creation of a detailed corpus of Old Egyptian multiword expressions to be published in PARSEME.
- 2) The training of a language parser in Old Egyptian.

The UniDive training school will give me the opportunity both to improve my skills in morphosyntactic annotation in Universal Dependencies and to learn how to use a CoNLL-U file in PARSEME. This will conveniently complement my research stay at the *Centro Interdisciplinare di Ricerche per la Computerizzazione di Segni dell'Espressione* in Milan from 1 May to 31 July, where Prof. Marco Passarotti will teach me how to train a language parser.

### Discussion of Open Questions

When annotating the first sections of the Pyramid Texts, I have come upon some issues that I would like to share with the UniDive trainers in order to get to know their opinion, for example:

1) Some sentences in Egyptian texts are ambiguous nowadays because the vowels were not written in the hieroglyphic script which makes the grammatical function of some words hard to understand. So far, I have dealt with such ambiguous sentences by registering to different annotations, for example:

UDE-214 = Pyramid Texts § 49+6 Nt

<i>Ḥw.ti</i>	<i>in</i>	<i>św</i>
Thoth-GN	bring:IMP	=3SG.M

“Thoth, bring him.”

UDE-214a = Pyramid Texts § 49+6 Nt

<i>Ḥw.ti</i>	<i>in</i>	<i>św</i>
Thoth-GN	bring:PTCP (M.SG)	=3SG.M

“It is Thoth who brought him.”

2) Egyptian has idiosyncratic constructions that are very different from modern European languages, such as the use of nisba adjectives derived from prepositions. Such words can be used as adjective modifiers linking two nouns (A - B) or as nouns governing a second noun. In the first case, I explain the nisba adjective as an adjectival modifier of noun A to which noun B is linked as a nominal modifier, for example the nisba *n.t* links two nouns in the so-called “indirect genitive”:

UDE-77 = Pyramid Texts § 13d N

<i>r'</i>	<i>n(.t)</i>	<i>Ppy</i>
mouth-M.SG	of:NISBA(M.SG)	Pepi-KN

“The mouth of (lit.: belonging to) Pepi.”

As a noun cannot modify an adjective, the name Pepi is not a noun modifier of the nisba *n.t*, but of *r'* “mouth”. The same happens with other nisbas. In this example, the nisba *im.(i)t* must be the adjective modifier of *ir.t* “eye” and the place name Dep its noun modifier:

UDE-240 = Pyramid Texts § 56b W

<i>ir.t</i>	<i>Hr.w</i>	<i>im.(i)t</i>	<i>Ṭp</i>	(...)
eye-F.SG	Horus-GN	in:NISBA(M.SG)	Dep-PN	

“The eye of Horus which is in Dep (...).”

However, a noun is linked as a nominal modifier to the nisba adjective when used as a noun, for example:

UDE-382 = Pyramid Texts § 120a–b W

<i>ḥr.(i)w</i>	<i>śt</i>
who are over- NISBA(M.PL)	baked food-M.SG

“Those who are over the baked food (i.e. those who are in charge of the baked foods).”

3) Egyptian texts sometimes contain errors which I enclose between curly brackets {} and I explain as typographical errors (Typo) in an unspecified dependency (dep), for example:

UDE-240 = Pyramid Texts § 56b W

{ <i>ir.t</i> }	<r>ḥ.n	<i>n</i>	<i>≠k</i>	<i>Hr.w</i>	<i>ir.t</i>	<i>≠f</i>
{eye-F.SG}	bring:PST	for:PREP	=2SG.M	Horus-GN	eye-F.SG	=3SG.M

“{Eye} Horus gave his eye to you.”

### Final statement

I am very interested in attending the UniDive training school and would be delighted if my application were accepted, for it would allow me to look deeper into learning concepts and tools of computational linguistics in order to develop new digital resources for Egyptian.