

# Wh-questions in Tocharian and implications for PIE reconstruction\*

Ryan Hearn  
April 24, 2017

## 1 Introduction

The goal of this project is to present a cohesive syntactic analysis of the *in-situ* behavior of wh-question words in Tocharian, as a first step toward reconstructing the state of wh-question syntax in Proto-Indo-European (PIE). The two well-attested Tocharian languages, Tocharian A and B, are sister SOV languages attested in northwestern China along the northern edge of the Tarim Basin from the 6th to 10th centuries CE. Tocharian is widely believed to be the second branch of Indo-European (IE) to diverge from the proto-language, which makes Tocharian languages invaluable for PIE reconstruction, as they can be directly compared with Anatolian<sup>1</sup> and the reconstructed ancestor of the remaining IE languages to produce the earliest possible IE proto-language.

The ancient Indo-European languages have long been considered strong wh-movement languages (Fortson 2004), in which wh-elements obligatorily raise to the highest spec-CP position in the clause. Recently, however, it has become evident that at least Anatolian, and possibly also Tocharian, display syntactic behavior that appears to be wh-*in-situ*, even taking into account IE languages' propensity for free word order (Pinault 1997; Adams 2015; Hoffner Jr 1995). Compare (1) and (2) below, from Tocharian B<sup>2</sup> and Hittite, respectively.

- (1) *somo-añyai*      *somo ytārye k̄ā*      *westār*  
single-traversable single road      **why** called

“**Why** is the single road called the only-traversable?” (29b1<sup>C</sup>) Adams (2015)

- (2) *šummeš=kan*      *kui-t*      *ney-ari*  
you.DAT.PL=LOC what-NOM.SG.N happen-3SG.PRS.MED

“**What** will happen to you?” (NH/NS (CTH 89.A) KUB 21.29(+) rev. iv 13-14) (Sideltsev 2014)

§2 presents the Tocharian wh-*in-situ* data for both wh-questions and relative clauses to gain a more complete picture of Tocharian wh-element behavior. §3 presents similar data and prior analyses

---

\*Thanks to my committee, Michael Weiss, John Whitman, and Miloje Despic, for their invaluable advice. Thanks also to my fellow research workshop classmates for their input and patience through many presentations and draft revisions.

<sup>1</sup>Assuming that Anatolian was in fact the first daughter language to diverge from PIE, as per Jasanoff (2003) and Melchert (2013).

<sup>2</sup>In this paper I will be following the dating system for Tocharian B used by Adams, in which a superscript capital letter after the manuscript identifier indicates the archaism of the text, after Peyrot (2008). A indicates “archaic”, E indicates “early”, C “classical”, L “late”, and Col “colloquial late”. Most of the examples from Tocharian B in the corpus are from the classical era, but this is likely due to the large number of documents dating from that period. Tocharian A also shows *in-situ* behavior, as seen in the Appendix, so I believe that it can be reconstructed for the proto-language.

from Hittite for comparison/contrastive purposes, and to inform PIE reconstruction. §4 gauges the extent to which the Hittite analyses are applicable to the Tocharian data and presents the support for my own conclusion that Tocharian does in fact show *wh-in-situ* behavior, and that, in the absence of evidence to the contrary, these *wh*-question words likely remain in their base-generated positions in the clause. §5 discusses the implications of the *in-situ* behavior of Hittite and Tocharian, vis-a-vis Hale (1987), for the reconstruction of *wh*-question syntax in PIE.

## 2 *Wh-in-situ* in Tocharian

### 2.1 *Wh*-questions

Adams (2015) provides many of the examples of *wh*-questions present in the Tocharian B corpus. He follows traditional IE syntactic thought in claiming that “like most Indo-European languages, Tocharian B has a rule like English *wh*-fronting whereby the interrogative pronoun is moved to the head of its clause, whatever its normal place might have been” (p. 47). He then notes that only for the interrogative pronouns *k<sub>u</sub>se* ‘who, what’, *k<sub>u</sub>ce* ‘whom, what’, *ket(e)* ‘whose’ and *k<sub>u</sub>tameṇ* *k<sub>u</sub>tameṇ* ‘where, whence’ is this rule almost exceptionless. The one exception of this subgeneralization he notes is the following:

- (3) *śraddhauñe k<sub>u</sub>se ste*  
 faith            **what** is  
 “**What** is faith?” (23a5<sup>C</sup>) Adams (2015)

For all other question words, however, including *kā* ‘why’, the *m*-words *mäkte* ‘how’ and *mäksu* ‘who, what’, and *intsu* ‘which’, Adams notes that exceptions to *wh*-initial word order in *wh*-questions are so numerous, as high as 25-30% of the attested corpus, “that one might wonder whether these words are subject to *wh*-fronting at all” (p. 48). It seems to me that positing *wh-in-situ* as the default across the board might be preferable to positing, as Adams does, *wh*-movement for all *wh*-words, but then pointing out that *kā* ‘why’ and all other non-*k*-initial question words all act exceptionally, especially when the *k*-initial *wh*-words that should obligatorily move also show at least one *in-situ* exception.

Here I shall briefly list the examples from Adams (2015) showing non-initial behavior<sup>3</sup>. Examples of *kā* ‘why’:

- (4) *somo-añyai somo ytārye kā westār*  
 single-traversable single road    **why** called  
 “**Why** is the single road called the only-traversable?” (29b1<sup>C</sup>) Adams (2015)
- (5) *sū kā swāsaṇ*  
 it    **why** rain  
 “**Why** will it rain?” (140b4) (Adams 2015)

Examples of *mäkte* ‘how’:

- (6) *tumeṇ no stamäṣṣälñe mäkte yentets*  
 therefrom but establishment    **how**    wind.GEN.PL  
 “But therefrom **how** is the establishment of the winds?” (41b5<sup>C</sup>) Adams (2015)

---

<sup>3</sup>Additional examples from the CEToM corpus will be discussed later in the paper.

- (7) *y(e)s mäk̄te maś(c)e(r)*  
 you.PL **how** be  
 “**How** will you act?” (108a5<sup>L</sup>) Adams (2015)

Examples of *intsu* ‘which’

- (8) *posa śpālmeṃ rṣāke intsū ste*  
 all.PERL best rishi **which** is  
 “**Which** seer is best?” (107b3<sup>L</sup>) Adams (2015)

As mentioned above, Adams claims that wh-words with initial *k*-, except *kā* ‘why’, show obligatory movement, and that all others show optional *in-situ* behavior. Based off the data Adams presents, I originally believed that a different generalization could be made: only arguments of a (non-copular) verb obligatorily raise. The following example from the CEToM corpus, however, seems to strongly indicate that all question words in Tocharian, including *k*-initial words, may show *in-situ* behavior:

- (9) *tusa yweru nuwaṃ k<sub>u</sub>se*  
 thus.PERL.SG with.rage shout.3SG.SUBJ.ACT **who.N.SG**  
 “Who roars with more rage than that?” (PK AS 7M.a4<sup>C</sup>, CEToM)

## 2.2 Wh-relatives

Adams (2015) also provides a section on the behavior of wh-relatives. As this subject is slightly orthogonal to the aim of the current paper, it will be discussed only briefly.

Adams notes, citing Pinault (1997), that relative pronouns almost universally occur in clause-initial position in prose, but that poetry shows an alarmingly large number of exceptions to this generalization, to the extent that “the failure of relative pronoun movement [...] is not just permitted in poetry but rather is a **marker** of poetry” (Adams 2015: 29). Pinault goes so far as to claim that clause-initial position of wh-relatives is due to the influence of Buddhist Hybrid Sanskrit (BHS), and notes that 6% of relative clauses in Tocharian A showed *in-situ* behavior. Adams claims that the numbers for Tocharian B are likely similar.

Also, it is noteworthy that *k*-words (which according to Adams normally do show obligatory wh-movement) and other wh-question words (which he claims do not) can both act as wh-relatives. (10) below demonstrates this *in-situ* behavior of Tocharian relatives in poetry. Braces {} demark the relative clause.

- (10) *kleśanmaṣṣi sta(na {no bha}wakārne kem̄t witsko}*  
 afflicted trees {then highest.existence-LOC **whose** root}  
 “*kleśa*-trees whose root [is] in the highest existence” (554a3/4<sup>E</sup>) Adams (2015)

Poetry is known to preserve archaisms, so chances are good that these *in-situ* relative pronouns reflect the old state of affairs. I would not go as far as Pinault and claim that all relative clauses prior to contact with BHS had non-initial *in-situ* relative pronouns, but I do think it likely that *in-situ* relative pronoun behavior in Tocharian is old. Hopefully future work will shed more light on the precise nature of these relative constructions, but for our purposes it is sufficient to note that systemic *in-situ* behavior extends beyond the wh-question domain.

In the next section we will take a look at the IE language whose *in-situ* behavior has seen the most study recently, to see if any of these generative analyses translate well to our Tocharian data.

### 3 Wh-*in-situ* in Hittite

Hittite is an SOV language of the Anatolian family, attested from the 16th to the 13th centuries BCE, widely believed to be the first branch of Indo-European to diverge from the proto-language, before Tocharian (c.f. Jasanoff (2003) and Melchert (2013) for discussion).

According to the question word corpus in Hoffner Jr (1995), containing an exhaustive list of approximately one hundred wh-questions, wh-question words in Hittite seem to appear in three different locations, the lowest of which is immediately preverbal. As mentioned above, Hittite is an SOV language, with verbs almost always appearing clause-finally. Auxiliary verbs follow main verbs, and in rare sentences the main verb can be followed by certain adverbs, as seen in (11) below.

- (11) *nu kun memiya-n kuwat iya-tten QATAMMA*  
 CONN this.ACC.SG matter-ACC.SG **why** do-2PL.PRET this.way  
 “Why have you treated this matter in the same way?” (KBo III 3 + KUB XXIII 126 + KUB XXXIII 36 (+) 1459/u iii 2-4) (Hoffner Jr 1995)

The wh-words in this low position, however, always show up directly before the verbal complex, as seen in examples like (11) above.

The next-highest position in which we find wh-question words in Hittite is immediately pre-negation, exemplified in (12) below. Negation itself directly precedes the clause-final verbal complex, however, and question words never occur between negation and the verbal complex, so we may be tempted to lump pre-negation and pre-verbal question words into the same syntactic structure, but for now I present pre-negation wh-question words as a separate case for the sake of comprehensiveness.

- (12) *n=an EGIR-pa [k]uwat UL piš-teni*  
 CONN=him back **why** not give-2PL.PRS  
 “Why don’t you give him back?” (HKM 58:21-22) (Hoffner Jr 1995)

The final position in which we find wh-question words is clause-initially, hosting the Hittite second-position clitic chain. These clitic chains, when present, always occur in second position in the clause, and consist of up to six ordered morphemes conveying discourse, argument, and other grammatical information about the clause. They must be hosted initially by either a dummy host *nu* ‘and’ or an argument or other element in the clause. As seen below, wh-question words may act as this host.

- (13) *kui-š=war=an hara-n <sup>D</sup>Pirwa[-i] <sup>URU</sup>Haššuw-aza*  
 who-NOM.SG=QUOT=him eagle-ACC.SG Pirwa-DAT.SG Hassu-ABL  
*uwate-z[zi]*  
 bring-3SG.PRS  
 “**Who** will bring the eagle from the city of Hassu to Pirwa?” ((OH/NS (CTH 337.1.A) KUB 48.99 obv 6-7)) (Sideltsev 2014)

Many diverse analyses of this apparent *in-situ* behavior of wh-elements have been put forward throughout the relatively short history of Hittite syntax. Here I will briefly present three such analyses, from the wh-movement analysis of Garrett (1994) to the base-generated *in-situ* argument of Huggard (2011) and the *in-situ* left-periphery movement analysis of Sideltsev (2014).

### 3.1 Wh-movement - Garrett 1994

Garrett’s work is more concerned with relativization than with wh-questions, but the structures and movements he discusses point out potential landing sites for wh-words in question formations, and have strongly influenced later research into wh-word behavior in Hittite.

He divides Hittite relative clauses into two general categories: preposed and postposed relatives. For preposed relatives he adopts the analysis of Held (1957), who argued that Hittite preposed relatives take universal quantification interpretation. These preposed relative clauses are then further subdivided into determinate and indeterminate categories, with corresponding semantic interpretations, and each with their own distinct syntax. (14) and (15) below show examples of determinate and indeterminate clauses, respectively.

- (14) *namma NINDA.[GUR<sub>4</sub>].RA UD-MI **kuiēš** ēšanzi n=at parkuwaiš*  
 but bread daily who make.3PL CONN=them clean  
*ašandu*  
 be.3PL.IMP  
 “Furthermore, (those) **who** make the daily loaves, let them be clean.” (KUB 13.4 i 14) Garrett (1994)

- (15) ***kuiš**=an=šan EGIR-pa tarnai n=an šakuwanzi*  
**who**=him back allow.3S CONN=him imprison.3PL  
 “**Whoever** lets him back, they will imprison him” (KUB 13.2 iii 16) Garrett (1994)

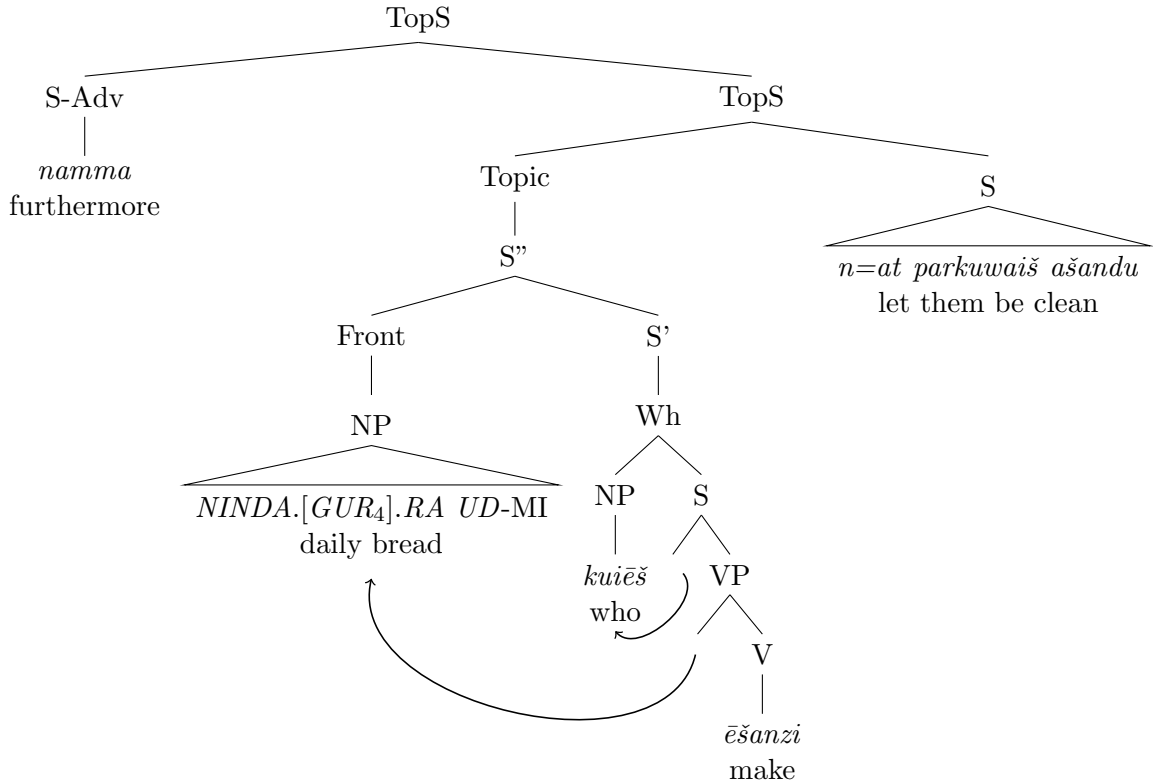
According to Garrett, determinate wh-words are never clause-initial within the relative clause and are normally preceded by only one constituent, while indeterminate wh-words are always clause-initial, an observation borne out by the data above.

Garrett, in accordance with the prevailing theories of the day as expressed in Kiparsky (1995), assumes that all relative clauses, both preposed and postposed, are adjoined to the main clause<sup>4</sup>. Preposed relative clauses occupy a position in the main clause that Garrett calls “Topic”, a syntactic operator position that allows quantificational interpretation of the relative clauses. Within the relative clauses themselves, Garrett assumes an obligatory “Fronting” operation: indeterminate relative wh-words must first raise to Wh-position, then must Front to initial position in the clause, while determinate relative wh-words stay in Wh-position, with another constituent moving to fill the Fronting position. The structures associated with each type are shown below.

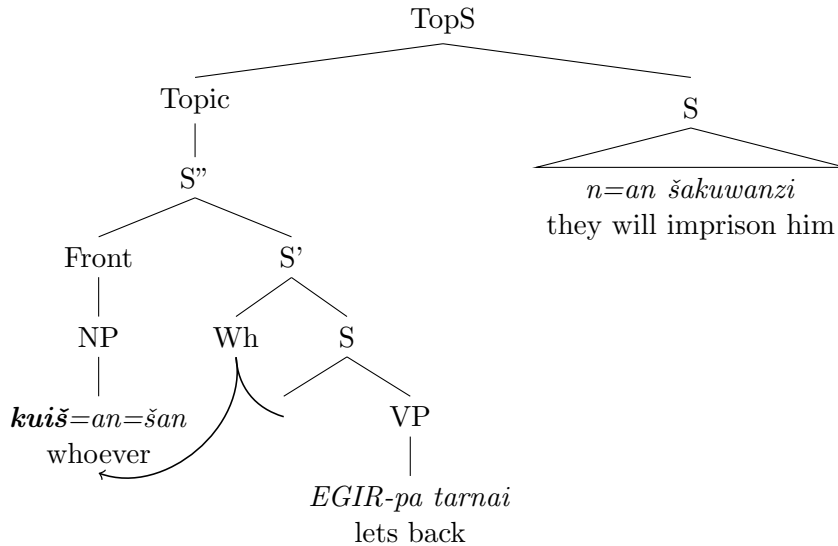
---

<sup>4</sup>Modern research, however, most notably Probert (2014), has shown that there is no reason to assume that Hittite (or Proto-Indo-European for that matter) lacks clausal embedding.

(16) Determinate Relative Clause, from (14) above



(17) Indeterminate Relative Clause, from (15) above



Garrett’s postposed relative clauses, divided into indefinite and nonrestrictive relative clauses, show parallel syntax within the relative clauses to the determinate and indeterminate clauses shown above. Garrett thus proposes obligatory wh-movement to a left-peripheral Wh-position in all relative clause subtypes, followed by further movement to “Front” position depending on the specific subtype in question.

### 3.2 “Wh-non-movement” - Huggard (2011)

Huggard instead presents a head-raising analysis of Hittite relative clauses based on Kayne (1994), within the articulated left periphery of Rizzi (1997).

Huggard states that in a true wh-movement language we would expect wh-elements to fill the highest spec of the expanded CP, and that no element in the clause should precede it. He then cites Goedegebuure (2009)’s survey of one hundred Hittite wh-questions, in which thirty-five of the fifty-four questions with more than two constituents had non-initial wh-question words. Huggard concludes that Hittite must therefore be a wh-*in-situ* language.

Other types of movement, however, may result in a wh-element surfacing in non-*in-situ* position. He again cites Goedegebuure, who states that the location of the wh-element in Hittite questions is dependent upon whether the wh-element is focused, and what type of focus the element has.

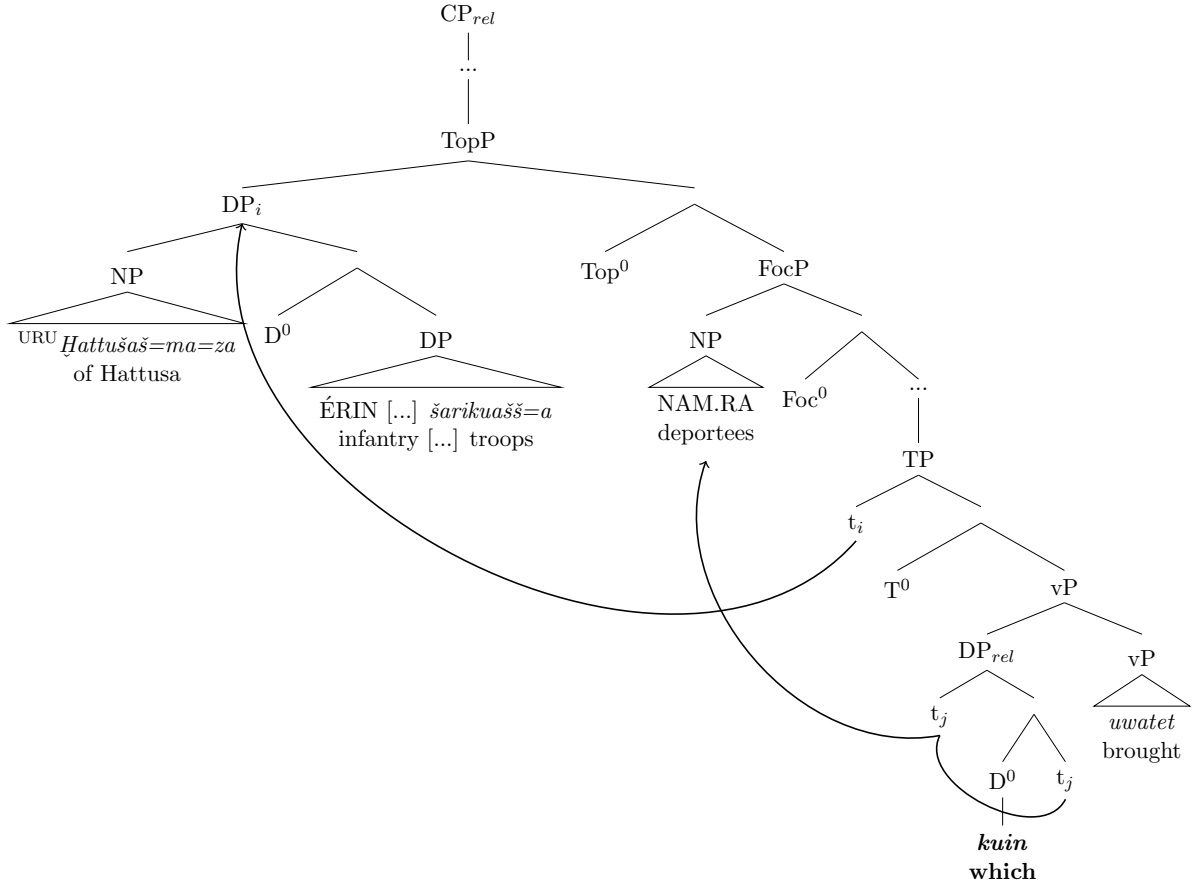
Huggard then points out examples like the correlative constructions in (18) and (19) below, which provide strong evidence for wh-*in-situ* in relative clauses in Hittite as well as in questions. Note that not only is Garrett’s Front position filled in (19), but the Topic position as well.

- (18) *nu=za*                    <sup>d</sup>UTU-ŠI    **kuin**    NAM.RA    INA    É.LUGAL    *uwatenun*  
 CONN=REFL Majesty=my **which** deportees in    Palace    bring-1SG.PRET  
*n=aš*                    1.SIG<sub>7</sub>.LIM 5.ME    NAM.RA    *ēšta*  
 CONN=those 15,000            500    deportee    be-3SG.PRET  
 “The deportees **whom** My Majesty(I) brought into the palace, they were 15,500. (KBo 3.4 ii 3842) Huggard (2011)

- (19) <sup>URU</sup>*Hattušaš=ma=za*                    ÉRIN.MEŠ ANŠE.KUR.RA.MEŠ    ÉRIN.MEŠ *šarikuwašš=a*  
 Hattusa-GEN.SG=but=REFL infantry    cavalry                    sarikuwas.troops=conj  
 NAM.RA **kuin**    *uwatet*                    *nu=ššan*                    *kappūwar*    NU.GÁL    *ēšta*  
 deportees **which** bring-3SG.PRET CONN=of.them counting    none            be-3SG.PRET  
 “The deportees **whom** the infantry and cavalry and *sarkuwas* troops of Hattusa brought, there was no counting of them. (KUB 14.16 iii 1522) Huggard (2011)

Huggard then proposes the following head-raising syntactic account of these *in-situ* relative clauses. In his analysis any constituent, not just the wh-element, may be topicalized or focused according to the requirements of the discourse, yielding the various surface orderings described in Garrett, Goedegebuure, and Huggard.

(20) Wh-*in-situ* relative clause, from (19) above



### 3.3 Movin' on up (to the left periphery) - Sidel'tsev (2014)

Sidel'tsev begins his Hittite analysis by arguing persuasively for right-headedness within the TP domain, *pace* Huggard. Sidel'tsev reaches this conclusion due to the SOV nature of Hittite, combined with the observation that auxiliary verbs in Hittite inevitably occur clause-finally, after the main verb.

To determine the exact syntactic position of wh-elements in Hittite, he first adopts the adverb hierarchy of Cinque (1999). Sidel'tsev argues that the linearization of “low adverbs” like *kiššan* ‘in this way’ and *mekki* ‘much’ point to a syntactic structure in which wh-elements cannot be in their base-generated positions.

- (21) *našma ANA AWAT KUR* <sup>D</sup>*U-tašša* ***kui-t*** *kiššan* EGIR-*an*  
 or to matter land Tarhuntassa **something** as.follows then  
*iyan*  
 do.PRTC.NOM.SG

“Or concerning the problem of the land of Tarhuntassa **something** is stipulated as follows”  
 (NH/INS (CTH 106.A.1) Bo 86/299 rev. iii 2-3) Sidel'tsev (2014)

If these adverbs are adjuncts located within VP, Sidel'tsev argues, then all arguments generated



within VP, including objects and wh-phrases, must then be located at least as high as Spec-vP (effectively becoming a low focus position), and possibly as high as Spec-FocP.

To determine whether the wh-phrase landing site is in Spec-vP or Spec-FocP Sideltsev turns to preverbs, which he believes heads a PrvP phrase dominating either vP or TP, marking the left edge of the vP/TP domain. He then provides evidence like (22) below to show that wh-phrases are higher than preverbs:

- (22) *n=[aš]ta*      <sup>m</sup> *Tarul[iʔ]y[a]šʔ*    *tuzzi-n*      <sup>m</sup> *Zilapiyašš=a*      ÉRIN<sup>MES</sup> GIBIL  
 CONN=LOC Taruliya.GEN.SG army-ACC.SG Zilapiya.GEN.SG=and troops    new  
*maḥḥan* *šarā* *uwat-er*  
 how      up    bring-3PL.PRET  
 “How could they have brought up the army of Taruliya and the new troops of Zilapiya?”  
 (MH/MS (CTH 186<sup>?</sup>) HKM 43 obv. 1’-5’) Sideltsev (2014)

He takes this word order as definitive evidence that all objects/wh-elements must end up outside the vP/TP domain. The one apparent counterexample to this claim, (23) below, in which the preverb occurs before the wh-element, he explains as topicalization of the preverb due to “D-linking to the previous stretch of discourse” (p. 215).

- (23) *zik=wa=kan*      *apūn*      *anda kuwat auš-ta*  
 you=QUOT=LOC that.ACC.SG into why    look-3SG.PRET  
 “Why did you look at that woman?” (MH/NS (CTH 42.A) KBo 5.3+ rev. iii 56’) Sideltsev (2014)

He also cites verb movement as further evidence of the vP external nature of wh-phrases in Hittite. Verbs in Hittite sometimes appear immediately before a low adverb in clause-final position, which Sideltsev interprets as indicative of verb movement, due to his head-final interpretation of Hittite syntax within the TP domain. When the verb raises in this manner, wh-elements still appear immediately before the verb.

- (24) *nu*    *k[ū]n*      *memiyan*      *kuwat iya-tten*      QATAMMA  
 CONN this.ACC.SG matter-ACC.SG why    do-2PL.PRET in.this.way  
 “So, **why** have you handled this matter in this way?” (NH/NS (CTH 63.A) KUB 19.31+ rev. iii 27”-31”) Sideltsev (2014)

He also concludes that wh-question words and wh-relative pronouns target different syntactic positions, as shown by the relative clause example below.

- (25) *apāt=ma*      HUL-*lu*      *uttar*      *iya-t*      *kuiš*  
 that.ACC.SG=but evil.ACC.SG thing.ACC.SG do-3SG.PRET **who**  
 “The one **who** did that evil thing...” (NH/INS (CTH 383) KUB 21.19+ obv. ii 9) Sideltsev (2014)

In what appears to be a parallel example of verb raising, we now see that the relative pronoun appears after the verb, instead of before it like the wh-question word.

Based on this evidence, Sideltsev concludes that Hittite does show wh-*in-situ* behavior, but that wh-question words obligatorily move first to Spec-FocP in the left periphery, and then optionally move further to Spec-ForceP, while relative pronouns merge in Spec-Q(uantifier)P and then have the option of moving further to Spec-TopP or Spec-ForceP.

### 3.4 Summing Up

In this section, we have seen that syntactic analyses of wh-element behavior in Hittite have varied parametrically between wh-movement and wh-*in-situ*, and that even the headedness of Hittite in general has been called into question. Finally, within the *in-situ* analyses themselves, the authors differ on whether subsequent movement of the wh-element into the left periphery is obligatory or optional.

In the following section, we will see how far these analyses account for the Tocharian data, as a starting point for developing our own analysis of wh-question behavior in Tocharian.

## 4 Toward a syntactic analysis of Tocharian wh-questions

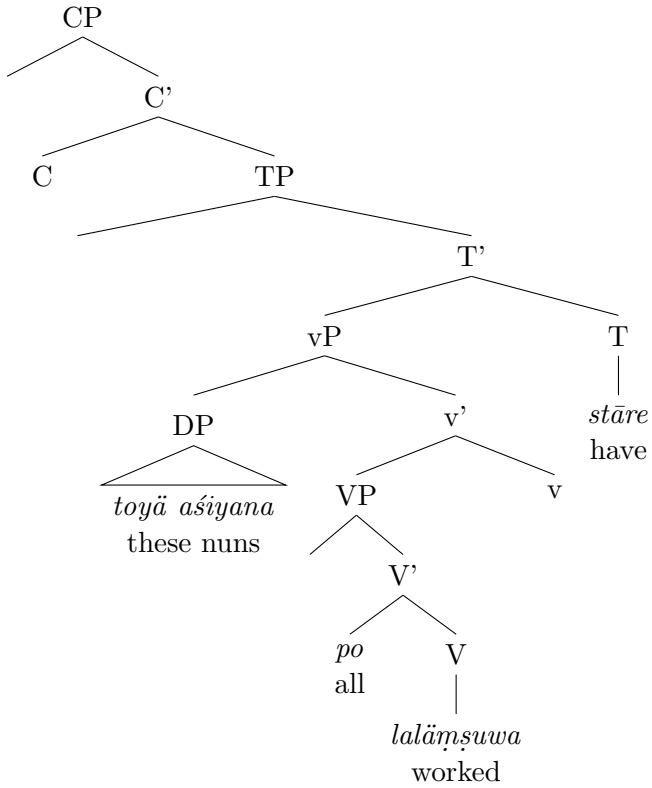
First, let us attempt to determine a few characteristics of the Tocharian syntactic system. Overall word order appears to be staunchly SOV. In poetic texts and literary prose word order is much freer, but Adams (2015), citing a few examples of unliterary business prose texts from Pinault (2008), notes that practically every non-copular sentence is verb-final.

Our next concern is headedness. Taking a cue from Sideltsev (2014), I looked for auxiliary constructions in Adams (2015) to determine whether auxiliaries follow main verbs sentence-finally in neutral clauses. There are a few constructions with modal-esque verbs like *cämp-* ‘be able to’, *yāt-* ‘id.’, and *skāy-* ‘try’ which take infinitives, but their syntax seems closer to that of control verbs than modal/auxiliaries, and would likely be better served by a multi-clause analysis.

Tocharian does, however, have periphrastic perfect, future, necessitive, and potential constructions that should serve our purposes nicely. Each of these are a combination of participle/gerundive + inflected copula, similar to auxiliary constructions in English. Notably, the overwhelming majority of the examples of these constructions cited by Adams place the auxiliary clause-finally, *after* the main verb. In this Tocharian shows surprising similarity with Hittite in overall syntactic structure. As a result, it seems most fitting to classify Tocharian as left-headed above the TP domain, as complementizers and topicalized elements in the sentence appear to the left of the clause proper, and right-headed within the TP domain. Note that this is the same structure as that posited for Hittite by Sideltsev. Here is one example of such an auxiliary construction, and how it might be derived.

- (26) *toyä asíyana po lalāmṣuwa stāre*  
these nuns all worked be.3PL.PRET  
“These nuns have worked everything” (MSL.19.160) Adams (2015)

(27)

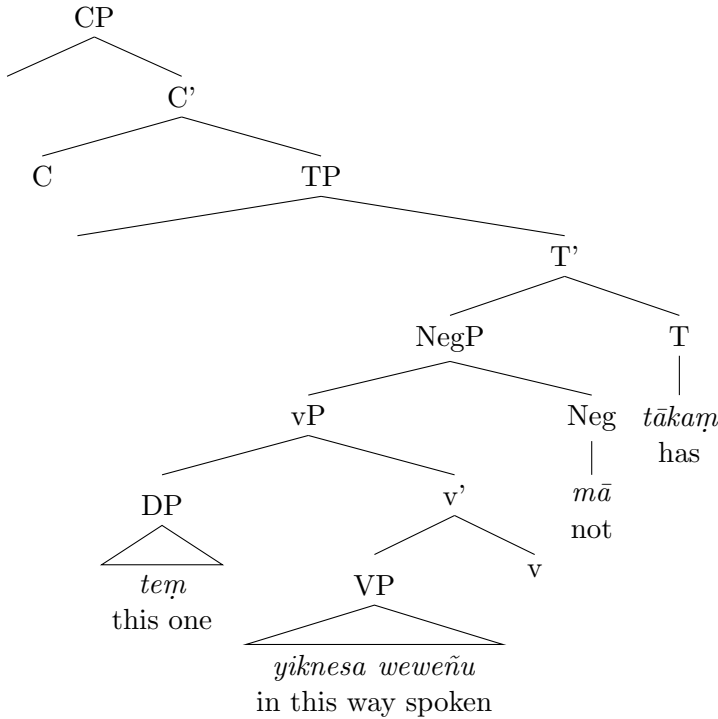


#### 4.1 Negation

Tocharian has both clausal and lexical negation mechanisms. By far the most common clausal negator is *mā*, which in Tocharian B acts both as a simple negator and prohibitive. While the other, much rarer, sentential negators appear only clause-initially, *mā* also appears preverbally as well, implying some sort of movement process when it appears clause-initially. In fact, the specific behavior of this preverbal negator when “*in-situ*” provides further evidence of the right-headedness of the TP domain in Tocharian B. I was able to find one instance of negation interacting with a verbal auxiliary complex in Adams (2015), and the word order is quite telling:

- (28) *teṃ yiknesa weweñu mā tākaṃ*  
this way spoken not be.3SG.SUBJ  
“(If) he has not spoken in this way,” (331b3/4<sup>L</sup>) Adams (2015)

(29)



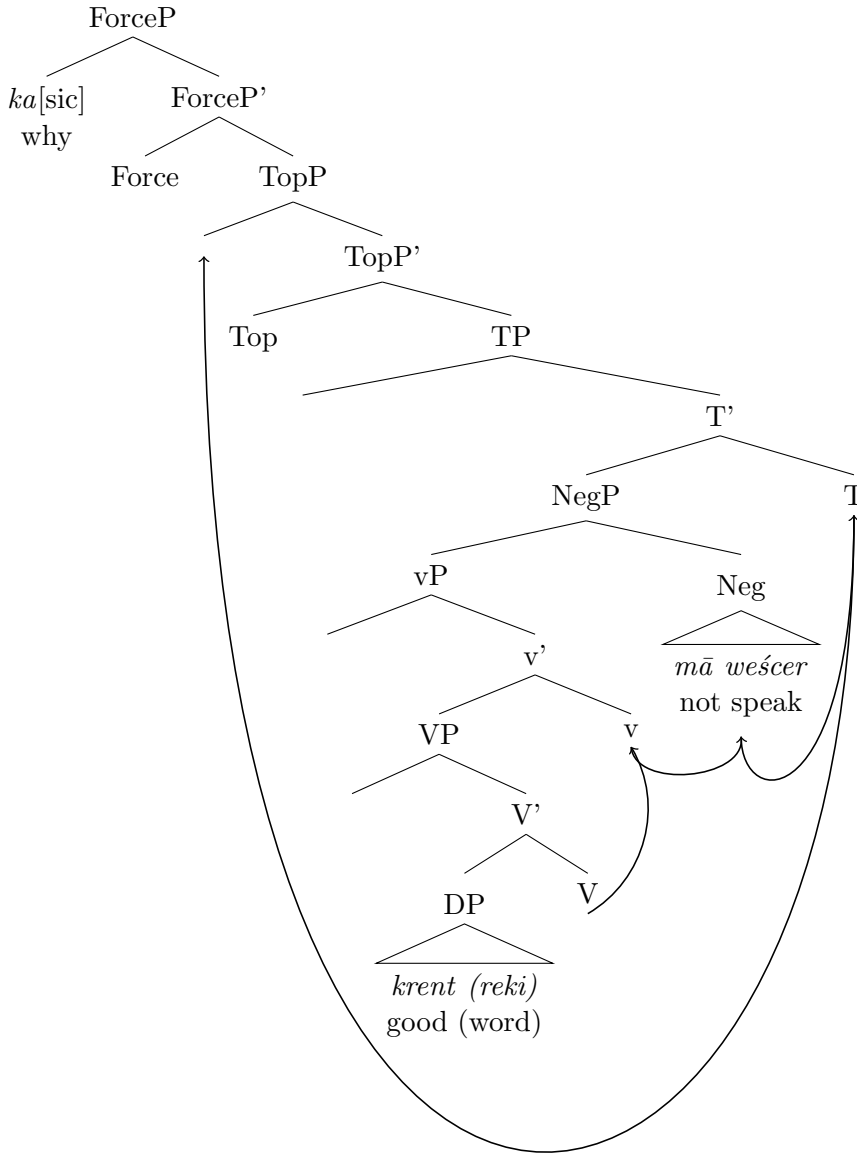
Note how the negation appears precisely between the participle and the copula. With our posited right-headed TP domain, we would expect our right-headed NegP to be located between the TP and vP layers, and that’s exactly where we find it.

Recall, however, that negation also occurs immediately in front of an inflected clause-final verb. If Neg is a head in Tocharian, wouldn’t we expect it to block head-movement to T? To account for this, I argue that inflected verbs in Tocharian move up and merge with Neg, and that the resulting verbal complex then itself moves up to T<sup>5</sup>. And, in fact, we see evidence of negation and the inflected verb acting as a single constituent elsewhere, in sentences like (30) below. Here, adopting the expanded left periphery of Rizzi (1997), the inflected verb has merged with Neg, moved up to T, and the entire complex has then been topicalized below the wh-question word in spec-ForceP.

- (30) *ka[sic] mā wešcer krent (reki)*  
 why not say.you good word  
 “Why do you not say the good (word)?” (20b6<sup>C</sup>) Adams (2015)

<sup>5</sup>Thanks to Miloje Despic for this idea

(31)



As far as negation interacting with *wh*-questions *in-situ*, however, I was not able to find any examples. Either negation would front, or the *wh*-word, or both, as seen in (30) above.

To sum up this subsection, though negation plays an important role in establishing the exact nature of Tocharian right-headedness, and may yield information about the exact syntactic locations of fronted elements, without more data it unfortunately has little to say about the exact surface positioning of *in-situ* *wh*-elements.

## 4.2 Verb-raising

Next let's take a look at verb movement in Tocharian, and see what it tells us about *wh*-question word positioning. As all other constituents in Tocharian seem to be able to do, verbs too can front all the way to the beginning of the clause. Adams (2015) notes that even in Pinault's nonliterary

prose corpus at least two verbs front, though it should be noted that both of them are imperatives.

Adams also mentions a process he calls “right-detachment”, in which, after (Clackson 2007: 167), “appositional phrases and other adjuncts are tacked onto the end of a grammatical sentence”. While this is an apt description of the data, if we could account for this apparent rightward movement in a more traditional leftward manner, we would certainly like to do so. The few examples we see in nonliterary prose involve simple measurement sentences likely of little interest, but in more literary prose SVO and even VSO word order is more common:

- (32) *ysāre rine plyasi wāya Tonke šarmire śak cakaṃnma*  
 grain city.LOC to.sell brought Tonke novice ten *caks*  
 “The novice Tonke brought ten *caks* of grain to sell in the city” (Otani II.12, lines 8/9)  
 Adams (2015)

There is a lot going on in this example. The constituent “grain to sell in the city” has fronted to what is likely the specifier of the highest CP. The subject and object have likely remained within TP, but the verb has nowhere to go from T other than somewhere within the left periphery, perhaps FocP.

As for verb raising interfering with *wh*-question word syntax specifically, Tocharian has a few sentential adverbs that tend to cluster immediately before a clause-final verb. These likely would interact with *in-situ* *wh*-question words in the event of a fronted verb, but unfortunately there aren’t any relevant examples in Adams (2015). A more thorough search of the entire Tocharian corpus may give us the data we need to determine the precise ordering of these elements.

### 4.3 Indirect questions in embedded clauses

Interestingly, according to Adams (2015), embedded questions in Tocharian do show obligatory *wh*-movement to the front of the clause without exception:

- (33) *wālo preksa cī kā nai śintsai mā [św](āsta)*  
 king ask.3SG.PST you why then fodder not eat.2SG.PST  
 “The king asked thee, why didst thou not eat the fodder” (THT-1540a3) Schmidt (2007)

This asymmetry between matrix and embedded clauses is not uncommon, and similar effects are seen elsewhere in Indo-European. One such example is the ‘optional’ *in-situ* behavior of matrix *wh*-questions in modern French, in which intonation alone can indicate a *wh*-question in matrix clauses, but not embedded clauses. Compare (34), (35), and (36) below, from Cheng & Rooryck (2000).

- (34) *Quel livre Jean a-t-il acheté?*  
 which book Jean has-he bought  
 “Which book did Jean buy?”
- (35) *Jean a acheté quoi?*  
 Jean has bought what  
 “What did Jean buy?”
- (36) *\*Je me demande que Jean a acheté quoi?*  
 I REFL wonder that Jean has bought what  
 \*“I wonder what Jean bought.”

The *in-situ* behavior of (35) is accompanied by a rising intonation matching that of yes/no questions in French, whereas the wh-movement example shows a distinct “wh-intonation”. According to Cheng and Rooryck, *in-situ* questions indicate a “presupposed context”; that is, the asker presupposes that the answer to the question is affirmative. Wh-movement questions, on the other hand, have no such presupposition.

To account for the yes/no intonation of wh-*in-situ* questions, as well as for the lack of wh-movement, Cheng and Rooryck propose an underspecified intonation morpheme in overt syntax compatible with both wh-questions and yes/no questions that can check the question feature in C. Thus, either the wh-element may overtly move to spec-C to check the Q-feature, as seen in (34), or the intonation morpheme in C itself may check the feature in (35). The former shows no question intonation due to the absence of the intonation morpheme, but the latter does due to the morpheme’s presence.

A further requirement of this intonation morpheme, in French at least, is that it be a *root morpheme*; that is to say, it may only appear in matrix clauses and only takes matrix scope. This accounts for the ungrammaticality of (36). Both the underspecified and root natures of this intonation morpheme are parameters that may vary cross-linguistically, as seen in Cheng and Rooryck’s Portuguese examples that show *in-situ* behavior even in embedded questions:

- (37) O João perguntou se tu compraste o quê.  
 João asked whether you bought what  
 “João asked what you bought.”

Cheng and Rooryck therefore conclude that the intonation morpheme in Portuguese is, as in French, underspecified, but, unlike French, is not a root morpheme and may therefore take embedded scope.

Their analysis is straightforwardly applicable to our Tocharian (and Hittite) data. In both languages we have seen both *in-situ* behavior and seemingly optional raising to the highest specifier of CP. And, though we have (to my knowledge) no data indicating question intonation in either Tocharian or Hittite, it seems very reasonable to assume that a special intonation to indicate the interrogative nature of *in-situ* wh-questions would have existed, as they would otherwise be indistinguishable from declarative sentences.

And, in Tocharian and Hittite we see the exact same parametric variation Cheng and Rooryck noted in French and Portuguese: Tocharian does not allow wh-*in-situ* in embedded questions (as seen in 33 above), while Hittite does:

- (38) *šumeš=wa* [D]INGIR.MEŠ *UL uškatteni / kiššan=wa=mu* *kuiš iyan ħarzi*  
 you=QUOT gods not see thus=QUOT=1SG.DAT who done has  
 “Don’t you gods see who has done thus to me?” (Hoffner, KUB LIV 1 i 20-21)

So, we see that Tocharian likely has the root variation of the intonation morpheme, like French, while Hittite’s mirrors Portuguese in also taking embedded scope.

Also, as linguists are generally not fans of true optionality in language, it’s important to note that Goedegebuure (2009)’s analysis of Hittite wh-questions demonstrates that semantic and pragmatic factors can account for the variation in question word location. Briefly, she found that: “If the counterexpectancy of the Q-word question is high, the Q-word appears in preverbal position. But if the counterexpectancy of the Q-word question is low or zero, [...] then the Q-word appears in initial position or in pattern position.” (p. 19) A similar analysis of Tocharian is beyond the scope

of the current paper, but I think it likely that a similar, if not exactly the same, generalization can be made.

Goedegebuure’s counterexpectancy findings in Hittite sound surprisingly similar to the “presupposed contexts” analysis of Cheng and Rooryck to explain the apparent optionality of *wh-in-situ* in French. This presence or absence of counterexpectancy in Hittite (and also likely Tocharian) could also therefore trigger the insertion or absence of the intonation morpheme, eliminating the need for optionality in Tocharian and Hittite *wh-in-situ* question syntax.

Finally, Cheng and Rooryck’s analysis of French is but one of many analyses of similar phenomena cross-linguistically, and others may account for the Tocharian data equally well: it is sufficient for our purposes to demonstrate that such analyses which maintain the *in-situ* status of matrix clause *wh*-elements can account for the Tocharian data.

#### 4.4 Conclusion

As it stands, the data considered for this paper point toward a Tocharian *wh*-question word analysis that falls somewhere between that of Huggard (2011) and that of Sideltsev (2014). Auxiliary and negation data supports SOV constituent ordering, exhibiting left-headedness in the CP domain and right-headedness in the TP domain, reminiscent of Sideltsev’s analysis of Hittite. However, without conclusive adverb or verb raising data in Tocharian pointing toward obligatory movement of *wh*-question words into lower left periphery positions, I see no reason to conclude that they move out of their base-generated *in-situ* positions.

If indeed they do stay in their base-generated positions in Tocharian, but obligatorily raise in Hittite, the culprit in Hittite may be its other complex left-periphery constructions, particularly its clitic chains. With so much other pragmatic information taking advantage of the left periphery in Hittite, there may be additional pressure for Hittite to resolve its *wh*-element situations in the left periphery as well. With all of this additional machinery absent from Tocharian, however, it likely has much less cause to force its question words to move all the way up.

## 5 Implications of Tocharian and Hittite for PIE reconstruction

As the first languages to break off from the proto-language, Anatolian and Tocharian have huge implications for the reconstruction of PIE, as each can be directly compared with the reconstructed ancestor of the remaining eight branches to produce the oldest possible stage of the proto-language.

### 5.1 *Wh-in-situ* elsewhere in IE: Hale (1987)

One attempt has already been made to describe the *wh*-question behavior of the ancient Indo-European languages: Mark Hale’s exemplary dissertation. Hale takes an in-depth look at the ancient Indo-Iranian languages with the goal of explaining why *wh*-question words in these languages sometimes occur as the second constituent in their clause. Take, for example, the following line from the Rig Veda:

- (39) *rátham kó nír avartayat*  
chariot who prepared  
“Who prepared the chariot?” (RV 10.135.5b)



Hale concludes that examples like this across the Indo-Iranian languages must be “topicalization around the COMP node”, i.e. that obligatory wh-movement across the board may then be followed by a topicalization process around the wh-word. The occurrence of Wagernackel clitics to the right of the wh-word in these situations further evidences that this topicalization process occurs late in the derivation.

- (40) *índraḥ kím asya sakhyé cakāra*  
 indra what his friendship did  
 “What did Inda<sub>i</sub> do in his<sub>j</sub> friendship?” (RV 6.27.1b)

This argument bears surprising similarity to that of Sideltsev (2014), who also seems to be saying that wh-question words undergo obligatory movement into the left periphery, and that this process may then be followed by a further fronting process of the wh-word itself or another constituent higher in the left periphery. Sideltsev, however, calls this behavior wh-*in-situ* instead of wh-movement due to a stipulation that “wh-phrases which are not demonstrably in the highest Spec,CP are assumed to be *in-situ*”, after Cheng (2009), who defines wh-movement in this manner for theory-internal purposes.

One would assume that, according to this definition, Sideltsev would consider the Indo-Iranian data cited by Hale to be *in-situ* as well, and that Hale would consider Sideltsev’s Hittite data as supporting his wh-movement analysis. In fact, Hale *does* cite Hittite data in his dissertation, though it is relative clause data, and not wh-question data. He concludes that wh-elements in Hittite, like those in Sanskrit, are never preceded by more than a single element.

If this is the solution, and both Sideltsev and Hale (effectively) agree that Indo-Iranian and Hittite are wh-movement languages that then show further topicalization of a constituent, then why have I argued for wh-*in-situ* treatment for Hittite and Tocharian for twenty pages?

First, I believe that Hale has mischaracterized an important piece of the data. Hale follows Held (1957) in assuming that the clause-initial connective particle *nu* is effectively extrasyntactic as far as clausal syntax is concerned, and that initial position in the clause is to the right of *nu*, or *nu* plus the clitic chain. Further, when *nu* is not present, the clitic chain would instead attach lower in the syntactic tree, on the first constituent.

I agree with Sideltsev’s take on this matter: I believe that the syntactic position of the clitic chain is static (whether in ForceP as Sideltsev argues, or somewhere lower in the left periphery), and that constituents lower in the clause raise to the specifier of this position to anchor the clitic chain. If the pragmatic requirements for this fronting process are not met, then we see the appearance of the dummy clitic anchor *nu*. As such, I would argue that some of Hale’s relative clause examples do in fact display *in-situ* behavior, since the host of the clitic chain, whether *nu* or otherwise, is a constituent of the sentence as far as topicalization across wh-elements is concerned:

- (41) *nu=za ANA DINGIR.MEŠ kuit arkuwar iyami*  
 now.REFL for gods which prayer make.1SG  
 “which prayer I make for the gods” (H122)

But, of course, these are just two different treatments of the same data. The definitive evidence in arguing that Hittite does show *in-situ* behavior would be to find a question in which two constituents precede the wh-element. Given the relatively small nature of Hoffner’s wh-question corpus and the

odd semantics that such a sentence would require, this is not an easy task, but I believe that there is exactly one such example:

- (42) *zik=wa=kan*                    *apūn anda kuwat aušta*  
 you=QUOT=COMP that in why look.2SG.PRET  
 “Why did you look at that (woman)?” (KBo V 3+ iii 71)

Here, we see that *zik* ‘you’ anchors the clitic chain instead of *nu*, so even under Hale’s analysis this would count as a topicalized constituent. The wh-word doesn’t appear until after the object of the verb, as the third constituent of the sentence. I argue that this example shows that Hittite does in fact have wh-*in-situ* behavior.

What about Tocharian? The same logic would apply: if all we find are questions where only one constituent precedes the wh-element, then there is no reason to assume that Tocharian behaves any differently than the rest of ancient Indo-European, with topicalization over wh-movement. In Tocharian too, however, I argue that I have found an example where two arguments of the verb precede the wh-word:

- (43) *somo-añyai*                    *somo ytārye kā westār*  
 single-traversable single road why called  
 “Why is the single road called the only-traversable?” (29b1<sup>C</sup>) Adams (2015)

This sentence is a bit complex, but it seems clear that both arguments do occur higher in the syntax than the wh-word. As such, I would argue that this example constitutes evidence that Tocharian does show true wh-*in-situ* behavior, alongside Hittite.

Second, recall the behavior of Wagnackel clitics in Sanskrit as seen in (40) above: Hale used the occurrence of the clitics to the right of the non-initial wh-word as further evidence that the movement across the wh-word was topicalization late in the derivation. Note, however, that in the Hittite examples the clitic chain attaches to the “topicalized” item itself instead, and is then subsequently followed by the wh-word. The specifics of this construction are not yet fully understood, but it is clear that this is not a case of the same topicalization-over-wh process put forward by Hale.

## 5.2 Conclusion and future work

With the thousands of years and miles that separate the attestations of Tocharian and Hittite, their syntactic similarities are remarkable. Both show left-headed C domains and right-headed T domains, and both clearly demonstrate wh-*in-situ* behavior in wh-questions, with landing sites for the wh-elements preverbally and clause-initially in both languages, indicating their shared syntactic antiquity. There are confounding factors for this analysis, however. First, Tocharian’s earliest attestation is almost two thousand years after that of Hittite and the other Anatolian languages, giving Tocharian much more time to innovate wh-*in-situ* behavior on its own. Or, even if it did inherit its *in-situ* behavior from PIE, it had much more time to undergo other syntactic operations that obscure the original Proto-Tocharian situation inherited from PIE.

The other major confounding factor is language contact with Chinese. Being on the western edge of the East Asian sprachbund in which *in-situ* wh-question behavior is the rule, it cannot be ruled out that the behavior we see in Tocharian is the result of close contact with these East Asian wh-*in-situ* languages. This contact is evidenced by lexical borrowings both from Chinese into Tocharian and

from Tocharian into Chinese.

However, I argue that the structural similarities in the *in-situ* constructions of these two languages are far too similar to be chance. Further, I believe that these similarities provide further evidence of Tocharian's archaism, such that two thirds of Indo-European at this stage of reconstruction shows clear *in-situ* behavior. Majority rules would dictate that we reconstruct wh-*in-situ* for PIE as well, but of course the situation is not that simple.

There are two possible developmental pathways: either PIE was originally wh-*in-situ* and this situation was maintained in Hittite and Tocharian while the rest of IE innovated wh-movement, or PIE was originally wh-moving, and Hittite and Tocharian separately developed wh-*in-situ* behavior in parallel. With the surprising syntactic similarities exhibited by Tocharian and Hittite, I am inclined to believe that it's the former.

In fact, it's even possible that, if Anatolian did break off from PIE earlier than Tocharian, we might be able to find evidence in Tocharian of the beginnings of wh-movement syntax. We do, for example, see obligatory wh-movement to the very beginning of the clause in embedded questions in Tocharian, but not in Hittite, an innovation shared with the rest of ancient Indo-European. Relative clauses in Tocharian, however, at least in poetry and literary language likely to be old, do not have this requirement. There may be other syntactic indicators of the development of wh-movement in Indo-European as well, but if there are, we haven't found them yet.

Future work, however, will hopefully provide a more satisfying and conclusive answer to this question. The behavior of the other ancient Indo-European languages needs to be double-checked to determine whether they have any residual wh-*in-situ* behavior that could strengthen the arguments in favor of reconstructing wh-*in-situ* for the PIE. Hale's thorough analysis of Indo-Iranian seems to indicate that this branch at least likely does not show residual wh-*in-situ* behavior.

I also plan to take another look at Sideltsev's analysis of Hittite. I feel that many of his arguments are persuasive, but I'm not completely convinced that an analysis that does not require obligatory wh-movement into the left periphery doesn't exist. His strongest argument for this obligatory movement is the "verb-raising" data where a low adverb occurs to the right of the verb in a clause that would otherwise have normal surface word order. Low adverbs seem to be the only objects that show this behavior in relation to verbs in Hittite, and I believe that a postsyntactic phonological explanation, or some other syntactic explanation for this data likely exists. Furthermore, his use of ForceP as the site of the clitic chain seems at odds with the intended use of ForceP by Rizzi, and, I believe, is the result of his conclusion that wh-question words must move up so high into the left periphery. If we could convincingly argue that these wh-words are truly *in-situ* within the TP layer, then we clear up a lot of space lower in the left periphery for the clitic chain to sit.

Next, I plan to research the development of wh-*in-situ* and wh-movement behavior in languages that have innovated it over the course of their attested histories (Hindi from Sanskrit, for example) in order to better understand the processes by which a language switches from one state to the other. Ultimately, I hope to provide a conclusive answer not only to the question of whether PIE had wh-*in-situ* in wh-questions, but also to the question of how wh-movement behavior evolved in its daughter languages. And, after all this is done, I hope to tackle exactly the same questions about wh-*in-situ* behavior for relative constructions in PIE.

## References

- A *Comprehensive Edition of Tocharian Manuscripts*. <http://www.univie.ac.at/tocharian/>. Retrieved: April 11, 2017.
- Adams, Douglas Q. 2015. *Tocharian B: a Grammar of Syntax and Word-formation*. Inst. für Sprachen und Literaturen der Univ. Innsbruck, Bereich Sprachwiss.
- Cheng, Lisa Lai-Shen. 2009. Wh-in-situ, from the 1980s to Now. *Language and Linguistics Compass*, **3**(3), 767–791.
- Cheng, Lisa Lai-Shen, & Rooryck, Johan. 2000. Licensing wh-in-situ. *Syntax*, **3**(1), 1–19.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.
- Clackson, James. 2007. *Indo-European linguistics: an introduction*. Cambridge University Press.
- Fortson, Benjamin. 2004. *Indo-European language and culture: An introduction*. Cambridge Univ Press.
- Garrett, Andrew. 1994. Relative clause syntax in Lycian and Hittite. *Die Sprache*, **36**(1), 29–69.
- Goedegebuure, Petra. 2009. Focus structure and Q-word questions in Hittite. *Linguistics*, **47**(4), 945–969.
- Hale, Mark R. 1987. *Studies in the Comparative Syntax of the Oldest Indo-European Languages*. Ph.D. thesis, Harvard University.
- Held, Warren H. 1957. The Hittite relative sentence. *Language*, **33**(4), 3–52.
- Hoffner Jr, Harry A. 1995. About questions. *Studio historiae ardens: Ancient Near Eastern Studies Presented to Philo H. J. Houwink ten Cate on the Occasion of His 65th Birthday*, 87–104.
- Huggard, Mattyas. 2011. On Wh-(non)-movement and internal structures of the Hittite preposed relative clause. *Pages 104–126 of: Proceedings of the 22nd Annual UCLA Indo-European Conference*. Bremen: Hempen, vol. 83.
- Jasanoff, Jay H. 2003. *Hittite and the Indo-European verb*. Oxford University Press.
- Kayne, Richard S. 1994. *The antisymmetry of syntax*. MIT Press.
- Kiparsky, Paul. 1995. Indo-European origins of Germanic syntax. *Pages 140–169 of: Battye, Adrian (ed), Clause structure and language change*. Oxford.
- Melchert, H Craig. 2013. The position of Anatolian. *Handbook of Indo-European Studies, Oxford: Oxford University Press*.
- Peyrot, Michaël. 2008. *Variation and change in Tocharian B*. Vol. 15. Rodopi.
- Pinault, Georges-Jean. 1997. Sur l'assemblage des phrases (<<Satzgefüge>>) en tokharien. *Berthold Delbrück y la sintaxis indoeuropea hoy. Actas del Coloquio de la Indogermanische Gesellschaft (Madrid, 21- 24 de septiembre de 1994)*. Madrid/Wiesbaden: Reichert, 449–500.
- Pinault, Georges-Jean. 2008. *Chrestomathie tokharienne: textes et grammaire*. Peeters.
- Probert, Ph. 2014. Relative clauses, Indo-Hittite and Standard Average European. *Pages 137–64 of: Proceedings of the 25th Annual UCLA Indo-European Conference*.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. *Pages 281–337 of: Elements of grammar*. Springer.
- Schmidt, Klaus T. 2007. THT 1540. *Instrumenta Tocharica*, 223–254.
- Sideltsev, Andrej. 2014. Wh-in-situ in Hittite. *Pages 199–222 of: Typology of Morphosyntactic Parameters. Proceedings of the International Conference Typology of Morphosyntactic Parameters*.