Mixed Headedness in Tocharian
and Its Implications for PIE Reconstruction*

RYAN HEARN
Cornell University

1. Introduction

The study of constituent structure in the Indo-European (IE) languages is both a
very old science and a very young one. Age-old syntactic phenomena have only
relatively recently seen rigorous theoretical analysis, which has given us more
detailed knowledge of the synchronic syntax of these ancient languages, as well
as improved our syntactic reconstructions of the proto-language. This paper is an
attempt to continue this trend of grounding robust word-order generalizations
throughout the IE languages in generative syntactic theory.

The Tocharian languages have not yet seen extensive generative syntactic
study, so this paper builds a foundation for future work by providing an analysis
of the overall clause structure of Tocharian B, focusing on the headedness of the
CP and TP domains.1 Section 2 uses the behavior of complementizers, Wacker-
nagel clitics, and wh-questions to argue in favor of left-headedness for CP in
Tocharian. For the TP domain, I show that the positioning of auxiliaries and ne-
gation clearly points toward a right-headed analysis. Section 3 briefly compares
the Tocharian data with other old IE languages in order to determine the extent
to which this analysis of Tocharian could also account for the behavior of its sis-
ters. It is demonstrated that the CP and TP behavior of Tocharian as described in
this paper parallels that of the oldest IE languages quite closely, and that these
similarities warrant further study. Section 4 discusses prior right-headed VP

* Thanks to Michael Weiss, John Whitman, and Miloje Despić for their invaluable advice on
this project and to attendees of the 29th Annual UCLA Indo-European Conference for their
excellent feedback.

1 I will be working within the Principles and Parameters framework first introduced in Chomsky
1981, in which headedness is a parameter of language whereby a given phrase may be either
head-initial (left-headed) or head-final (right-headed).

Throughout the paper I use “CP” to refer to all projections in the clause above TP, includ-
ing the left periphery or “expanded CP” as described by Rizzi (1997). Similarly, “within TP”
refers to all tense and verbal projections below CP.
reconstructions of Proto-Indo-European phrase structure, and suggests extending this right-headedness to the entire TP domain.

2. Mixed headedness in Tocharian

2.1. Left-headedness in CP

Left-headedness within CP across the IE languages is effectively the default stance taken in the literature, in that most syntactic analyses of old IE languages assume or adopt a left-headed CP (e.g., Huggard 2011:105, Goldstein 2016:25, Hale 1996:193, Hale 2018); it is therefore unsurprising that Tocharian would also exhibit left-headedness within CP. However, since this has not been formally shown in the literature to my knowledge, this section examines the data supporting left-headedness within CP in Tocharian and lays out the details of Tocharian phrase structure overall.

First and foremost, the location of the complementizers themselves provides the best evidence for the location and headedness of CP in a given language. From Hackstein (2013:117–8), who compares the development of complementizer behavior from relative constructions across the ancient IE languages, we see that the most common means of marking complement clauses in Tocharian is with a null complementizer. This unfortunately does little to show us its syntactic location. We do, however, see TB (kₚce) and TA kucne ‘whom, what’ showing up as complementizers in a minority of clause-taking predicates, and, when used in this fashion, they always show up at the beginning of the clause:

\[
\text{Do thou thus by my command that the Kashgarian P. goes through (LP-1a1/3Col; Adams 2013:193)}
\]

\[
\text{twe ŋi yaitkorsa maṇṭ pyāṁ kuce kaṣake Puttamitre parra yam}
\]

There are no clause-final elements in Tocharian that show complementizer behavior, so though complementizers are relatively rare, when they do show up they exclusively appear initially.

In addition, grammatical particles appearing early in the clause in Wackernagel position show behavior indicating that they head their own left-headed projections in the left periphery. For example, Koller (2013:4–7) locates Tocharian A ne (and its Tocharian B cognate nai) in the head of FocP since it immediately follows wh-phrases (which Koller places in Spec-FocP) clause-initially.

Further, according to the Complementizer Attraction Universal of Bresnan (1972:42), only languages with a clause-initial Comp permit a Comp attraction transformation. Since wh-movement in Tocharian (and across most of the archaic...
Indo-European languages) is to a position at the left edge of the clause, CP should be left-headed.

Taken together, these phenomena (clause-initial complementizers, second-position discourse clitics, and leftward wh-movement to clause-initial position) strongly support the position that CP in Tocharian was left-headed.

2.2. Right-headedness in TP

Adams (2015:15–6) claims that “neutral” word order in Tocharian B is SOV, since most clauses in the language show this surface ordering of constituents. In fact, Adams (2015:16), citing a few examples of non-literate business prose texts from Pinault 2008, notes that practically every non-copular sentence is verb-final, though poetic texts and literary prose have much freer word order. Verb-finality is a good start when looking for right-headedness in general, but is inconclusive, as there are many ways of deriving verb-final order with left-headed clause structure. One easy way is to simply move everything to a position above the verb’s final landing site, a trivial set of transformations with the advent of the expanded left periphery of Rizzi 1997. As a consequence of the availability of this theoretical machinery, a right-headed analysis not only has to fit the data, but must do so better than a left-headed analysis.

2.2.1. Tocharian Auxiliaries

I checked for auxiliary constructions to determine whether auxiliaries follow main verbs sentence-finally in neutral clauses. There are a few modal-like verbs such as cāmp- ‘be able to’, yāt- ‘id.’, and skāy- ‘try’ that take infinitives, but since their syntax could be multi-clausal, I will only be looking at constructions with a participle or gerund and an inflected copula. Tocharian has periphrastic perfect, future, necessitive, and potential constructions that fit this criterion. Notably, the overwhelming majority of these constructions cited by Adams place the auxiliary clause-finally, after the main verb.

(1) toyā aṣiyana po lalāṃṣuwa stāre

These nuns have worked everything. (MSL.19.160) Adams 2015:124

As shown in Figure 1, this auxiliary evidence fits nicely with a right-headed TP analysis, since it allows us to generate the auxiliary in T clause-finally for φ-feature purposes, while leaving the participle below in the verb phrase. Note that right-headedness explains the ordering of object and verb nicely as well, with the feminine plural preterite participle lalāṃṣuwa appearing in V to the right of its object.
This evidence does not completely rule out the possibility of a left-headed TP domain, however, as every other constituent in the sentence could simply be topicalized around the auxiliary sitting in the left-headed T. We need another element with a well-established syntactic position to help us triangulate the location of the auxiliary.

2.2.2. Negation

According to Adams 2015:53–4, mā is the most common clausal negator and prohibitive, by itself accounting for 87% of all negated sentences, with all other instances of negation consisting of mā plus another morpheme. These negation compounds only appear clause-initially, but mā itself may occur either clause-initially or immediately before the inflected verb lower in the clause. With most

---

2 As it is not necessary to depict the relevant structural relationships, vP is not shown in the (simplified) tree. Also, the current paper does not address the question of whether the Tocharian TP possesses an EPP feature, a strong syntactic feature that forces movement of the subject to the specifier position of TP.

3 While an interesting question, the behavior of initial mā is outside the scope of this paper. When used clause-initially, mā is often, though not always, prohibitive, and as a result likely heads a projection in the left periphery, potentially Force. If the correct analysis is along these lines, this behavior of mā provides further evidence that Tocharian is left-headed within the expanded CP.
analyses since Pollock 1989 locating negation in either the head or specifier of a NegP between TP and vP, interaction between negation and auxiliaries should shed light on their underlying syntactic structure. I was able to find one instance of mā collocated with a verbal auxiliary complex in Adams 2015:131:4

(2) tem yiknesa weweñu mā tākam

(If) he has not spoken in this way … (331b3/45)

Fig. 2. Derivation of (2)

Note how the negation in (2) appears precisely between the participle and the copula. With our posited right-headed TP domain, we would expect a right-headed NegP to be located between the TP and vP layers, and that’s exactly where we find it. Further, this word order cannot easily represent an underlying left-headed structure, because mā should be in NegP, which is located below T, but it is appearing above tākam, which must be at least as high as T for φ-feature

4 Here is another potential example from the CEToM corpus, though again, this verb+infinitive construction is likely indicative of a multicausal structure:

(kuśalamū)Inta karsatsi mā yāt(ām)

They will not be able to recognize the roots of evil. (IOL Toch 109)
expression. Example (2) must therefore represent a right-headed TP, with the copula being generated in T for φ-features above mā sitting in Neg.

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.5 And, in fact, we see evidence of negation and the inflected verb acting as a single constituent elsewhere, in sentences like (3) below. Here, the inflected verb has merged with Neg, moved up to T, and the entire complex has then been topicalized below the wh-question word.

(3) ka[sic] mā wešcer krent (reki)

Why do you not say the good (word)? (20b6c) Adams 2015:48

Again, this word order cannot represent underlying head-initial TP: since mā is appearing non-initially (indicating that this is not a prohibitive usage), it would have to be somehow located above the inflected verb in T if (3) were indeed representative of a left-headed syntactic structure.

Thus, we see that a right-headedness proposal for Tocharian within the TP domain not only provides us with an elegant account for the surface order of the auxiliary constructions we see in the language, but also gives us a straightforward account of the behavior of preverbal negation.6

3. Mixed headedness evidence elsewhere in old Indo-European

3.1. Anatolian

Most syntactic analyses in the Anatolian literature either avoid the topic of headedness, or seemingly default to a head-initial analysis (e.g., Garrett 1994, Huggard 2011). Sideltsev (2014:206–8), however, argues specifically for right-headedness within TP and left-headedness above TP for Hittite. He bases this claim primarily on the “rigidity” of clause-final verbs and the rarity of postverbal subjects and objects. He also notes the behavior of the auxiliaries ḫark- ‘have’ and ēš- ‘be’, which always follow the main verb, as seen below:

5 I am indebted to Miloje Despić for this idea.
6 Due to space and time constraints, the following analyses are based on limited data and secondary sources. I hope to give a much more in-depth treatment of the behavior of these other daughter languages in the near future.
We see then that the word order data from Hittite matches that of Tocharian surprisingly closely for our purposes, especially if the Hittite clitic chain is heading (or at least cliticizes to) a left-headed functional projection within the expanded CP.

3.2. Sanskrit

One of the most complete theoretical treatments of word order and local headedness in Sanskrit is Schaufele 1991. He follows most of Western scholarship in assuming base SOV word order, and claims that the majority of phrases are head-final. PPs seem to show mixed headedness, with the majority being head-final, behavior consistent with Tocharian and the other ancient IE languages. Hock (1984:94) notes that 97% of the clauses in prose texts of the later Vedic period are verb-final, compared to 63% of poetic texts, though as we indicated earlier, verb-finality is not enough to prove TP right-headedness.

Another potential piece of evidence for right-headedness within TP is the periphrastic constructions we see appearing in later Vedic, mirroring those seen in Tocharian and Hittite above. The earliest attestation of the periphrastic perfect, for example, seems to be gamayāṃ cakāra ‘he went’ from Atharvaveda 18.2.27. If these auxiliary constructions do indicate right-headedness within TP once they appear in late Vedic, their development could constitute a more overt realization of TP head-finality that was already present in earlier, more surface-ambiguous Vedic as well.

As far as evidence for left-headedness in the CP domain goes for Vedic, the clitic, complementizer, and wh-movement data seems to mirror that of Tocharian. Complementizers like yād appear on the left, wh-movement is to the left, and Hale (1996:192) places Rigvedic second position clitics in the C head.

I would be remiss to leave out clause-final quotative iti in this discussion, but note that Hock (1982:42–62) explains at length how its complementizer-like behavior did not fully evolve until the classical period, leaving Vedic Sanskrit without any clause-final complementizers.

3.3. Italic

The most thorough work on phrasal headedness in Italic is undoubtedly Ledgeway 2012, which describes in detail the gradual change of the headedness
parameter from head-initial to head-final in the development of the Romance languages. Interestingly, however, his position for overall clausal headedness is that both TP and CP emerged over the (pre-)history of Latin and Romance, as opposed to merely switching headedness like the rest of the language.

Ledgeway’s argument against the existence of CP is based on the idea from Kiparsky 1995:141 that PIE lacked clausal embedding, but see Probert 2014 for an argument that clusal embedding is present in all of the oldest IE languages, and should be reconstructed for the proto-language. For my purposes, the complementizers seen in the Latin data do appear clause-initially, and second-position clitics in Latin show much of the same behavior that we see in its sister languages (Hale 2018). I therefore see no reason to believe that the structure of Latin’s left periphery is appreciably different from the left-headedness we see in Tocharian.

Ledgeway also claims that the development of TP within Latin corresponds to the rise of the left-headed auxiliary constructions that we see in later Romance. But, auxiliary constructions already existed in Latin itself clause-finally, both with the copula and with habere. If the existence of auxiliary constructions constitutes evidence for the presence of TP (as an overt, in-situ expression of the φ-features in T), then the clause-final auxiliaries of Latin should suffice as evidence of the presence of TP as well. Furthermore, these auxiliary constructions show similar behavior to the others we have seen in the old IE languages, with the majority of auxiliaries occurring clause finally, after the participle:

(5) [cum cognitum habeas] quod sit summi rectoris […] numen

[When you realize] the will of the supreme lord (Cic. Fin. 4.11, Ledgeway 2012:131)

I would therefore argue that the major innovation from Latin to Romance was not the development of TP, but the switch of TP-headedness from clause-final to clause-initial, seen on the surface in the shift from Latin’s clause-final auxiliaries to the near-initial auxiliaries Ledgeway describes for later Romance.

3.4. Greek

As Fortson (2010:160–3) mentions, Greek displays second position clitic and wh-movement data similar to the other IE languages, which, combined with the fact that the Greek complementizers ὅτι and ὡς occur initially in their clauses, strongly suggests that Greek was head-initial within CP.

Headedness within TP for Greek is a bit trickier. As Goldstein (2016:18) states, “Ancient Greek is unique in its degree of word-order variation.” This
combinatorial freedom leads him to posit a largely flat clausal structure for Ancient Greek at the time of Herodotus, with clausal word order determined instead by semantics and pragmatics. Taylor (1994:33–4) states that, during the time of Homer, Greek is primarily verb-final, but by the time of the New Testament, Greek has become verb-medial. She instead concludes that Herodotus is composed of “a head-final base” and “a head-medial base.” For our reconstructive purposes, we will be looking only at the word order of verbal constructions in Homer.

Per Bentein 2016:107, the earliest periphrastic constructions have already shown up in Homer, with 58 examples composed of the copula or ἔχω + the perfect participle. I went through each of these examples, and found that 43 of them, 74%, placed the inflected verb clause-finally, with all but one also placing the participle immediately before the verb, mirroring the behavior seen in the other old IE languages. Ceglia (1998:29) shows us that by the time of Herodotus the participle generally follows the copula, but even if, as Goldstein states, there is not enough evidence to make an explicit decision about the behavior of TP in Herodotus, the Homeric evidence does favor a right-headed TP analysis.

4. Mixed headedness in PIE?

Delbrück (1900:82–3) was the first to attempt an overall clause structure reconstruction for the proto-language, concluding that PIE must have been SOV based mainly on Sanskrit word-order evidence. Here, I would like to go a step further, and tie this reconstruction of word order to a specific underlying syntactic structure.

Per Hackstein (2013:117–8), overt complementizer behavior seems to develop relatively late in the prehistory (or even history) of many of the IE daughter languages. He claims that zero-embedding is the most securely reconstructible method of sentential embedding across the earliest attestations of the daughter languages. However, even though we cannot securely reconstruct any of the individual complementizers for the proto-language, I argue that we can reconstruct their shared underlying syntactic structure, especially given their overwhelming similarities when they do emerge in the daughter languages.

Also, if, as Hale and others have suggested, some second-position clitics head their own projections in the left periphery, then we have further support that PIE CPs were left-headed if we reconstruct this widely shared behavior as well. Finally, we have Bresnan’s Complementizer Attraction Universal, mentioned in Section 2 above. With wh-movement being so ubiquitous across the Indo-European languages, it makes sense to reconstruct this behavior for the proto-language,
adding further to the pile of evidence pointing toward left-headedness in the PIE CP.

As far as right-headedness lower in the clause goes, Sapp (2016:396–9) reconstructs head-final VPs for PIE due to the Germanic evidence when considered alongside the SOV word order of the other IE languages. He does not, however, go so far as to reconstruct right-headedness for TP in PIE. Similarly, Krisch (2017:115) reconstructs head-final VPs due to SOV evidence from across the Indo-European languages, but does not reconstruct TP at all for the proto-language.

In a similar vein, I contend that, although morphological equations among the periphrastic constructions among the daughter languages are absent, we can reconstruct their shared underlying syntactic structure back to PIE. The separate innovation of such strikingly similar auxiliary constructions across so many of the IE daughter languages is unlikely to be due to chance, and should be taken as evidence of a common underlying inherited structure in the absence of evidence to the contrary. We can and already do reconstruct SOV word order and head-final VP behavior for PIE; the evidence in favor of going one theoretical step further and reconstructing a right-headed TP for the proto-language as well is, in my opinion, worth serious consideration.

References


