Wh-question word syntax in Tocharian and Hittite, and their implications for PIE reconstruction

Ryan Hearn
June 3, 2017

1 Introduction

- The ancient Indo-European languages have long been considered strong wh-movement languages (Fortson 2011).
- However, at least Anatolian and Tocharian display syntactic behavior that appears to be, if not wh-in-situ, at least non-initial (Pinault 1997; Adams 2015; Hoffner 1995; Sideltsev 2014).
- Compare (1) and (2) below, from Tocharian B and Hittite, respectively.

(1) *somo-aiya* *somo yl gyr kā westār*  
  single-traversable single road *wh* called

  “Why is the single road called the only-traversable?” (29b1C) Adams (2015)

(2) *šummaš=lan 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)

- In this paper I will explore various syntactic analyses of these non-initial wh-question constructions in Tocharian and Hittite, and discuss their implications for the reconstruction of the proto-language.

2 The Tocharian data

- Adams (2015) provides many examples of Tocharian B wh-questions.
- He claims 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).
- But, only for the *k*-initial pronouns (except for *kā ‘why’) is this rule mostly exceptionless. The one exception he cites:

(3) *śraddhaus kusu ste*  
  faith *what is*

  “What is faith?” (23a5C) Adams (2015)

- For all other question words, Adams notes “that one might wonder whether these words are subject to wh-fronting at all” (p. 48).

3 The Hittite data

- According to the question word corpus in Hoffner (1995), wh-question words in Hittite seem to appear in two different locations, the lower of which is immediately preverbal.
- The wh-words in this low position, however, always show up directly before the sentence-final verbal complex, but also always precede negation:

(9) *n=an EGR-pa [k]uwat UL pš-te[ni]*  
  CONN=him back *why* not give=2PL.PRS

  “Why don’t you give him back?” (HKM 58:21-22) (Hoffner 1995)

- The other position in which we find wh-question words is clause-initially, hosting the clitic chain.
  - These clitic chains 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.

(10) *kui-i=war=an hara-n D Pirwa[-i] URU Haššiwa=aza uuatu=zi[l]*  
  who-NOM.SG=QUOT=him eagle-ACC.SG Pirwa-DAT.SG Hassu-ABL 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)

- Exceptions to wh-initial word order is as high as 25-30% of the corpus.
- Here are his examples showing non-initial behavior. Examples of *kā ‘why’:

(4) *somo-aiya* *somo yl gyr kā westār*  
  single-traversable single road *wh* called

  “Why is the single road called the only-traversable?” (29b1C) Adams (2015)

(5) *sū kā swāsām*  
  it *why* rain

  “Why will it rain?” (140b4) (Adams 2015)

Examples of *mākte ‘how’:

(6) *tumen no stamāssīgīne mākte gentets*  
  therefrom but establishment *how* wind.GEN.PL

  “But therefrom *how* is the establishment of the winds?” (41b5C) Adams (2015)

(7) *y(e)s mākte maš(c)e(r)*  
  you.PL *how* act


Examples of *intsu ‘which’

(8) *posa śpāmnen rsāke intsu ste*  
  all.PERL best *rishi which* is

  “Which seer is best?” (107b3C) Adams (2015)
3.1 The Hittite wh-in-situ analysis of Sideltsev (2014)

- One recent analysis of the Hittite data, that of Sideltsev (2014), concludes that the Hittite data does constitute in-situ behavior, as the wh-question word does not obligatorily raise to the highest specifier of CP.

- Sideltsev begins by arguing for right-headedness within the TP domain.
  - This is due to the SOV nature of Hittite, combined with the observation that auxiliaries in Hittite inevitably occur clause-finally, after the main verb.

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

- He also cites verb movement as further evidence of the vP external nature of wh-phrases in Hittite.

He also cites verb movement as further evidence of the vP external nature of wh-phrases in Hittite.

He takes this word order as evidence that all wh-elements must end up outside the vP/TP domain.

The counterexample to this claim, in which the preverb occurs before the wh-element, he explains as topicalization of the preverb due to D-linking, i.e. that the answer is drawn from the speaker and hearer’s common ground:

- Sideltsev interprets this 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.

(14) \text{nu} \quad \text{k[u]n} \quad \text{memiyan} \quad \text{kuwat} \quad \text{ya-tten} \quad \text{QATAMMA}

\text{CONN} \quad \text{this.ACC.SG} \quad \text{matter-ACC.SG} \quad \text{why} \quad \text{do-2PL.PRET} \quad \text{in.this.way}

“So, why have you handled this matter in this way?” (NH/NS (CTH 63.A) KUB 19.31 + rev. ii 27”–31”) Sideltsev (2014)

- Based on this evidence, Sideltsev concludes that Hittite does show wh-in-situ behavior.

- Wh-question words obligatorily move first to Spec-FocP, and then optionally move further to Spec-FoC.

4 Three interpretations of the Hittite and Tocharian data, and their implications for PIE reconstruction

4.1 Late topicalization across obligatory wh-movement

- Hale (1987) took an in-depth look at Indo-Iranian languages to explain why wh-question words sometimes occur as the second constituent in their clause, e.g. the following line from the Rig Veda:

(15) \text{ráthaṅ kó} \quad \text{nir avartayat}

\text{chariot who prepared}

“Who prepared the chariot?” (RV 10.135.5b)

- Hale concludes that examples like this across the Indo-Iranian languages indicate that obligatory wh-movement across the board may then be followed by a topicalization process around the wh-word.

- The occurrence of Wackernagel clitics to the right of the wh-word in these situations further evidences that this topicalization process occurs late in the derivation.

(16) \text{indr̄a kīṁ} \quad \text{asya sakhýe} \quad \text{cakāra}

\text{indra what his} \quad \text{friendship did}

“How could they have brought up the army of Taruluiya and the new troops of Zilapiy?”

MH/MS (CTH 186’) HKM 43 obv. 1’–5’ Sideltsev (2014)

- For Hittite specifically, Hale follows Held (1957) in assuming that the clause-initial connective particle \text{n}u is effectively extrasyntactic, and that initial position in the clause is to the right of \text{n}u plus the clitic chain.

- Sideltsev (2014) 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.

- Thus, both Hale and Sideltsev both seem to agree that Hittite shows obligatory wh-movement followed by optional topicalization of one element across the wh-word.

- The vast majority of the Tocharian evidence also conforms to this analysis.
4.1.1 Counter-evidence

- Evidence in favor of Hittite showing true in-situ behavior would therefore require a question in which two main constituents precede the wh-element.
- I believe that I’ve found exactly one such example in the Hittite corpus:

  (17) \texttt{zik=wa=kan aπín anda kuwət ušətə}
  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 \texttt{zik ‘you’} anchors the clitic chain instead of \texttt{nu}, so even under Hale’s analysis this would count as a topicalized constituent.

- For Tocharian as well, I believe that I have found an example that cannot easily be accounted for by the topicalization argument:

  (18) \texttt{somo-aṇga somo yārə ye kā wəstər}
  single-traversable single road why called
  “Why is the single road called the only-traversable?” (29b1C) Adams (2015)

- Further, recall the behavior of Wackernagel clitics in Sanskrit as seen in (16) 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, and is then subsequently followed by the wh-word.
  - It seems clear that this Hittite behavior is not a case of the same topicalization-over-wh process described by Hale.

4.2 True in-situ behavior

- If one concludes that the data mentioned above constitutes strong enough evidence against the topicalization-over-wh analysis for Tocharian and/or Hittite, then an in-situ analysis may be preferable, in which the wh-words remain in their base-generated/case-assigned positions within TP.
- The following is a brief attempt at such a syntactic in-situ analysis.
- To start, according to Adams (2015), embedded questions in Tocharian show obligatory wh-movement to the front of the clause without exception:

  (19) \texttt{wələ preksə ci kə nai ʃəntsər mà [św]/fəsta]}
  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 (20), (21), and (22) below, from Cheng & Rooryck (2000).

  (20) \texttt{Quel livre Jean a-t-il acheté?}
  which book Jean has-he bought
  “Which book did Jean buy?”

  (21) \texttt{Jean a acheté quoi?}
  Jean has bought what
  “What did Jean buy?”

  (22) \texttt{*Je me demande que Jean a acheté quoi?}
  I REFLECT wonder that Jean has bought what
  “*I wonder what Jean bought.”

- 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.
- Either the wh-element may overtly move to spec-C to check the Q-feature, as seen in (20), or the intonation morpheme in C itself may check the feature in (21).
- In French this intonation morpheme must also be a root morpheme: it may only appear in matrix clauses and take matrix scope.
- Both the underspecified and root nature of this morpheme are parameters that may vary cross-linguistically.
  - In Portuguese, Cheng and Rooryck argue that the morpheme is underspecified but not a root morpheme, as it may take embedded scope.
- This analysis is straightforwardly applicable to our Tocharian (and Hittite) data.
- In Tocharian and Hittite we see the exact same parametric variation Cheng and Rooryck noted in French and Portuguese:
  - Tocharian does not allow non-initial wh-question words in embedded questions (as seen in (19) above), while Hittite does.

  (23) \texttt{šumeš=wa [DJINGIR.MES UL uikattenni / kiššan=wa=mu kuiš iyaŋ hərzī}
  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)

- Further, according to Melchert (1994), there may be evidence for a special yes/no question intonation in Hittite, indicated by plene spelling of the clause-final verbs of two questions in Old Hittite.
  - To my knowledge, no examples of this plene spelling exist for non-initial wh-questions, but it’s certainly something to keep an eye out for.
- So, we see that there already exist syntactic analyses that can account not only for the apparently optional in-situ behavior that we see in Tocharian and Hittite, but also that these analyses can even account for their typological differences as well.
4.3 Asymmetrical in-situ behavior

- In every in-situ example in the Hittite and Tocharian corpora, with the possible exception of the example below, the wh-words in question are either how or why.
  - This example can potentially be explained by wh-movement, followed by topicalization of the entirety of TP as one constituent across it.

\[(24) \text{tusa yweru nwam} \quad \text{ka,se} \quad \text{thus.PERL.SG rage.OBL.SG shout.3SG.SUBJ.ACT who.N.SG}

“Who roars more rage than that?” (PK AS 7M.a4)

- Many languages show an asymmetry between argument wh-behavior and adjunct wh-behavior (for example, Austronesian languages as shown by Aldridge (2002)).

- It’s possible that Hittite and/or Tocharian could show this asymmetry, that this behavior is old, and that the development of across-the-board wh-movement was innovated after Anatolian and Tocharian broke away from the proto-language.

- Of course, this effect could also be due to the relatively odd pragmatics needed for two arguments of a verb to precede a third wh-argument, and our in-situ sample size may just be too small.

5 Conclusions and directions for future work

- There are multiple possible conclusions that can be drawn from the Hittite and Tocharian wh-question data:

  1. The in-situ sentences in both languages can be explained away by other movement processes: Hittite and Tocharian behave like other ancient IE languages in that they have wh-movement, followed by optional topicalization across the wh-word.
    - The number of sentences with two or more constituents before the wh-word is a tiny percentage of the corpus, with only one or two examples in Hittite and Tocharian.
  2. Obligatory wh-movement in Tocharian (and possibly Hittite as well) is asymmetric.
    - All of the sentences with two or more constituents before the wh-word in both Tocharian and Hittite have an adjunct wh-word, either how or why.
    - If this behavior contrasts with the across-the-board obligatory wh-movement elsewhere in Indo-European, this could lead to a reconstruction of this asymmetry for the proto-language, and the separate innovation of across-the-board wh-movement in the younger daughter languages.
  3. The in-situ sentences in Tocharian and/or Hittite cannot be explained away by other fronting processes, leading to the conclusion that one or both languages do show truly in-situ behavior.
    - If both Tocharian and Hittite show this behavior, we should likely reconstruct it for the proto-language as well.
    - For non-in-situ behavior to be old, Anatolian and Tocharian would either have to share a common ancestor unrelated to the other IE languages (unlikely), or have separately innovated their in-situ behavior (also unlikely).

- Ultimately, it appears that whichever conclusions we draw currently rest on the analysis of the very small number of Hittite and Tocharian sentences that do not seem to fit Hale’s topicalization-across-wh picture, along with the Wackernagel clitic data from Hittite.

References


