Fragments and Formula

An annotated corpus for the Phrygian language

Oggi Peeters

Presented in fulfilment of the requirements for the degree of
Master of Arts in Linguistics

Supervisor: dr. Alek Keersmaekers
Co-supervisor: prof. dr. Toon Van Hal

Academic year 2023-2024

76.755 characters

KU LEUVEN
FACULTY OF ARTS
BLIJDE INKOMSTSTRAAT 21 BOX 3301
3000 LEUVEN, BELGIË

I hereby declare that, in line with the Faculty of Arts' code of conduct for research integrity, the work submitted here is my own original work and that any additional sources of information have been duly cited.

KU LEUVEN
FACULTY OF ARTS
BLIJDE INKOMSTSTRAAT 21 BOX 3301
3000 LEUVEN, BELGIË

Samenvatting

Deze thesis stelt een geannoteerd corpus van het Nieuwfrygisch voor. Het eerste hoofdstuk biedt een inleiding tot de taal zelf door de aandacht te vestigen op de externe geschiedenis en genetische status, gevolgd door een grammaticale basisschets en een discussie over de fragmentaire attestatie van de taal. Het tweede hoofdstuk bespreekt de prominentie van het Frygisch in de secundaire literatuur. Daarbij wordt een blik geworpen op de hoopvolle attitude van taalkundigen tegenover onderzoek van de taal in de 21ste eeuw, en op de nauwe verwantschap van het Frygisch met het Oudgrieks. In het derde hoofdstuk komt het annotatieproces aan bod en wordt beschreven hoe de data tot stand zijn gekomen in de FileMaker-omgeving van de KU Leuven. Als laatste wordt nagegaan in hoeverre de gegevens van het geannoteerde corpus beantwoorden aan de beschrijvingen van het Frygisch in de secundaire literatuur inzake basismorfologie, -syntaxis en het lexicon. De resultaten van het onderzoek suggereren dat de grammaticale beschrijvingen in de literatuur weliswaar over het algemeen correct zijn, maar dat er desalniettemin voldoende afwijkingen zijn die verder onderzoek verantwoorden.

KU LEUVEN
FACULTY OF ARTS
BLIJDE INKOMSTSTRAAT 21 BOX 3301
3000 LEUVEN, BELGIË

Contents

KU LEUVEN
FACULTY OF ARTS
BLIJDE INKOMSTSTRAAT 21 BOX 3301
3000 LEUVEN, BELGIË

List of abbreviations

Abbreviation	**Definition**
Acc.	Accusative case
AG	Ancient Greek
B	Provenance Bithynia
Dat.	Dative case
Fig.	Figure
Gen.	Genitive case
IE	Indo-European
M	Provenance Midas City
MPhr.	Middle Phrygian
Nom.	Nominative case
NPhr.	New Phrygian
OPhr.	Old Phrygian
Phr.	Phrygian
POS	Part(s) of speech
PIE	Proto-Indo-European
Tab.	Table
TM	Trismegistos (number)
UD	Universal Dependencies
W	Provenance West-Phrygia

KU LEUVEN
FACULTY OF ARTS
BLIJDE INKOMSTSTRAAT 21 BOX 3301
3000 LEUVEN, BELGIË

Tables and figures

KU LEUVEN
FACULTY OF ARTS
BLIJDE INKOMSTSTRAAT 21 BOX 3301
3000 LEUVEN, BELGIË

Preface

The purpose of this thesis is to demonstrate of what use an annotated corpus for Phrygian is. In particular, it explores how the data measure up against, as well as add to, existing research on the language. The structure is as follows. Firstly, secondary literature on Phrygian itself takes centre stage, ranging from external history and genetic status to basic grammar and its fragmentary attestation. Afterwards, the corpus is introduced and put to its first test.

I would like to thank my supervisors Alek Keersmaekers and Toon Van Hal. I also extend my gratitude to Mark Depauw for providing the Trismegistos identifiers, an invaluable addition to the data. It goes without saying that any and all mistakes and oversights are my own.

KU LEUVEN
FACULTY OF ARTS
BLIJDE INKOMSTSTRAAT 21 BOX 3301
3000 LEUVEN, BELGIË

1. Introduction to the Phrygian language

This chapter provides a general introduction to Phr. First, the external history of Phr. is discussed (§1.1), followed by a brief overview expounding it as a distinct IE language (§1.2).[1] Lastly, its characteristics as a 'fragmentary language' are particularly pertinent to the annotated corpus (§1.3).

1.1 Phrygian among IE languages

Phr. was an IE language spoken in ancient Anatolia (Fortson, 2010, pp. 460-461). The most probable scenario for its origins is that its speakers, the Phrygians, immigrated from the Balkan Peninsula to form their own branch of IE between 1200 and 800 BCE (Anthony, 2007, pp. 53, 57; Drews, 1993, p. 9; Fortson, 2010, p. 461),[2] whence they settled in the somewhat vaguely defined region named after them, Phrygia, and at least partly filled a power vacuum left by the Hittites (Diakonoff & Neroznak, 1985, p. xi). By the 9th century BCE, they had established a state on the peninsula, exercising great influence on neighbouring Anatolian and AG cultures through economic and political might (Mallory, 1989, p. 31; Mallory & Adams, 2006, p. 37; Roller, 1983, p. 301). Although their independence was lost to the Lydians, Persians, Greeks and Romans, their language remained intact until the 5th century CE and perhaps even until the Arab conquest in the 7th century (Brixhe, 1997, p. 177), though the attestations are not attested this late (cf. §1.3).

Regarding closer relations to other IE languages, the current consensus holds that AG has a close connection to Phr.[3] This Graeco-Phrygian relationship also carries an implicit yet close connection with Macedonian and Thracian, however not enough is known of either language to assert anything specific (Drews, 1993, pp. 9-10; Diakonoff & Neroznak, 1985, p. x are more confident in this opinion). Though in the minority, some researchers have emphasised close relations to Armenian (Diakonoff, 1984, pp. 109-112; Diakonoff & Neroznak, 1985, pp. xi, 41-42; Mallory, 1989, p. 34; Hämmig, 2017). Kuhrt (1998, p. 566) has even gone so far as to posit that "it is still not certain whether [Phr.] belongs to the Indo-European family". This position is untenable, as it is apparent that at an early stage Phr. was "dialectally close to" AG, Armenian, and Indo-Iranian "or even formed a dialectal group at the time of the Indo-European dispersal" (Martirosyan, 2014, pp. 4, 7). Apart from isoglosses, another important point

[1] For an overview of secondary literature on the matter, cf. §2.
[2] There are also classical accounts of the Phrygians' Balkan origins. Cf. Herodotus 7.73.2: οἱ δὲ Φρύγες, ὡς Μακεδόνες λέγουσι, ἐκαλέοντο Βρίγες χρόνον ὅσον Εὐρωπήιοι ἐόντες σύνοικοι ἦσαν Μακεδόσι, μεταβάντες δὲ ἐς τὴν Ἀσίην ἅμα τῇ χώρῃ καὶ τὸ οὔνομα μετέβαλον ἐς Φρύγας ("The Phrygians were, as the Macedonians claim, called Brigians back when they lived in Europe and were neighbours to the Macedonians. When they had migrated to Asia, they changed their name, as they did their homeland, to Phrygians."); Strabo 7.3.2: […] καὶ αὐτοὶ δ' οἱ Φρύγες Βρίγες εἰσί, Θρᾴκιόν τι ἔθνος […] ("The Phrygians are themselves Brigians, a Thracian tribe of some sort"). Such accounts, however, merely serve as antecedents to modern theories and should not form the basis of one's opinions on the matter.
[3] Brixhe (2008, p. 72) provides an overview of isoglosses that argue in favour of this assumption; Obrador-Cursach (2019) has an article dedicated to the matter.

for this classification is what is known in the scientific literature as 'Lautverschiebung' – the development of the PIE plosives in the above mentioned daughter languages (cf. §1.2). For the purpose of this thesis, AG is assumed to be Phr.'s closest relative.

1.2 The character of Phrygian[4]

It has previously been claimed that Phr. is a satem language (cf. Diakonoff & Neroznak, 1985). Current opinion holds that it is a centum language (Brixhe, 2008, p. 72; Ligorio & Lubotsky, 2018, p. 1824; Obrador-Cursach, 2020, p. 70), but Phr. might exemplify how the two-way distinction is not really useful. As Fortson (2010, p. 461) points out, there are conflicting outcomes for PIE **ǵʰ* as both /g/ in γλουρεος (W-11) 'golden' and /z/ in ζεμελως 'human (dat.pl.)'. The origin of /z/ is attributed to palatalisation secondary to the levelling of plain, labialised and palatal velar plosives in pre-Phr. (Ligorio & Lubotsky, 2018, p. 1824; Obrador-Cursach, 2020, p. 73). Perhaps a more apt phrasing is that Phr. underwent centum developments, some of which became less apparent due to later developments associated with satem languages. To me, Phr. is thus, in a sense, as much a satem language as French, which descends from the centum language Latin. For instance, the French *cent* 'hundred' is pronounced with an initial /s/, which goes back to Latin /k/ after palatalisation before front vowels (Allen, 1978, p. 14; Fortson, 2010, p. 288). Another major phonological development is *Lautverschiebung* (cf. §1.1). It specifically entails the merger of PIE voiced and voiceless plosives into the voiceless row and the transition from murmured plosives[5] into the voiced row.[6]

PIE	Voiceless	Voiced	Murmured/breathy
Phr.	Voiceless		Voiced

Tab. 1.1: The development of PIE plosives into Phr.

The shift is observed in words as *petes* 'feet' (< PIE **ped-*; cf. AG πόδες; Obrador-Cursach, 2020, p. 71) and βεκος 'bread' (< PIE **bʰhₓǵ-os-*; Lubotsky, 2004, p. 233). Phr. phonology otherwise preserves the PIE sound system quite conservatively (Brixhe, 2008, p. 72).

The consensus on Phr. morphology, regardless of how little is known, greatly resembles AG. With regard to nouns (cf. §4.1), there are four cases (nom., acc., gen. and dat.) and three genders (masculine, feminine and neuter); only two numbers are attested, singular and plural. There are multiple declensions including a thematic, also called o-stem, declension with the nom.sg. *-os*/-ος. Our understanding of the verbal system is limited due to the nature of the inscriptions. Major evidence for identifying these categories derives from their PIE heritage. Due to the nature

[4] A detailed grammatical account is not provided since clear and up-to-date grammatical overviews on Phr. already exist, such as Brixhe (1997, pp. 167-180; 2008, pp. 69-80), Ligorio & Lubotsky (2018) and Obrador-Cursach (2020, pp. 62-118); Fortson (2010, pp. 460-463) is good for a cursory overview, but inextensive.
[5] Traditionally, but inaccurately labelled the 'voiced aspirated' plosives (cf. Ladefoged & Maddieson, 1996, pp. 48 *Table 3.2*, 57-63).
[6] Lubotsky, 2004 provides an excellent discussion on this development.

of the inscriptions, only the third person is attested, along with only three 'tenses' (present, aorist and perfect). There was a binary voice distinction between active and mediopassive (e.g. resp. αδδακετ vs. αδδακετορ 's/he does'), and the indicative, subjunctive, optative and imperative moods are also attested. There is no evidence for a Phr. injunctive mood (cf. §4.2).

On a lexical level, Phr. shows a great number of apparent loanwords beside inherited vocabulary. Plenty of loans are of AG origin, although due to the close relation between the languages, it is often difficult to discern early borrowings from native Phr. words. Cf. M-01a *Midai lavagtaei vanaktei* 'to Midas the leader and king': *vanaktei* 'king (dat.sg.)' is only considered a loanword because the preceding *lavagtaei* 'leader (dat.sg.)' cannot be natively Phr. and thus must also be a loanword (Obrador-Cursach, 2020, p. 133). There also seems to have been some influence from the Galatians, a Celtic tribe that migrated to Anatolia in the 3rd century BCE (Fortson, 2010, pp. 309, 312). Such influence may be evidenced, for instance, by an apparent loanword from Galatian: τευτου (TM 867070|2|26) 'lack' (Obrador-Cursach, 2020, p. 356). Lastly, a small portion of loans into Phr. derive from various Anatolian languages, Aramaic, Old Persian, Elamite, Assyrian and Latin.

1.3 Phrygian data

A major obstacle with the study of Phr. is the scarce amount of attestations. As Pitts (2018, p. 8) has noted, "there is some diversity of terminology" on how to classify these languages, and I will follow his terms: 'corpus languages' being attested languages that lack living L1 speakers, whereas 'fragmentary languages' form a subset of corpus languages, being poorly attested. The corpus, central to these definitions, is made up of several hundred inscriptions. These texts are accounts of the Phrygians themselves, as opposed to the glosses of the Alexandrian scholar Hesychius (fl. 5th century CE), in which he attempts to explain the meaning of not only obscure AG words, but also those of Phr., Thracian and Macedonian origin (Fortson, 2010, pp. 461, 463, 464).

The Phr. corpus is subdivided into multiple subcorpora, the exact number of which depends on the researcher. Most often, a two-way distinction is made between Old and New Phrygian (Mallory & Adams, 2006, p. 37; Brixhe, 2008, p. 71; Fortson, 2010, p. 461; Ligorio & Lubotsky, 2018, pp. 1817-1818; Fritz & Meier-Brügger, 2021, p. 38). The former dates from the 8th until the 4th century BCE and is attested in inscriptions throughout Central Anatolia. It is composed in the native Phr. alphabet, which has word dividers. The latter dates from the 1st century CE onward and is attested in a more concentrated area of Central Anatolia. It is written in the Greek alphabet in *scriptio continua*, a striking difference with OPhr. A third subcorpus, MPhr., has also been proposed, whose singular certain attestation from the city of Dokimeion is closer to OPhr. in terms of chronology and phraseology (ca. 4th century BCE), but is written using Greek letters, and whose orthography is more akin to NPhr., especially in terms of phonological developments (Obrador-Cursach, 2020, pp. 7-9). Its position in this twilight zone enables some researchers to disregard MPhr. as a useful category (cf. Brixhe 2008, 71; Meier-Brügger 2021, 38).

Research into the Phr. language is by no means a simple endeavour, but is making rapid strides (Obrador-Cursach & Adiego, 2022, p. 9). The fragmentary status of Phr. provides both scarcity of materials as well as a steadily increasing number of attestations to work with (Obrador-Cursach, 2020, pp. 1-2). Thus, the edition on which the digital corpus data have been based is, with the discovery of new inscriptions, outdated or at the very least incomplete (Obrador-Cursach & Adiego, 2022, p. 10). On the other hand, the limited amount of data also limits to what extent we can generalise observed patterns. For instance, certain constructions are not "*categorically* banned from occurring" merely because of poor attestation; they are, at best, "extremely unusual" (Stefanowitsch, 2006, p. 71). As I attempt to demonstrate in §4, a corpus-linguistic approach thus requires grave caution, as practised throughout this thesis.

2. Phrygian: a fragmentary maverick

Phr. has been studied in a modern sense for about two hundred years with varying degrees of in-depth discussion.[7] This chapter provides a brief overview of the literature, focussing on Phr.'s position in the IE family tree.[8] §2.1 deals with Phr. as discussed in general literature on IE linguistics, from introductory to slightly more technical handbooks. §2.2 investigates literature specialised in Phr. itself.

2.1 Skimming the surface: general literature

Mallory and Adams (2006, p. 37) as well as Fortson (2010, pp. 460-463) grant Phr. a special status among fragmentary languages due to its relatively good documentation. Fortson in particular provides a sketch of Phr.'s historical grammar and a text sample. He discusses other fragmentary languages, but no other language receives such a grammatical overview, and only in his discussion of Venetic and Messapic does he include a text sample. Attestation is not enough, however, as a lack of deeper understanding probably prevented Tichy from mentioning Phr. in her *Indogermanistisches Grundwissen* (2009) to focus primarily on the better attested branches. Mallory, Adams and Fortson are, thus, clear outliers in their opinion since most scholars remain silent on Phr. Such is, of course, not without good cause: fragmentary languages play a comparatively less reliable role in reconstruction.

In case researchers do mention Phr., three themes stand out. Firstly, Phr. words are used as additional support for well-established PIE reconstructions, such as verbal morphology (cf. the index of Beekes, 2010, p. 414: e.g. the PIE middle participle suffix **-mh₁nos* attested in NPhr. τιττετικμενος 'accursed') and laryngeal theory (cf. Mayrhofer, 2004, pp. 24-25, 28: e.g. the preservation of initial PIE **h₂* in NPhr. αναρ 'man' < PIE **h₂nḗr*, cf. AG ἀνήρ). The second is basic information, historical or linguistic, on the Phrygians, and not unsurprisingly so. An important part of understanding this language and its origins is understanding the history of its speakers. It is thus not uncommon to encounter a brief summary of the settling in modern-day Turkey and regional importance prior to Alexander's conquests (cf. Mallory & Adams, 2006, p. 37; Fritz & Meier-Brügger, 2021, pp. 38-39; Clackson, 2007, pp. 7, 9; and Fortson, 2010, p. 461). Topics include the migration from the Balkan Peninsula into Anatolia, the influence of the Phr. kingdom and the legendary king Midas, the capital city of Gordium, and the continued existence of the Phr. people long after Macedonian and later Roman takeover. The history of the Phr. people is, however, also easily omitted in literature that prioritises the language itself. Lastly, there is attention to subgrouping the IE daughter languages. Wherever mention is made of Phr.'s genetic status, emphasis lies on its kinship to AG and Armenian, though Thracian, Messapic, Albanian, and Indo-Iranian are also named (cf. Fortson, 2010, p. 462; Olsen & Thorsø, 2022, pp. 209, 216-217; Hyllested & Joseph, 2022, pp. 237-238). This, of course, comes with an

[7] For an overview of the history of research on Phr., cf. Woudhuizen, 2009.
[8] For the presumptions on this matter for the purposes of this thesis, cf. §1.1.

indispensable methodological consideration when studying the Phr. language. Given the scarcity of its linguistic material, subgrouping the language is done under great scrutiny, ideally in accordance with a Brugmannian methodology focused on shared innovations between related languages – relying merely on "a large amount of similarities" is a luxury which can be afforded with, say, Germanic and Romance languages, but not with Phr. (cf. Clackson, 2022, p. 22; for Phr.'s subgrouping, cf. §2.2.2).

2.2 Delving deeper: specialised literature

With regard to literature that focuses on Phr., two major themes are important to discuss: the first are methodological considerations, which have briefly been touched on in the previous section, and the second the genetic status of Phr. At the end of this section both themes are related to each other.

2.2.1 Progress on Phrygian

Obrador-Cursach shares the aforementioned optimism of Mallory, Adams and Fortson. He quotes the former two in his introduction that Phr. has "the greatest claim to consideration" (2020, p. 1), and affirms this sentiment in his foreword with Adiego (2022, pp. 9-10). Brixhe was of the same mind (T. Van Hal, personal communication, August 4, 2024). It is notable that such optimism seems to be a recent development. Other experts are by no means vocal in this regard, but rather keep quiet on the difficulty of extracting much information from little data. Indeed, it is the "near-Sisyphean" (Obrador-Cursach, 2020, p. 1) nature of research into Phr. that warrants optimism. Progress of our understanding of Phr. is, namely, on the one hand incremental, but on the other hand steady. Optimism should not, however, overshadow the real obstacle that is scarcity of data. As Fortson so prudently warns, one ought to exercise caution when dealing with fragmentary languages (2010, pp. 459-460).

2.2.2 Phrygian's genetic status

When it comes to classifying Phr. as an IE language, as specialists do not doubt its place in that family, one overarching *communis opinio* is prevalent among all: Phr. was part of a subgroup spoken in the Balkans, in close contact with some IE languages and separate from others. That in itself is not very telling, but there are a few languages that recur as candidates for other members of the subgroup, some of them more probable than others, in particular AG, Indo-Iranian, Albanian, Thracian, and even Balto-Slavic. It is in all likelihood safest to adopt the position of a common subgroup when it comes to genetically classifying Phr., although that does not preclude considering closer relations. Three routes are presented here.

Firstly, there is the idea of Thraco-Phrygian unity. Haas (1966) is a proponent of this position, though it is quite flawed. For one, Thracian is classified as a satem language, whereas Phr. is centum. It is unlikely that those two types should be grouped together (Brixhe, 1997, p. 178; Brixhe, 2008, p. 72) when another member of the

proposed subgroup, namely AG, is a more likely candidate on the basis of also being centum language. Furthermore, since Thracian is more scantily attested than Phr., grouping two fragmentary languages so closely together is methodologically suspect (cf. Fortson, 2010, p. 464). As a result, there is both less data to disprove the idea and more opportunity to argue for it from silence.

Secondly, there is the emphasis on Phr.'s relationship with Armenian. Adherents of this position include Diakonoff (1984), Neroznak (1985 with Diakonoff) and more recently Hämmig (2017). Considering the *communis opinio* that the original subgroup diverged as some constituent peoples migrated eastwards (though Drews, 1993 has argued against this), it is tempting to subscribe to this idea on a geographic basis, since both Armenian and Phr. are attested across the Hellespont from the Balkan Peninsula. The scarcity of Phr. data, however, remains a major obstacle to proving this as Hämmig (2017, p. 54) has noted herself, as well as the dubious development of PIE intervocalic **p* to Phr. *v* – a similar development to Armenian (Hämmig, 2013, p. 136) – singularly attested on B-05 (Obrador-Cursach, 2020, p. 72).

Lastly, there is the mainstream view that Phr. has a particularly close relationship to AG. This position can be sorted on a spectrum, with one end where it is merely observed that Phr. and AG share quite a few similarities (the soft position), and the other end where there must have been a Graeco-Phrygian ancestor that was part of the Balkanic subgroup (the hard position). Orel (1997) is a clear proponent of the soft position, whereas Brixhe (softer in 1997, but harder in 2008), Ligorio and Lubotsky (2018) and Obrador-Cursach (2020; his 2019 publication is dedicated to the matter) clearly lean to the hard position.

Research into the genetic status of Phr. may yield informative results, not only for Phr., but also for the posited related languages. However, Phr. being a fragmentary language of course limits the extent to which this is possible. As a result, making radical conclusions based on scarce evidence might prove detrimental to Phr. studies unless the evidence is clear cut. What counts as too radical is, of course, subjective. If the Phr. language, promising in the eyes of Mallory, Adams, Obrador-Cursach and Adiego, is to be uncovered, it will surely be in no small part thanks to additional archaeological and historical research, not merely linguistic endeavours such as this thesis.

3. Annotating the corpus

This chapter describes the annotation process for the NPhr. subcorpus. Since the main focus of this project are the annotated data, this description is key to understanding the corpus. After some preliminary remarks (§3.1), I provide insight into the application of UD conventions to tokenization, morphology and syntax (3.2.0-3.2.3), as well as some difficulties along the way (§3.2.4) and, lastly, some examples (§3.3).

3.1 Preliminary remarks

Phr. texts can be sorted, as previously mentioned (cf. §1.3), into two or three subcorpora. OPhr., alternatively called Palaeo-Phrygian, is already preserved in the TITUS database,[9] which uses an edition by Brixhe and Lejeune (1984). Practicality, however, was the main motivator for me to annotate only the NPhr. subcorpus. The task of annotating both OPhr. and NPhr. inscriptions, let alone the dubious MPhr. ones, would have demanded too much time and effort for a thesis to be completed in merely two semesters. OPhr. would also have been more problematic regardless of availability due to its highly fragmentary nature, which in turn would have made most annotations tentative at best. Since most NPhr. inscriptions are bilingual, often also featuring AG, it was thus important to further isolate the Phr. data during annotation to ensure that only those data come up as search results when using the corpus.

For the purpose of annotation, an up-to-date edition of the inscriptions had to be selected. Thus, the text from Obrador-Cursach (2020) was chosen, which offers not only the inscribed text, but also a translation wherever possible. In addition, lemmas are provided, next to a general overview of Phr. grammar. This edition was available digitally, viz. as an e-book edition. However, there was no automatic way to copy the inscriptions, meaning I had to make use of regular expressions and manual corrections in an excel file in order to import the texts into the annotation environment. A tremendous grammatical aid was the chapter by Ligorio and Lubotsky (2018). For it concisely presents information that is sometimes absent in Obrador-Cursach's otherwise comprehensive work. Due to the open nature of the Phr. corpus, however, this annotated corpus in a sense starts out flawed. Epigraphical editions that compile all available texts quickly become obsolete as new texts are published. Obrador-Cursach's edition is no exception in that regard (cf. Obrador-Cursach & Adiego, 2022, pp. 9-10). Instrumental to the annotation was a FileMaker environment that had previously been used to annotate AG data (cf. Keersmaekers et al., 2019) and, in the case of Pitts (2018), the Sabellic languages.

To help future researchers link the texts to pertinent metadata such as provenance and approximate date, each sentence is connected to a stable Trismegistos identifier which serves to identify each inscribed artifact.[10] The

[9] https://titus.fkidg1.uni-frankfurt.de/texte/etcs/phrygian/phryg.htm
[10] For more on TM, cf. Depauw & Gheldof, 2014.

identifiers are accessible online.[11] The structure of such an identifier is straightforward: TM followed by a string of numbers which represents the physical artifact, followed by an ordinal representing the sentence of the inscription, followed by an ordinal representing the token in the sentence. Thus, TM 867450|2|5 (.) is the fifth token in the second sentence of artefact TM 867450.

3.2 Annotating the corpus

Since no annotated corpus of the Phr. language exists, it fell on me to devise some conventions using Universal Dependencies.[12] In order to define a UD standard for Phr. (language code xpg), the parameters were based on the standard for AG (grc). The decision to use UD standards for this project instead of, say, the ones employed in the Perseus Project was informed by its widespread application to many different languages outside the classical sphere, thus broadening its namesake universality.

3.2.1 Tokens

In the annotation environment, each token was subject to the same possible parameters. That is, each had a field for the token itself, its lemma, its syntactic category, its syntactic head, and several morphological tags. Among the 1722 tokens were not only the attested words and punctuation, but also (1) strings of characters without clear segmentation and/or meaning and syntax, and (2) *hederae*, decorative inscribed ivy leaves represented in the data by fleurons (❧; cf. Fig. 3.1), which are consistently annotated by UD standards as a symbol (SYM).

Figure 3.1: A hedera *from a NPhr. inscription (TM 984517|2|1) (Drew-Bear, Lubotsky & Üyümez, 2008, p. 110).*

The annotation as a whole was performed manually.

When going through each token, it was important to look up each word, as a single lemma can vary greatly throughout the corpus. I catalogued such variation in a list containing relevant information on each lemma where it is available in the above mentioned source publications. The list will be included in the dataset on GitHub.[13] It is structured into four columns containing the lemma, its variations, translation and miscellaneous remarks such as etymological information, its relation to other Phr. words and notable epigraphic problems.

[11] https://www.trismegistos.org/lang/detail.php?language_id=87.
[12] https://universaldependencies.org/.
[13] https://github.com/OggiPeeters/UD_Phrygian-KUL.

3.2.2 Morphology

When it comes to morphology, UD distinguishes POS tags from features. In my annotated data, I recognise 16 POS tags (cf. Tab. 3.1). The only difference from the AG POS tags is the explicit absence of the 'other' tag (X), which is not used already in UD version 1.[14]

POS	UD abbreviation	# of attestations
Adjectives	ADJ	21 tokens
Adpositions	ADP	67 tokens
Adverbs	ADV	16 tokens
Auxiliaries	AUX	*No auxiliary verbs attested*
Coordinating conjunctions	CCONJ	110 tokens
Determiners	DET	110 tokens
Interjections	INTJ	*No interjections attested*
Nouns	NOUN	347 tokens
Numerals	NUM	1 token
Particles	PART	82 tokens
Pronouns	PRON	150 tokens
Proper nouns	PROPN	106 tokens
Punctuation	PUNCT	212 tokens
Subordinating conjunctions	SCONJ	*No subordinating conjunctions attested*
Symbols	SYM	67 tokens
Verbs	VERB	325 tokens

Table 3.1: POS tags for Phr. UD.

The features present for Phr. are 14 (cf. Tab. 3.2). The difference from AG here is more expansive. For one, UD AG also features a category for polarity (Polarity; e.g. for negation particles like οὐ and μή), which is absent from the Phr. standard. Secondly, the features do overlap somewhat. As a result, VerbForm, which covers participles, is brought under Mood, reflecting traditional grammar where participles are considered a non-finite mood like infinitives. Also, NumType cannot meaningfully be used due to a lack of attestation: πινκε 'five' (TM 984535|4|3) is the only attested numeral and is an ordinal.

[14] https://universaldependencies.org/grc/pos/X.html.

Feature	UD abbreviation	Possible values
Case	Case	Nominative (Nom) Accusative (Acc) Genitive (Gen) Dative (Dat)
Definiteness	Definite	Definite (Def) Indefinite (Ind)
Degree of comparison	Degree	Positive (*zero-marking*) Comparative (Cmp) Superlative (Sup)
Form of verb or deverbative	VerbForm	Finite (Fin) Participle (Part)
Gender	Gender	Masculine (Masc) Feminine (Fem) Neuter (Neut)
Mood	Mood	Indicative (Ind) Subjunctive (Sub) Optative (Opt) Imperative (Imp) Infinitive (Inf) Participle (Part)
Number	Number	Singular (Sing) Plural (Plur)
Numeral type	NumType	Ordinal (Ordinal)
Person	Person	First person (1) Second person (2) Third person (3)
Possessive	Poss	Not possessive (*zero-marking*) Possessive (Yes)
Pronominal type	PronType	Demonstrative (Dem) Emphatic (Emp) Indefinite (Ind) Relative (Rel)
Reflexive	Reflex	Not reflexive (*zero-marking*) Reflexive (Yes)
Tense	Tense	Present (Pres) Aorist (Aorist) Perfect (Perfect)
Voice	Voice	Active voice (Act) Middle voice (Mid)

Table 3.2: Features and values for Phr. UD.

The feature ‘Tense’ is UD terminology. In §4.2, however, this label is called into question with regard to the NPhr. data.

3.2.3 Syntax

Despite the formulaic nature of the NPhr. inscriptions, the range of possible syntactic relations is rather extensive. The relations present in the annotated data are the following (Tab. 3.3).[15]

[15] For a list of all possible relations, cf. https://universaldependencies.org/u/dep/index.html.

Syntactic relation	UD abbreviation	# of attestations
Adjectival modifier	amod	28 tokens
Adnominal clause	acl	7 tokens
Adverbial clause modifier	advcl	2 tokens
Adverbial modifier	advmod	20 tokens
Appositional modifier	appos	1 token
Auxiliary	aux	82 tokens
Case marking	case	67 tokens
Clausal subject	csubj	84 tokens
Conjunct	conj	82 tokens
Coordinating conjunction	cc	105 tokens
Copula	cop	4 tokens
Determiner	det	118 tokens
Dislocated element	dislocated	1 token
Fixed multiword expression	fixed	78 tokens
Flat expression	flat	1 token
Goes with	goeswith	3 tokens
Indirect object	iobj	113 tokens
Marker	mark	2 tokens
Nominal modifier	nmod	20 tokens
Nominal subject	nsubj	138 tokens
Object	obj	117 tokens
Oblique nominal	obl	90 tokens
Orphan	orphan	1 token
Punctuation	punct	209 tokens
Root	root	135 tokens
Unspecified dependency	dep	3 tokens

Table 3.3: Syntactic relations for Phr. UD.

3.2.4 The annotation process

The UD conventions for AG formed the guiding principles for annotating Phr. inscriptions. Both secondary literature as well as the grammatical overview included in the text edition provided enough insight to annotate the tokens for their inflections, whereas knowledge of AG suffices to understand the syntax of the texts, which are simple and

mostly formular. I rarely refrained from annotating, only if I could not find a source that would aid in interpreting the inscriptions. Morphology and syntax thus presented relatively few problems.

Perhaps the biggest obstacle was dealing with uncertain readings, the result of the inscriptions' physical state. Because of my strong adherence to Obrador-Cursach's edition, I refrained from obfuscating the text as he presents it. Thus, no character is lost nor group of characters emended. As a result, plenty of tokens make no sense, containing either epigraphical representations for the many lacunae among the inscriptions or seemingly random Greek letters to which no meaning can be assigned. Furthermore, the emendations proposed by Hämmig (2022, pp. 94-96) of ουελασκετου and ουελασκοννου, which I am otherwise inclined to follow, are not included in the corpus directly.

[...] α Τι αδειτου **ουελασκετου** κε ισνου ασ[?]τοι παρτης (TM 984529)

[...] Τιε τι[τετικ]μενος ειτυ **ουελασκοννου** κ' ΗΝΚ [...] (TM 984534)

[...] [ζε]μελως κε [.][?] **ο[υ]ελασκοννου** κε ισνιο[υ[?] - - -] ΑΙ παρτης (TM 918557)

They are, however, mentioned in the 'notes'-column in the list of lemmata, which will be part of the published dataset.

3.3 Examples

In this section, I present some examples of annotated tokens. First, some common tokens will be shown. After that, some more obscure tokens will highlight the adversity encountered while annotating. Below is a screenshot of the annotation environment in FileMaker, in particular the record for TM 984524|3.

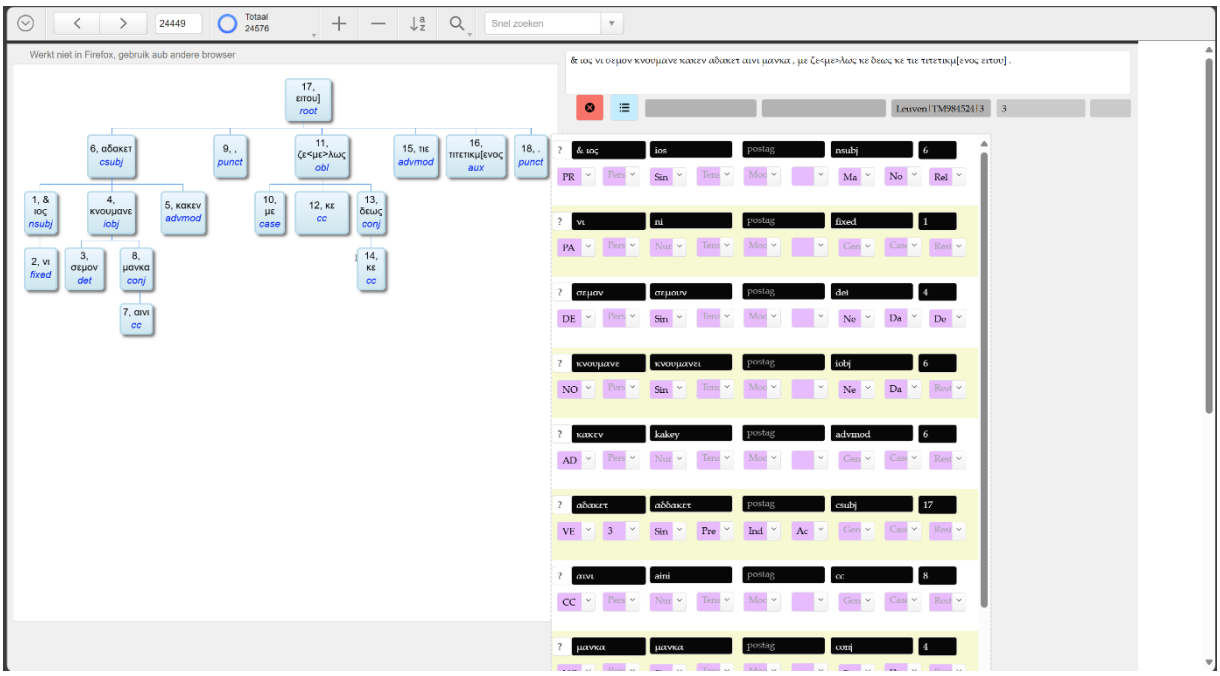

Figure 3.2: An annotated record (TM 984524|3) in FileMaker.

On the left, there is a dependency tree showing the relationship between each token along with a number indicating what position the token is in the record, the token itself and its syntactic function. In the upper right of the screen are all tokens of the record positioned according to their assigned number value, which itself corresponds to the original Phr. text. Directly underneath that are several fields, of which only the third and second to last are of note. The third to last field gives the Trismegistos identifier. The second to last field gives the sentence of that inscription, since a single text can of course contain multiple sentences, which need to be distinguished. Following are two examples of tokens of this record to demonstrate how each token in the corpus was annotated.

The first example shows ιος '(he) who', a commonly occurring word in curse formulas in general. Since many sentences start with ιος, it is often preceded by a dollar sign or ampersand to respectively mark the beginning of a new inscription or sentence in the same inscription. The symbols were added to aid in converting the data to the CoNLL-U format used by UD.

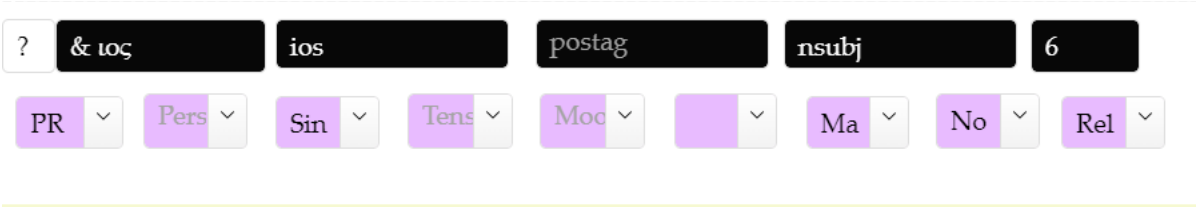

Figure 3.3: The token & ιος *(TM 984524|3|1) in FileMaker.*

From left to right, the upper row shows the token (& ιος) followed by its lemma 'ios'. The middle box is empty as it is a vestige of different annotations of AG texts, which happen to use the same annotation environment. The second to last box shows that our token is the nominal subject (nsubj) of its head in the final box, which is token 6 of this record. The lower row shows morphological information. In the first box, the POS tag is selected as PRON (a pronoun). The second box indicates person, which is irrelevant for this token. The third box marks our token's number as singular (Sing). The following three boxes indicate the tense, mood and voice respectively, which are in this case, again, irrelevant. The third to last box indicates the gender of the token as masculine (Masc), the second to last box indicates the nominative case (Nom). The final box indicates extra, secondary information. In this case, the token is marked as a relative pronoun (Rel). When converting to UD's CoNLL-U format, the data from this box can simply be added to the 'features'-column.

The second example shows ειτου 'let him be', a commonly occurring word in curse formulas of the type 'If one does harm to this monument, let him be accursed'. In this case, it is marked as being recognisable, but damaged on the inscription – hence the closing square bracket.

Figure 3.4: The token ειτου] *(TM 984524|3|17) annotated in FileMaker.*

The structure of this token is identical to the previous example, though of course filled in differently.

The second to last box shows that our token is the root of the record and it is thus independent from every other token, hence its head is named '0' in the final box. In the first box of the lower row, the POS tag is selected as VERB. The second box indicates that the verb is conjugated for the third person, followed by the singular for number. The third, fourth and fifth boxes show that ειτου is a present tense verb (Pres) in the imperative mood (Imp) and active voice (Act). As this is all relevant morphological information, the last three boxes are left empty.

The third example shows a fleuron, which indicates that there is a *hedera* in the actual inscription. Due to their commonality, these *hederae* are annotated as a token like any other where they could have been left out just as easily.

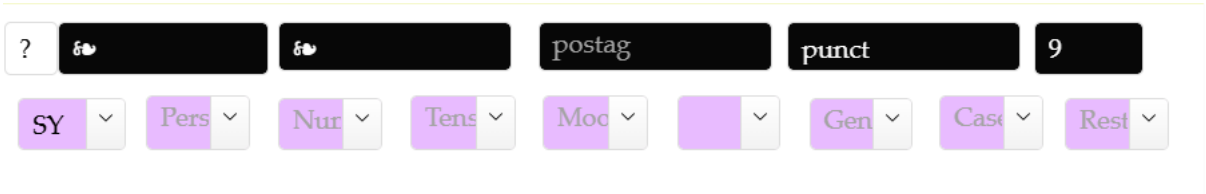

Figure 3.5: The token ❧ (TM 984525|2|15) annotated in FileMaker.

Since there is no morphological variation in the *hederae*, token and lemma are the same. They are consistently marked as punctuation (punct) dependent on the root of the record, in this case token 9. The UD POS tag is symbol (SYM) and further morphological tagging does not apply.

Below is another screenshot of the annotation environment. Its structure is identical to the previous one, but here it shows a particularly difficult record to annotate. The contents of the sentence were so impenetrable, I had no other option than to leave the annotations blank for this record. Had there been a computational way for me to put an entire record between *cruces desperationis*, I would have probably done that instead. This record is an outlier however. Out of the 1722 tokens in the corpus overall, 108 tokens (6.2%) are not annotated for POS, 36 tokens (2.1%) for lemma and 211 tokens (12.2%) for syntactic relation.

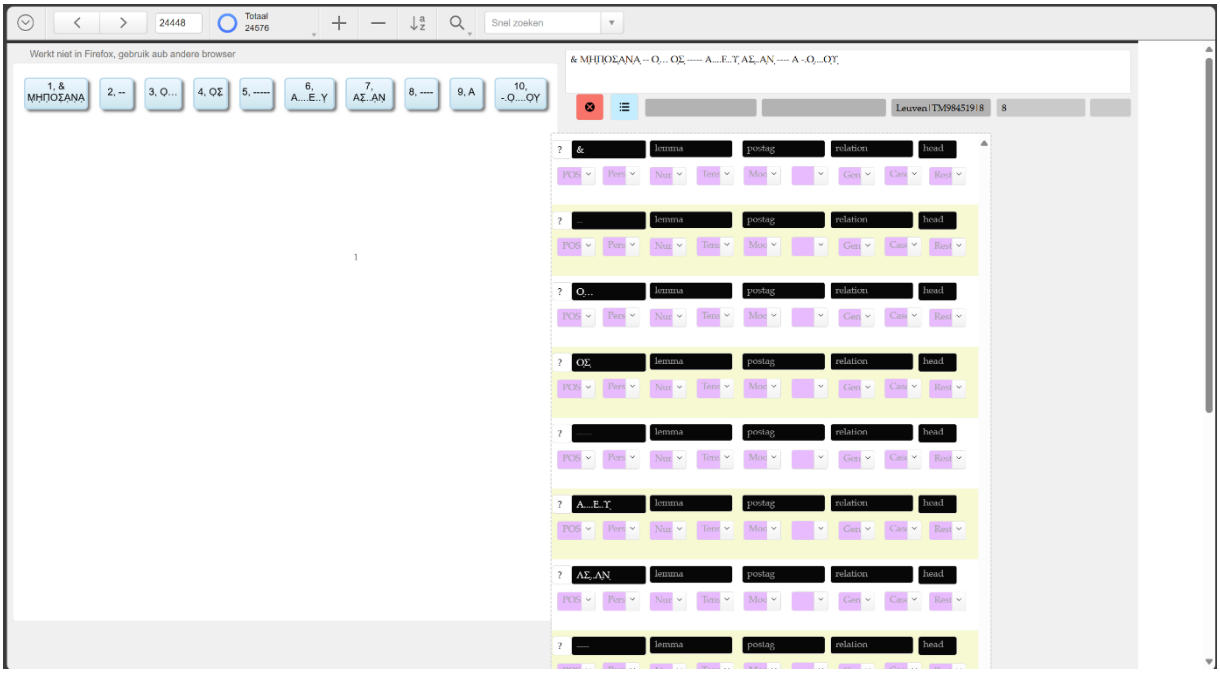

Figure 3.6: A record in FileMaker (TM 984519|8) showcasing the extent to how illegible Phr. inscriptions can be.

4. Towards a corpus-based typology of Phrygian

This chapter attempts to sketch a basic typology of NPhr. based on the annotated data. Two morphological topics are addressed: the case system (§4.1) and the augment (§4.2). Due to the formulaic nature of the inscriptions, there is very little variation in terms of word order. As a result, syntax is not examined in depth here, however very basic word order patterns can be observed (§4.3). Lastly, a brief investigation of the most frequent lemmas is presented (§4.4).

4.1 The case system

Like in AG, Phr. nominals are inflected for case, gender and number (Ligorio & Lubotsky, 2018, pp. 1824-1826; Obrador-Cursach, 2020, pp. 76-83). There is no morphological distinction between nouns and adjectives when it comes to the suffixes. The cases are nom., acc., gen. and dat. Due to the nature of the inscriptions, there is no vocative attested like in AG nor a distinct locative, of which Orel recognises a remnant in OPhr. *iman* (1997, pp. 387, 390). The three genders are masculine, feminine and neuter, and there is a distinction between singular and plural. Again, unlike in AG, no dual form is attested. The inflection for the three categories takes on the form of a suffix that distinguishes three basic declensions: o-stems, ā-stems and consonant (C-) stems. Semivocalic i- and u-stems also exist, as well as e-stems, but are poorly attested. The endings either closely resemble their AG counterparts or can easily be derived from PIE, as the table below (4.1) illustrates.

		o-stems		**ā-stems**		**C-stems**	
		OPhr.	**NPhr.**	**OPhr.**	**NPhr.**	**OPhr.**	**NPhr.**
Sg.	**Nom.**	-os	-oς	-a [f.]; -a(s) [m.]	-α [f.]	-s/-Ø	-ς/-Ø
	Acc.	-un	-ουν/-ον	-an	-αν	-an; -Ø [n.]	-αν; -Ø [n.]
	Gen.	? -ovo	-ου	–	-ας	-os	-ος
	Dat.	-oi/-oy	-ου	-ai-/ay	-αι/-α	-ei/-ey	-ε(ι)/-ι/-η
Pl.	**Nom.**	-oi	-α [n.]	–	-ας	-a [n.]	-ης
	Acc.	? -oys; -a [n.]	-α [n.]	–	-αις/-αης	? -ais	? -αις/-αης
	Gen.	–	-ουν	–	–	–	–
	Dat.	? -oys	-ως	–	–	–	–

Tab. 4.1: Phr. endings for the three declensions based on Ligorio & Lubotsky (2018, p. 1824).

Our knowledge on the uses of each case is severely limited due to the nature of the texts (Obrador-Cursach, 2020, pp. 114-115). What is known of their functions, however, reflects very basic IE syntax:

- Nom. indicates the subject of a sentence as well as the predicate noun;
- Acc. indicates the direct object of a transitive verb;
- Gen. indicates attributive dependency on a syntactic head;

- Dat. indicates the indirect object of a ditransitive verb as well as oblique case functions, such as in prepositional phrases.

Prepositional phrases also require a specific case. Only the nom. does not occur with a preposition.

With the annotated corpus, it is possible to take a close look at the Phr. case system. The parts of speech (POS) that inflect for case are adjectives, determiners, nouns, pronouns, proper nouns, and verbs (participles to be precise). There are only categories for the four attested cases. Furthermore, a number of tokens remain uncertain for POS and case due to text-critical uncertainty. The table below (4.2) illustrates the correspondence between POS and case.

	ADJ	DET	NOUN	PRON	PROPN	Part	Uncertain	Total
Nom	12	2	17	119	22	83	3	258 (27.5%)
Acc	5	5	136	8	8	13	7	182 (19.4%)
Gen	1	1	7	2	17	0	0	28 (2.9%)
Dat	2	102	174	21	55	0	3	357 (38%)
Uncertain	1	0	13	0	4	0	95	113 (12%)
Total	21 (2.2%)	110 (11.7%)	347 (36.9%)	150 (15.9%)	106 (11.3%)	96 (10.2%)	108 (11.5%)	**938**

Tab. 4.2: The 938 tokens that inflect for case per POS.

As will become apparent, the nature of the inscriptions plays a big part in the distribution of the cases. The following curse formula is particularly prominent:

ιος νι σεμουν κνουμανει κακουν αδδακετ, τιττετικμενος ειτου

ιο-ς νι σεμουν κνουμαν-ει κακ-ουν αδδακετ, τιττετικμεν-ος ειτου

REL-NOM PART DEM:DAT tomb-DAT harm-ACC does accursed-NOM become

'whosoever does harm to this tomb, let him become accursed'

This prominence colours our observation as non-formulaic language is not represented in the data.

There are 258 nom. in the corpus (27.5%), where two POS stand out in particular. Firstly, the pronoun has the greatest share with 119 tokens. 101 of those comprise the relative pronoun ιος '(he) who' in the relative clause of the curse formula. Secondly, 83 participles are in the nom. They include 75 tokens of the lemma τιττετικμενος 'accursed', which is also part of the curse formula in the main clause. As a result, nearly all participles function as

an auxiliary to ειτου ‘let (them) become’. The only exception ε[..]γενṭουμενος[16] is the root of the following corrupt sentence (Obrador-Cursach, 2020, p. 526).

ε[..]γενṭουμενος νιοισιος ναδροτος ειτου (TM 867450)

ε[..]γενṭου-μεν-ος νιοισι-ος ναδροτ-ος ειτου

?-PTCP.NOM ?-NOM impotent-NOM become

‘let him become X-ed and νιοισιος and impotent’

It is not clear here whether the participle should be interpreted as an adjective seeing as it is followed by two more adjectives and ειτου as a copula. Cf. the distribution of all nom. tokens below (4.3).

	ADJ	DET	NOUN	PRON	PROPN	Part	Uncertain	Total
Amod	6	0	0	0	0	0	0	6 (2.3%)
Appos	0	0	0	0	1	0	0	1 (0.3%)
Aux	0	0	0	0	0	82	0	82 (31.8%)
Conj	6	0	1	0	4	0	0	11 (4.3%)
Det	0	2	0	1	0	0	0	3 (1.2%)
Flat	0	0	0	0	1	0	0	1 (0.3%)
Nsubj	0	0	11	117	10	0	0	138 (53.4%)
Root	0	0	0	0	1	1	0	2 (0.8%)
Uncertain	0	0	5	1	5	0	3	14 (5.4%)
Total	12 (4.7%)	2 (0.8%)	17 (6.6%)	119 (46.1%)	22 (8.5%)	83 (32.1%)	3 (1.2%)	**258**

Tab. 4.3: The 258 nom. tokens for POS and UD relation.

As for the 182 acc. (19.4% of the inflected tokens), the vast majority of those are nouns (136). In turn, direct objects comprise most of those nouns (114). As part of the curse formula, the word for ‘harm’ κακουν occurs often (86 times). 3 of the 5 nouns marked as conjuncts (conj) are syntactically equivalent to direct objects: μμυρα (TM 984544|1|8), τεᾳμ[ας (TM 845306|1|9), and κακουν (TM 917779|1|7). The only other objects in the corpus are the relative pronouns ιον (TM 867098|2|4) and ιαν (TM 932138|1|9) (lemma *ios*). Cf. the distribution of all acc. tokens below (4.4).

[16] TM 867450|2|1.

	ADJ	DET	NOUN	PRON	PROPN	Part	Uncertain	Total
Amod	4	0	1	0	0	13	0	18 (9.9%)
Conj	0	0	5	0	0	0	1	6 (3.3%)
Det	0	5	0	2	0	0	0	7 (3.8%)
Dislocated	0	0	1	0	0	0	0	1 (0.5%)
Nmod	0	0	1	0	0	0	0	1 (0.5%)
Obj	0	0	114	2	0	0	0	116 (63.7%)
Obl	0	0	5	1	6	0	0	12 (6.6%)
Root	1	0	0	0	0	0	0	1 (0.5%)
Uncertain	0	0	9	3	2	0	6	20 (11%)
Total	5 (27.5%)	5 (27.5%)	136 (74.7%)	8 (4.4%)	8 (4.4%)	13 (7.1%)	7 (3.8%)	**182**

Tab. 4.4: The 183 acc. tokens for POS and UD relation.

The gen. is strikingly rare with only 28 attestations out of 938 declined tokens (just under 3%). 17 of those are proper nouns, all but one of which serve as nominal modifiers: εκατηας (TM 984535|2|15) 'Hecataean' is an adjectival modifier to the noun κναικος 'woman (gen.)' (lemma *knais*), whose syntactic relation is unclear. Furthermore, 10 of those nominal modifiers are in reference to the god Ti-,[17] whereas the other 6 refer to people.[18] With regard to gen. nouns, two points are of note. Firstly, there are 2 instances of αργου 'on account of, for the sake of', which is used as a postpositional case marker with the acc. noun ευκιν 'vow (acc.)' (from AG εὐχή). Something similar appears in AG, where χάριν can be used as postpositions with a similar meaning to Phr. after a gen. noun (van Emde Boas et al., 2019, p. 397), as well as in Latin, with the ablative nouns *grātiā* and *causā* (Obrador-Cursach, 2020, p. 180). The nom. of the Phr. noun is uncertain. Secondly, there is 1 attestation of the gen. being used as an iobj., a function otherwise reserved for the dat. The noun κνουμινος (TM 931700|2|4) 'tomb, memorial (dat.)' fills the slot of κνουμανει in the curse formula, thus its function is well-established: case morphology is its only striking feature.

ι⟨ο⟩ς κε σεμουν κ⟨ν⟩ουμινος ⟨κακουν⟩ αδακεν […] (TM 931700)

ι-⟨ο⟩ς =κε σεμουν κ⟨ν⟩ουμιν-ος ⟨κακ-ουν⟩ αδακεν

who-NOM =and this:DAT tomb-GEN harm-ACC do

'(he) who does harm to this tomb…'

[17] The nominative is attested nowhere, hence I follow Lubotsky (2004) in referring to him by stem.
[18] At least, they are assumed to be people's names.

Cf. the distribution of all gen. tokens below (4.5).

	ADJ	DET	NOUN	PRON	PROPN	Total
Amod	0	0	0	0	1	1 (3.5%)
Case	0	0	2	0	0	2 (7.1%)
Det	0	1	0	1	0	2 (7.1%)
Iobj	0	0	1	0	0	1 (3.5%)
Nmod	1	0	2	0	16	19 (67.8%)
Uncertain	0	0	2	1	0	3 (10.7)
Total	1 (3.5%)	1 (3.5%)	7 (25%)	2 (7.1%)	17 (60.7%)	**28**

Tab. 4.5: The 28 gen. tokens for POS and UD relation.

Most of the 938 inflected tokens are dat. (357, i.e. 38% of the inflected tokens). The dat. is not unexpectedly dominated by nouns (174) and determiners (102). The relative clause of the curse formula requires an entity being harmed, most often κνουμανει 'tomb, memorial (dat.)', as well as a demonstrative with that entity (σεμουν 'this [dat.]'). Thus, 100 nouns serve as iobj. with 92 determiners being dependent on an iobj. The curse formula also plays a role in the dat. obliques (75). 26 of the 30 oblique nouns are of the lemma ζεμελως 'humans (dat.pl.)' or *devos* 'god', which are common additions to the main clause as με ζεμελως κε δεως κε 'in sight of both men and gods' or a variant thereof. Furthermore, all 45 oblique nouns are in reference to Ti-, who is added as the agent in the main clause of a curse Τιε τιττετικμενος ειτου 'let him be(come) accursed by Ti-' or as a witness ατ Τιε τιττετικμενος ειτου 'let him be(come) accursed before Ti-'. Cf. the distribution of all dat. tokens below (4.6).

	ADJ	DET	NOUN	PRON	PROPN	Uncertain	Total
Advmod	0	0	0	0	1	0	1 (0.3%)
Amod	2	0	0	0	0	0	2 (0.5%)
Conj	0	0	37	1	3	0	41 (11.5%)
Dem	0	1	0	0	0	0	1 (0.3%)
Det	0	99	0	6	0	0	105 (29.4%)
Goeswith	0	0	1	0	0	0	1 (0.3%)
Iobj	0	0	100	9	3	0	112 (31.3%)
Obl	0	0	30	0	45	0	75 (21%)
Orphan	0	1	0	0	0	0	1 (0.3%)
Uncertain	0	1	6	5	3	3	18 (5%)
Total	2 (0.5%)	102 (28.5%)	174 (48.7%)	21 (5.8%)	55 (15.4%)	3 (0.8%)	**357**

Tab. 4.6: The 357 dat. tokens for POS and UD relation.

Of the 113 tokens of uncertain case (12% of the inflected tokens), 101 are also uncertain for UD relation. The majority of those (87) are also uncertain for POS. Considering the illegibility of some inscriptions (cf. Fig. 3.5) and the sheer limits on our knowledge of Phr., this is not surprising. E.g. ουεβρα (meaning uncertain; lemma *vebras*; TM 984534|1|3), τοᾳ (meaning uncertain; TM 984535|2|9), παρτυς (meaning uncertain; lemma *παρτης*; TM 846124|1|10), etc. One can only hope that some of the tokens, summarised in the table below (4.7), will be better understood as the language itself is better understood.

	ADJ	NOUN	PROPN	Uncertain	Total
Dep	0	0	0	1	1 (0.9%)
Goeswith	0	0	0	1	1 (0.9%)
Obj	0	1	0	0	1 (0.9%)
Obl	0	1	1	1	3 (2.6%)
Root	0	1	0	5	6 (5.3%)
Uncertain	1	10	3	87	101 (89.3%)
Total	1 (0.9%)	13 (11.5%)	4 (3.5%)	95 (84%)	**113**

Tab. 4.7: The 113 tokens of uncertain case for POS and UD relation.

To conclude, these data affirm the consensus that Phr. abided by basic IE syntax, as shown by the cases used in the curse formula. There is also overlap with other IE languages with regard to the postposition αργου 'on account of' following the gen., like in Lat. and AG, and the dat. functioning as an oblique case, like in AG. There is a limit, however, to what can be extracted from the limited data. For instance, marginal results are not necessarily representative for the entire system, as we are unable to distinguish between an infrequent use of a noun case that is grammatical versus one that is ungrammatical. The aforementioned usage of gen. κνουμινος for dat. κνουμανει as iobj exemplifies this: is this a one-off or can the gen. genuinely express an iobj in Phr.?

4.2 The augment

The PIE augment **(h₁)é* was a prefix to secondary tenses of the indicative mood, originating as a particle that indicated the past tense (Beekes, 2011, pp. 252-253; Clackson, 2007, p. 131; Fritz & Meier-Brügger, 2021, p. 188). It is attested in Indo-Iranian, AG, Armenian and Phr., though there is variation among them in use (Fortson, 2010, p. 101). In Old Armenian, the augment is present in past tense forms that would otherwise be monosyllabic (Beekes, 2011, p. 252; Fritz & Meier-Brügger, 2021, p. 189), and in Indo-Iranian and Homeric Greek, secondary tenses can appear both with or without it (Beekes, 2011, pp. 273-274). When the augment is absent from those tenses in Sanskrit, it is considered a separate 'injunctive' mood, which has its own functions that do not appear to correspond to the 'gnomic' forms in AG (Beekes, 2011, pp. 273-274; Clackson, 2007, pp. 131-132; Forston, 2010, p. 101).

The situation in Phr. is less clear due to the aforementioned scarcity of data. The literature recognises only three attested tenses: present, perfect and aorist (Ligorio & Lubotsky, 2018, p. 101; Obrador-Cursach, 2020, p. 98). This is based mainly on morphology, which may go back to PIE, and context clues provided by the inscriptions that hint at the meaning of the texts and particular obscure words. Because of this, very little is known about the functions these verb tenses served, such as differences based on aspect or even on temporal deixis. As a result, it may be safer to speak of 'verbal stems' instead of tenses. In the corpus, only the aorist stem is of concern when discussing the augment. Only there does the literature recognise the presence of the augment, which is, again, based on PIE morphology and epigraphical context. However, there is the problem of the uncertain readings of inscriptions. Because of this, words may be analysed and segmented a particular way to accommodate an interpretation, which may lead to circular reasoning.

There are only 7 aorist tokens: εγδαες[19] (TM 984531|1|29) 'he destroyed(?)', εναρκε (TM 984535|2|12; meaning unclear), ποσεκανες (TM 984535|2|17) 'he dug(?)', εδαες[20] (TM 984535|4|2) 'he did, he put', εγερετοι (TM 918640|1|15) 'he took, he brought(?)', ενεπαρκες[21] (TM 932138|1|16) 'he inscribed' and ουρνουσαεν (TM 844853|1|20; meaning unclear). All of them are in the third person and most share other morphological tags. The first 6 aorists are singular in person and in the indicative mood. It is unfortunate that no other moods are attested in the aorist to provide insight into its morphology. It does, however, grant us positive evidence, since the augment ε- is present in all forms. Like in AG, it is appended directly onto the verb stem (Smyth, 1920, pp. 145-146), which makes it that compound verbs have their augment follow their preverb, such as in ποσ-ε-κανες 'he dug (?)' and εν-ε-παρκες 'he inscribed' (cf. Obrador-Cursach, 2020, pp. 103, 109, 227, 335). On the other hand, the apparent exclusive attestation of the augment with the aorist in such a small yield tells us nothing of the motivation for the augment: is the augment really mandatory with secondary tenses or can it be omitted in certain environments, which are simply absent from this corpus? Cf. in Homeric Greek (De Decker, 2022):

1. Syntactic factors such as
    a) Drewitt-Beck's clitic rule: a past tense form followed by a word that cannot appear in first position (cf. Wackernagel's Law of clitics) will be unaugmented;
    b) Kiparsky's reduction rule: a series of past tense forms will have the tendency to augment only the first form in order to reduce superfluous markedness;
2. Semantic criteria like
    a) Background information does not use the augment (perhaps the main factor for its absence);
    b) The tendency to use fewer augments in narrative descriptions than in speeches;
    c) Emphatic or surprising information, or a new element in an enumeration prefers an augmented secondary verb form;
    d) Actions in the recent past or with relevance to the present (such as Homeric similes) feature the augment;

[19] Cf. AG φθί(ν)ω (Beekes, 2010, p. 1570; Gorbachov, 2005, p. 192).
[20] Cf. AG τίθημι (Beekes, 2010, p. 1482-1483; Rix et al., 2001, p. 136).
[21] Cf. Lithuanian peřšti (Rix et al., 2001, p. 475).

e) Actions in the remote or mythological past do away with the augment;
f) General truths and proverbs (the gnomic aorist) use the augment.

These criteria can only be applied to ενεπαρκες, since the reading and meaning of the inscriptions to which the other aorists belong are too unclear. The syntactic factors are not applicable to TM 932138. Thus, only the semantic criteria can be taken into account.

Πουκρος Μανισου ενεπαρκες δετουν. (TM 932138, line 4b-5)

Πουκρ-ος Μανισ-ου εν⟨ε⟩παρκ-ες δετ-ουν.

Poukros-NOM.SG Manisos-GEN.SG ⟨AUG⟩engrave-AOR.3SG monument-ACC

Poukros, son of Manisos, chiselled the monument.

The preceding lines (1-4a) tell us that the tomb to which the inscription belongs was set up for someone's brother. Criteria a) and d) seem most relevant, since the mention of who carved the funerary monument is both foreground information and denotes a past action with relevance to the moment of carving.

Alternatively, it is possible to calculate the probability of the 6 aorists occurring with an augment using the binomial distribution formula $P(X = k) = \frac{n!}{k!(n-k)!} \times p^k \times (1-p)^{n-k}$. For this, we must assume that the presence of the augment in these forms is as likely as its absence ($p = 0.5$) and that the number of trials and number of successes are equal to the aorists attested in the corpus ($n = k = 6$). Thus, we find the chance is 1 in 64 that the 6 attested aorists are augmented. This is statistically significant against the threshold 0.05 for our p-value (0.0156), suggesting that the 6 contexts in which the verbs appear favour the use of the augment, though how is not immediately obvious.

In any case, a tentative pattern is observed to describe the morphology of the aorist: preverb – augment – aorist stem – personal ending. Cf. the table below for a morphological breakdown (Tab. 4.8).

Preverb	Augment	Aorist stem	Personal ending	Description	Token
	ε-	-γδα-	-ες	act.ind.aor.3.sg.	εγδαες
	ε-	-ναρκ-	-ε⟨ς⟩	act.ind.aor.3.sg.	εναρκε
ποσ-	-ε-	-καν-	-ες	act.ind.aor.3.sg.	ποσεκανες
	ε-	-δα-	-ες	act.ind.aor.3.sg.	εδαες
	ε-	-γερ-	-{ε}τοι	mid.ind.aor.3.sg.	εγερετοι
εν-	-ε-	-παρκ-	-ες	act.ind.aor.3.sg.	ενεπαρκες

Tab. 4.8: The six aorists broken down into morphological slots.

Note that in εναρκε, the ending is considered misspelled and should yield *εναρκες (Obrador-Cursach, 2020, pp. 227). Similarly, the third ε in εγερετοι may be excessive, as it could be equated to OPhr. *egertoy* (W-01c) (Obrador-Cursach, 2020, pp. 105, 217-218).

The remaining aorist ουρνουσαεν is thus an outlier:

εινεα̣‿μδους, ξευνε πειρ αρεοπ αδεν⟨π⟩ατω κυ̣ο̣υμαν [ο]υ̣εκρω δαδωνει ονουεια̣ τε̣ αδ̣ενπα̣της δεουε̣ιας μα̣νεις δαδων κε οκκαυγοι ουρν̣ουσαεν. (TM 844853)

‘?’

It is a very unclear word with no known exact meaning: its mood and voice are unknown as well as its segmentation (e.g. Hämmig, 2019, p. 289 reads ουρ[...]υσαε(?)ν). The ending -εν points to a third person plural aorist, since it reflects the PIE secondary 3.pl. ending **-(e)nt* after a word-final reduction of consonant clusters (Fortson, 2010, p. 92; Obrador-Cursach, 2020, p. 324; Tichy, 2009, p. 90). However, it also lacks an apparent augment ε- and due to its obscure status, it is impossible to say that it does or does not feature one. For it is on the one hand possible that the augment is absorbed into a lengthened neighbouring vowel, as the ‘temporal augment’ does in AG (e.g. ἦγον < ἄγω). On the other hand, its mood is unknown, so it could be a non-indicative form that simply does not have an augment. Alternatively, if it is an indicative, the data are simply too few to conclusively say that the Phr. indicative aorist must have the augment.

It is difficult to analyse the morphology of the aorist stems without delving too far back to PIE. As far as the tokens above are concerned, they seem to belong to the same formation type, for which several possible origins have been posited (cf. Obrador-Cursach, 2020, p. 104 for a brief discussion). As mentioned above, the augment goes back to PIE **(h₁)é* ‘back then’, originally an adverb that got appended onto the verb (Tichy, 2009, pp. 125-126). The ending -ες results from pre-Phr. **-es-t* by final cluster reduction in polysyllabic words (Ligorio & Lubotsky, 2018, p. 1827; Orel, 1997, p. 384; also seen in ουρνουσαεν < PIE **-(e)nt*). The middle ending -τοι appears to be an older variant of -τορ (Ligorio & Lubotsky, 2018, pp. 1827-1828, Obrador-Cursach, 2020, p. 100).

4.3 Basic syntax

Whenever the literature discusses Phr. syntax, it is consistently, and understandably, of limited length (Brixhe, 2008, pp. 77-78; Ligorio & Lubotsky, 2018, pp. 1828-1830; Obrador-Cursach, 2020, pp. 114-118; Orel, 1997, pp. 402-407). Because of this, a handful of points are repeatedly brought up, on which I will briefly comment with data from the corpus.

Phr.’s basic word order is typologized as SOV (Ligorio & Lubotsky, 2018, p. 1828; Obrador-Cursach, 2020, p. 116). However, the data show 7 clear instances where the verb precedes a nominal subject: αδδακετ ορουαν (TM 867450) ‘the protector does/makes’, δακαρεν πατερης (TM 877307) ‘the parents did/made’, δεδασσιννι πατερης (TM 984519) ‘the parents placed’, (με) τοτοσσειτι Βας (TM 867154) ‘let Bas (not) give’, (με) τοτοσσειτι Σαρναν (TM

984531) 'let (Bas?) Sarnan (not) give', εναρκε Ερμωλαος (TM 984535) 'Hermolaos *X-ed*', τοτοσσειτι Βας (TM 942649) 'let Bas give'. However, the pronouns ιος '(he) who', τος 'that' (correlative with ιος; cf. §4.4), κος 'someone', τις 'whoever' and αυτος 'the very one' appear to always precede their verbs.

The indirect object (iobj) is attested already in OPhr. to precede the direct object (Ligorio & Lubotsky, 2018, p. 1828). In the data, however, if two iobj's are conjoined, the second iobj will follow the verb, including if it is expressed as a prepositional phrase: σεμον κνουμανε (first iobj) κακεν αδδακετ **αινι μανκα** (TM 984524) 'harms this tomb **or stele**', σεμουν κνουμανει κακουν αββερετοι **αινι ατεαμα** (TM 984526) 'bears ill to this tomb **or (part of) the monument**', σεμουν κνουμανει κακιν αδδακετ **αιν αδ τεαμας** (TM 984527) 'harms this tomb **or (part of) the monument**', etc. Such instances of an adjunct following the verb is notable considering the position of adjuncts in the main clause of the curse formula, which consistently precedes the verb. Namely, if Ti- is mentioned as the agent of τιττετικμενος ειτου 'let him be(come) accursed', Τιε precedes the participle, just like με ζεμελως κε δεως κε 'in sight of both men and gods'. There is, however, an attestation of two iobj's following the verb in the protasis: ιος κακον αββερετορ **κνουμανει αινι [---] […]** (TM 845532) 'he who does harm to the tomb or…'.

Phr. features a variety of prepositions, some of which are present in the data: *ad* + dat. 'to, ad, by', *as* + acc.(/dat.) 'in, at, by', εν + acc.(/dat.) 'in', με$_1$ + dat. 'before, in the sight of, among',[22] *por* + dat. 'for', *pos* + acc. '?'. There is also the twice attested postposition αργου + acc. 'on account of' (cf. §4.1). The etymology of αργου is unclear, but it is assumed to be derived from a noun in the gen. This is an interesting notion considering gen. nouns otherwise tend to precede their head noun.

The corpus shows a few NPhr. enclitics νι (whose exact meaning is not entirely clear; occurring 68 times; cf. §4.4), κε 'and' (< PIE **=k^{w}e*; cf. AG τὲ 'and'; occurring 71 times; cf. §4.4) and του/τι from the pronoun τος 'that' (occurring only 10 times). The former two can appear suffixed as the conjunctions αινι 'or' and ακκε 'and'. In that case they are no longer postpositive and precede whatever phrase they connect to what has been previously stated. As seen above, αινι, which occurs 26 times, can introduce a second indirect object. There are only 4 attestations of ακκε, all of which introduce a second main clause to the curse formula:

ακ⟨κ⟩ε οι ειροι α τιε τιττετικμενοι ειννου (TM 931697)

ακ⟨κ⟩ε οι ειροι α τιε τιττετικμενοι ειννου

and him:DAT heroes by Ti- accursed become

'and let his heroes(?) be accursed by Ti-'

and the rest

ακκε οι βεκος ακκαλος τιδρεγρουν ειτου (TM 867135, TM 845586 and TM 867070)

ακκε οι βεκος ακκαλος τιδρεγρουν ειτου

[22] As opposed to με$_2$, which is the prohibitive particle.

and him:DAT bread bread/food? unenjoyable become

'and let bread/food be unenjoyable to him'

The difference between κε and ακκε is paralleled in Latin by *-que* and *atque* respectively. Namely, the former is enclitic and thus follows the first word of the conjunct, whereas the latter introduces the conjunct in its entirety. Concerning the enclitic forms of τος, they are used seemingly interchangeably and exclusively following the dative demonstratives σεμουν, σα and σας: ιος νι **σ[εμ]ον τ[ου]** κνουμαν[ε] κακουν τι αδ[δ]ακετ (TM 984542) 'whoever does any harm **to this** tomb', ιος **σα τι** σκελεδριαι κακουν δακετ (TM 932037) 'who does harm **to this** *skeledria*', etc. Both του and τι are singular neuter forms, the former is dative and the latter nom.-acc., and their function is exclusively to reinforce the preceding pronoun (Obrador-Cursach, 2020, p. 366).

4.4 Lemmas

The data can also be used to examine individual words in NPhr. Excluding punctuation, only 10 lemmas are each attested more than 50 times.

Lemma	POS	Attestations
ios	PRON	105
σεμουν	DET	102
ituv	VERB	91
κακουν	NOUN	88
κνουμανει	NOUN	87
αδδακετ	VERB	85
τιττετικμενος	VERB	75
ke	CCONJ	73
ni	PART	68
τιε	PROPN	62

Table 4.9: The 10 most attested lemmas in NPhr.

All appear in the curse formula (cf. §4.1) except for *ke* when, for instance, it is not used in the adjunct με ζεμελως (κε) δεως κε 'in sight of (both) men and gods'. It is thus clear that the formula is prominent in the extant NPhr. inscriptions.

As mentioned in §4.3, τος 'that' (the lemma is attested 32 times) is correlative with ιος 'who'. This is backed up by the data, as this token refers back to a preceding relative clause introduced by ιος the 8 times it occurs. The remaining 24 attestations of the lemma are inflected differently and thus serve different functions, such as the enclitic function in §4.3.

There are 15 attestations of the lemma αββερετ ‘s/he bear’, 12 of which substitute αδδακετ in the protasis of the formula. There is a soft distinction in the data between the lemma appearing with the preverb and without. When it is present (13 attestations), it serves as the substitute; [ααβ]ęρετορ (TM 845568|5|7) is the exception to this, as it is present in the apodosis, of which the rest is obscure, however. The forms sporting the preverb are the only ones of this lemma to show variation in voice: active (αββερετ) versus mediopassive voice (αββερετοι/αββερετορ). There is no immediately apparent semantic difference between the two. When the preverb is absent, thus yielding βερετ (2 attestations: TM 919096|2|9, TM 984528|1|14), it is used in the phrase Βας ιοι βεκος με **βερετ** ‘let Bas not give him bread’, an alternative apodosis to the formula.

5. The annotated data as a whetstone

In this thesis, I have introduced the annotated corpus for Phr. and attempted to highlight its capabilities through comparison with established research. I have provided a basic overview of our current knowledge on the language (§1) and highlighted the consensus, both in general and in specialised literature, that its closest relation is with AG (§2). This relationship is important to understanding why the newly created UD conventions for Phr. are based on those for AG. Following a description of the annotation process (§3), a sample of the corpus' potential was presented, focusing on its application to basic morphology and syntax as well as the lexicon (§4). The results show that while descriptions of Phr. grammar generally do seem to hold up, there is plenty of deviation in such a limited corpus ripe for further inquiry. For instance, the gen. κνουμινος (TM 931700|2|4) is used as iobj where the dat. κνουμανει is expected, prompting a re-evaluation of the basic Phr. case system. Similarly, there seems to be a motivating factor to the presence of the augment in 6 of the 7 attested aorists. Certain criteria applicable to Homeric Greek may serve as recourse, but are insufficient.

The annotated data have thus showcased their usefulness. Firstly, it is evident that existing knowledge of Phr. can and should be refined, for which this corpus can serve as the required whetstone. Additionally, the corpus follows UD guidelines, thus aiding in developing them into a common standard. Since the data are also intended to be publicly accessible on GitHub,[23] there will be greater availability to participate in research on Phr. Furthermore, the corpus can be seen as the NPhr. counterpart to the TITUS database,[24] which preserves Brixhe and Lejeune's edition of OPhr. (1984). Lastly, the corpus paves the way for future research. Researchers are now able to both approach NPhr. computationally as well as base forthcoming resources on numerical data. Since errors may have slipped past me during annotation, there is of course room for improvement by correcting or editing this corpus, and there is opportunity for expansion whenever OPhr. or MPhr. be annotated as well. It might also be of interest to compare these data with other fragmentary corpora in order to expand on the research of fragmentary languages in general. These data, in short, thus serve to refine research more broadly.

[23] https://github.com/OggiPeeters/UD_Phrygian-KUL.
[24] https://titus.fkidg1.uni-frankfurt.de/texte/etcs/phrygian/phryg.htm

Reference list

Allen, W. S. (1978). *Vox Latina* (2nd edition). Cambridge University Press.

Anthony, D. W. (2007). *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppe Shaped the Modern World*. Princeton University Press.

Beekes, R. S. P. (2010). *Etymological Dictionary of Greek*. Brill.

Beekes, R. S. P. (2011). *Comparative Indo-European Linguistics: An Introduction* (2nd edition). John Benjamins Publishing Company.

Brixhe, C. (1997). Le phrygien. In F. Bader (Ed.), *Langues indo-européennes* (pp. 167-180). CNRS Éditions.

Brixhe, C. (2008). Phrygian. In R. D. Woodard (Ed.), *The Ancient Languages of Asia Minor* (pp. 69-80). Cambridge University Press.

Brixhe, C., & Lejeune, M. (1984). *Corpus des inscriptions paléo-phrygiennes*. Éditions Recherche sur les Civilisations.

Clackson, J. (2007). *Indo-European Linguistics: An Introduction*. Cambridge University Press.

Clackson, J. (2022). Methodology in Linguistic Subgrouping. In T. Olander (Ed.), *The Indo-European Language Family: A Phylogenetic Perspective* (pp. 18-32). Cambridge University Press.

De Decker, F. (2022). The use and absence of the augment in the forms ἔδωκ(ε)(ν) and δῶκ(ε)(ν) in the *Iliad*, compared and contrasted with the overall data of *Iliad* and *Odyssey*. *Acta linguistica Petropolitana*, *18*(1), 78-113.

Depauw, M., & Gheldof, T. (2014). Trismegistos. An interdisciplinary Platform for Ancient World Texts and Related Information. In Ł. Bolikowski, V. Casarosa, P. Goodale, N. Houssos, P. Manghi, J. Schirrwagen (Eds.), *Theory and Practice of Digital Libraries - TPDL 2013 Selected Workshops (Communications in Computer and Information Science 416)* (pp. 40–52).

Diakonoff., I. M. (1984). *The Pre-history of the Armenian People*. Caravan Books.

Diakonoff, I. M., & Neroznak, V. P. (1985). *Phrygian*. Caravan Books.

Drew-Bear, T., Lubotsky, A., & Üyümez, M. (2008). Three New Phrygian Inscriptions. *Kadmos*, *47*, 109-116.

Drews, R. (1993). Myths of Midas and the Phrygian Migration from Europe. *Klio*, *75*, 9-26.

Fortson, B. W. IV. (2010). *Indo-European Language and Culture. An Introduction* (2nd edition). Wiley-Blackwell.

Fritz, M., & Meier-Brügger, M. (2021). *Indogermanische Sprachwissenschaft* (10th edition). De Gruyter.

Gorbachov, Y. (2005). The Origin of the Phrygian Aorist of the Type *edaes*. In K. Jones-Bley, M. E. Huld, A. Della Volpe & M. Robbins Dexter (Eds.), *Proceedings of the Sixteenth Annual UCLA Indo-European Conference* (pp. 191-217). Institute for the Study of Man.

Haas, O. (1966). *Die phrygische Sprachdenkmäler*. Académie bulgare des sciences.

Hämmig, A. E. (2013). Nevotan niptiyan: Die Fluchformel der Stele von Vezirhan. *Indogermanische Forschungen*, *118*, 125-154.

Hämmig, A. E. (2017). What if Phrygian were the closest attested cognate of Armenian?. In A. Bayandur, & V. Kazaryan (Eds.), *XI International Conference on Armenian Linguistics. Dedicated to John A. C. Greppin (1937-2016)* (pp. 54-55). Areg Bayandur.

Hämmig, A. E. (2019). Deciphering Phrygian: Blind Alleys and Viable Ways. In G. R. Tsetskhladze (Ed.), *Phrygia in Antiquity: From the Bronze Age to the Byzantine Period* (pp. 287-304). Peeters.

Hämmig, A. E. (2022). A "new" Neo-Phrygian curse formula. In B. Obrador-Cursach & I.-X. Adiego (Eds.), *Phrygian linguistics and epigraphy: new insights* (pp. 89-102). Universitat de Barcelona.

Hyllested, A., & Joseph, B. D. (2022). Albanian. In T. Olander (Ed.), *The Indo-European Language Family: A Phylogenetic Perspective* (pp. 223-245). Cambridge University Press.

Keersmaekers, A., Mercelis, W., Swaelens, C., & Van Hal, T. (2019). Creating, enriching and valorising treebanks of Ancient Greek: the ongoing Pedalion-project. In M. Candito, K. Evang, S. Oepen, D. Seddah (Eds.), *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories* (pp. 109-117).

Kuhrt, A. (1998). *The Ancient Near East: c. 3000-330 BCE*, Vol. 2. Routledge.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Blackwell Publishers.

Ligorio, O., & Lubotsky, A. (2018). Phrygian. In J. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics* (pp. 1816-1831). De Gruyter Mouton.

Lubotsky, A. (2004). The Phrygian Zeus and the problem of Lautverschiebung. *Historische Sprachforschung*, *117*, 229-237.

Mallory, J. P. (1989). *In Search of the Indo-Europeans*. Thames and Hudson.

Mallory, J. P., & Adams, D. Q. (2006). *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford University Press.

Martirosyan, H. (2014). *Origins and historical development of the Armenian language*. Leiden University.

Mayrhofer, M. (2004). *Die Hauptprobleme der indogermanischen Lautlehre seit Bechtel*. Österreichischen Akademie der Wissenschaften.

Obrador-Cursach, B. (2019). On the place of Phrygian among the Indo-European languages. *Journal of Language Relationship*, *17*(3), 233-245.

Obrador-Cursach, B. (2020). *The Phrygian Language*. Brill.

Obrador-Cursach, B., & Adiego, I.-X. (2022). Foreword. In B. Obrador-Cursach & I.-X. Adiego (Eds.), *Phrygian linguistics and epigraphy: new insights* (pp. 9-12). Universitat de Barcelona.

Olsen, B. A., & Thorsø, R. (2022). Armenian. In T. Olander (Ed.), *The Indo-European Language Family: A Phylogenetic Perspective* (pp. 202-222). Cambridge University Press.

Orel, V. (1997). *The Language of the Phrygians: Description and Analysis*. Caravan Books.

Pitts, R. J. (2018). *Tense, Aspect and Modality in the Sabellic Languages* [Masters' thesis, KU Leuven].

Rix, H., Kümmel, M., Zehnder, T., Lipp, R., & Schirmer, B. (2001). *Lexikon der indogermanischen Verben: Die Wurzeln und ihre Primärstammbildungen* (2nd edition). Dr. Ludwig Reichert Verlag.

Roller, L. E. (1983). The Legend of Midas. *Classical Antiquity*, *2*(2), 299-313.

Stefanowitsch, A. (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistics Theory*, *2*(1), 61-77.

Tichy, E. (2009). *Indogermanistisches Grundwissen: für Studierende sprachwissenschaftlicher Disziplinen* (3rd edition). Hempen Verlag.

Van Emde Boas, E., Rijksbaron, A., Huitink, L., & de Bakker, M. (2019). *Cambridge Grammar of Classical Greek*. Cambridge University Press.

Woudhuizen, R. (2009). An overview of research on Phrygian from the nineteenth century to the present day. *Studia Linguistica*, *126*, 167-188.