

Universal Dependencies for Hittite

7th Pavia Indo-European Summer School
8 September 2026

The project aims to develop a Universal Dependencies (UD) treebank for Hittite, the most ancient Indo-European language attested in writing (cuneiform, II millennium BC). As a case study, the letter of the Hittite Queen Puduheba (CTH 176, KUB 21.38) was parsed and annotated for UD, containing 1,254 words (193 clauses). The previous research (Inglese et al. 2018) yielded another 313 clauses pre-annotated in the dependency framework. The development of Hittite UD treebanks naturally meets certain challenges that we discuss below, such as annotation of clitic chains, complex sentences, null auxiliaries. This work builds on previous projects [1], [2], [3], [4].

Authors

Maria Molina (1)*,
Emma Yavasan (2)
mashenke@tauex.tau.ac.il
emma.yavasan@msmail.ariel.ac.il

Affiliations

1 – Tel Aviv University, Department of Ancient Near East
Cultures and Archaeology
2 – Ariel University, Digital Pasts Lab

*corresponding author

Introduction

Universal Dependencies (US) is a framework for consistent annotation of grammar across different human languages. Hittite, however, still lacks a proper UD annotation. For now, there is only one treebank of Hittite built upon some data retrieval from the classical grammar of Hittite [5]. This project is an on-going work for developing a consistent treebank of the Hittite letters and instructions.

As a first approach, the Letter of the Queen Puduheba to the Egyptian pharaoh Ramses II has been chosen – as one that is long enough and better preserved than most of the other letters and instructions from the Hittite Kingdom.



The King
Hattusili III
and his wife
Puduheba

Photo:
Krähenstein
CC BY-SA
3.015

Challenges

There are several challenges in the annotation of the Hittite UD. To mention a few:

- Hittite **correlatives**: a part of the complex sentence or an independent clause?
- Hittite **clitic chains**: independent wordforms or one part together with the host?
- Hittite **lemmatization**: why on Earth did they write in three languages, Hittite, Akkadian and Sumerian?

Methodology

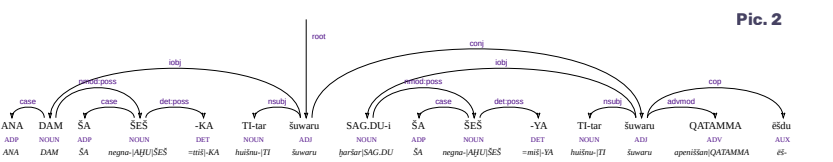
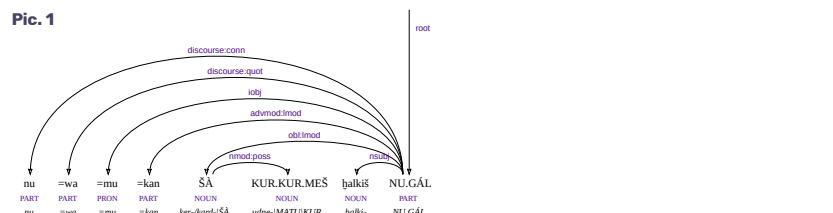
The corpus of texts includes Hittite letters, as published by [6] and represented in [2]. Files are being annotated in CONLL-U format. Visualization of the treebank: ArboratorGrew ([7], [8]).

Annotated files and screenshots of generated trees can be downloaded from GitHub:
https://github.com/mashenkeisraeli/hittite_ud

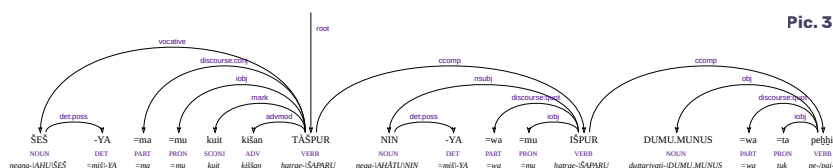
Examples

- nu=wa=mu=kan ŠA KUR.KUR.MEŠ [ha]kiš NU.GÁL* “In my lands I do not even have barley” (Pic. 1.)
- [ANA] DAM ŠA ŠEŠ-KA TI-tar šuwaru SAG.DU-i ŠA [ŠEŠ-YA TI-tar šuwaru QATAMMA ešdu “The wife of your brother (i.e., Puduheba, the wife of Hattusili) enjoys full life. May the person of my brother likewise enjoy full life!” (reconstruction by Hoffner 2009:282) (Pic. 2.)
- ŠEŠ-YA=ma=mu kuit kiššan TAŠPUR, NIN -YA=wa=mu IŠPUR DUMU.MUNUS=wa=ta pehhi “My brother, as you wrote to me in this manner, my sister wrote to me: ‘I will give my daughter to you.’” (Pic. 3.)
- ANA ŠEŠ-YA ammel aššulan ammel unuwašhan uppaḥḥun “I have sent my greetings (and) my ornaments to my brother”
- ammuk=ma ANA ŠEŠ-YA kuin DUMU.MUNUS nepiāš KI-a-āš=a pehhi n=an =kan kuedani ḥandami “The daughter of heaven and earth whom I will give to my brother—to whom should I compare her?” (Pic. 4.)
- nu=wa* NAM.RA.MEŠ kuin GU, MEŠ UDU.HI.A *pé ḥarkanzi ... nu=war=an=kan arḫa daškandu ... nu=war=an* parnawiškandu “The captives, cattle, and sheep that they are holding – let them be taken away and let them be housed.” (Pic. 5.)

Clitic chains are separate syntactic words



Reported discourse: ccomp or parataxis

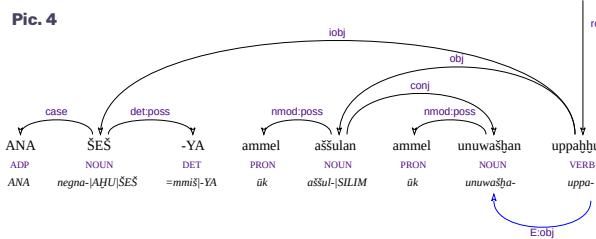


Three languages in one LEMMA

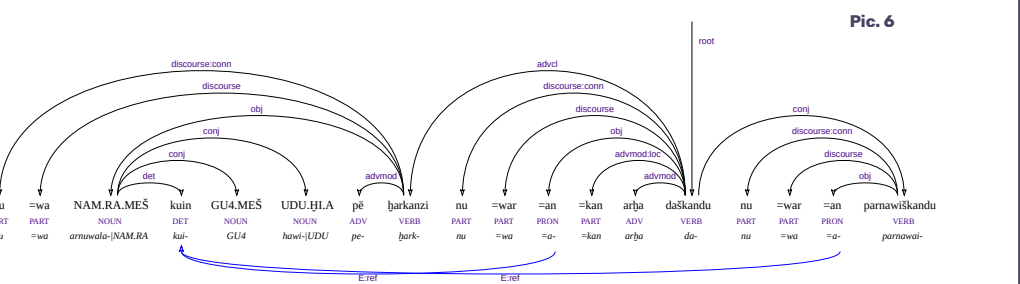
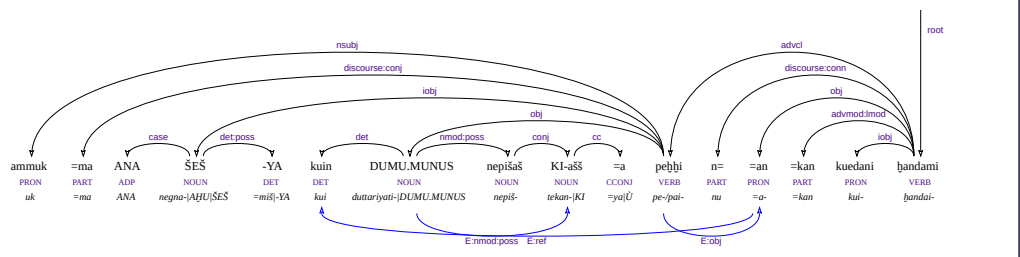
A Sumerogram or Akkadogram spells a Hittite word. The project records all three and puts the Hittite lexeme first, separated by vertical bars, with the written realisation marked in FEATS (Pic. 2)

LEMMA = mahḥan|GIM ‘when’
LEMMA = negna-|AHU|ŠEŠ ‘brother’
FEATS = Language=Sum / Language=Akk

Coordination and Akkadographic prepositions



Correlatives in the enhanced graph



The seal of Hattusili III
and Queen Puduheba

Conclusion

For the time being the corpus of Hittite clauses consists of 105 texts, 2,197 clauses, which include approximately 10,000 words. The work is ongoing, with a promising perspective to eventually create the first Hittite treebank based on the language of real documents.

The results of the UD annotation for the Hittite letters have allowed for the development of new guidelines for Hittite annotation without requiring many changes to the basic UD guidelines. This will facilitate the easier inclusion of Hittite into the Universal Dependencies system and in general will open Hittite texts to automated typological comparison.

References

- G.Inglese. 2015. Towards a Hittite Treebank. Basic Challenges and Methodological Remarks. In: *Proceedings of the Workshop on Corpus-Based Research in the Humanities (CRH)*.
- M.Molina. 2015-. Annotated Corpus of Hittite Clauses. <http://hittitecorpus.com>
- G.Inglese, M.Molina, H.Eckhoff. 2018. Incorporating Hittite into PROIEL: a pilot project. In: *Proceedings of the Second Workshop on Corpus-Based Research in the Humanities (CRH-2)*.
- E.Andersen, B.Rozonoyer. 2020. A Small Universal Dependencies Treebank for Hittite. In: *Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020)*.
- H.Hoffner, H.C.Melchert. 2024. A Grammar of the Hittite Language. Part 1. Reference Grammar. Eisenbrauns, Winona Lake.
- H.Hoffner. 2009. Letters from the Hittite Kingdom. Atlanta: Society of Biblical Literature.
- ArboratorGrew. 2020-. Accessed August 31, 2026. <https://arborator.grew.fr/>.
- Guibon, Gaël, Marine Courtin, Kim Gerdes, and Bruno Guillaume. 2020. “When collaborative treebank curation meets graph grammars.” *Proceedings of the Twelfth Language Resources and Evaluation Conference: 5291–5300, Marseille, France. European Language Resources Association.*

