Disharmonic headedness in Indo-European

Evidence from innovation: reconstructing disharmonic headedness for Proto-Indo-European

Ryan Hearn
Cornell University
rdh238@cornell.edu

October 12, 2018
NYU Syntax Brown Bag
Main goals

1. Establish feature-based syntactic reconstruction as a reliable application of the Comparative Method

2. Examine complementizer development across the early Indo-European (IE) languages
   - Reconstruct a left-headed CP domain for Proto-Indo-European (PIE)

3. Examine auxiliary construction development across the early IE languages, especially Tocharian and ancient Greek
   - Reconstruct a right-headed TP domain for PIE
Once we determine that two languages are genetically related, we can begin to reconstruct their common ancestor.

In the modern era, this often takes the form of feature-based phonological and morphological reconstruction.

<table>
<thead>
<tr>
<th>'hundred'</th>
<th>Latin</th>
<th>Greek</th>
<th>Sanskrit</th>
<th>Lithuanian</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>*kʰmtóm</td>
<td>centum</td>
<td>(he)katón</td>
<td>šatam</td>
<td>šimtas</td>
<td>hundred</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>'blood'</th>
<th>Latin</th>
<th>Greek</th>
<th>Sanskrit</th>
<th>Lithuanian</th>
<th>Old English</th>
</tr>
</thead>
<tbody>
<tr>
<td>*krewh₂-</td>
<td>cruor</td>
<td>kréas</td>
<td>kravis</td>
<td>kraujas</td>
<td>hræw</td>
</tr>
</tbody>
</table>

Note that setting up the right correspondence set is of the utmost importance for arriving at the right reconstruction.
Historically, syntactic reconstruction worked very similarly to phonological/morphological reconstruction.

Set up correspondence sets of either exact phrases, or of general word order.

- Greek *kléos áphthiton* and Sanskrit *śrávas ákṣitam* ‘undying glory’
- “Most old Indo-European sentences end in a verb”
Limits of Syntactic Reconstruction?

Anything beyond this rudimentary reconstruction is often thought to be impossible

Usually due to the belief that appropriate correspondence sets cannot be created (e.g. Lightfoot 2002a)

- Lexical information can be reconstructed since it is stored intact in a mental lexicon that is transferred to new generations of speakers

- Most assume that syntax, on the other hand, is procedurally built separately for every utterance

  - Very few sentences then, if any, are stored in the lexicon in their entirety
Feature-based Syntactic Reconstruction

- With the advent of the Minimalist program of Chomsky (1995), however, we now have a strong theoretical foundation for syntactic reconstruction
  - The computational component of syntax is universal
  - Synchronic and diachronic variation is due to featural differences of lexical items and functional heads
- As mentioned by Hale (1998) and others, we can ignore the syntactic building process that Lightfoot was concerned about
  - The elements that drive the variation in this process are stored in the lexicon
  - We can therefore reconstruct syntax using the same tools we use for phonology and morphology
Theoretical assumptions of the current paper

1. I will be working within Minimalism and will assume that the computational component of the syntax is universal
   ▶ Synchronic and diachronic variation is due to featural differences of lexical items and functional heads
   ▶ This assumption allows me to unify phonological, morphological, and syntactic reconstruction as different manifestations of the same process

2. Headedness is parametric, and part of the featural content of functional projections
   ▶ This assumption is much less important: the method of reconstruction I outline in this paper should give equally good results in a non-parameterized model

3. Disharmonic headedness, where functional projections in a language have different headedness, is allowed by the syntax
So, what if a given function word is not reconstructible for the proto-language?

▶ What if the words that fill a functional role aren’t cognate

▶ Or, what if each daughter language cognate developed this usage separately during each language’s attested history?
A relatively straightforward example: the IE complementizer

- Complementizers in early IE are not all cognate:
  - Latin *quod* and Tocharian *kuce/kucne* from *kw-o* (the PIE interrogative stem)
  - Gothic *batei* from *to*- (a PIE pronominal stem)
  - Sanskrit *yád* and Greek ὅτι and ὣς, from *Hjo*- (the PIE relative stem)

- More importantly, as shown by Hackstein (2013), complementizer behavior mostly developed within the attested history of these languages
  - Latin *quod* was only extended from use with factive verbs in the Classical period
  - Sanskrit *yád* develops its own complementizer usage from relative usage during the Classical period
  - Hittite *kuit* and Tocharian *kuce/kucne* start as adverbial adjuncts which later develop complementizer usage.
An overt complementizer cannot be reconstructed for PIE

- As shown by Hackstein, **zero-embedding** is likely the only reconstructible method for embedding sentential complements after verbs of utterance and cognition for PIE.

- So, assuming complementizers fill a functional head (let’s call it C), without any reconstructible complementizer, we have no way of locating C in the syntax, right?
  - The C domain in PIE could be either left-headed or right-headed
  - But without a reconstructible complementizer in C, how can we know which?
Setting up a correspondence set

▶ Even if PIE used zero-embedding, it still has a functional head C
  ▶ C in PIE is just filled with a null complementizer, or a complementizer that we can no longer reconstruct

▶ So, instead of trying to reconstruct both the phonological form *and* position of C, I reconstruct just the position itself, regardless of what phonological form this position takes in the daughter languages

▶ I therefore set up a correspondence set for the underlying syntactic structure, and ignore the specific complementizers used
Parallel complementizer innovation

- When we ignore the form of the complementizers, we immediately see striking similarities in the daughter languages’ syntax
  - Every single innovated complementizer, when it ultimately shows up, appears clause-initially

- These languages aren’t all independently innovating a left-headed C domain
  - They’re innovating a phonological form to fill the left-headed C domain that they already share

- We see that our correspondence set unilaterally points toward a null clause-initial C for the proto-language
Not structural innovation

- Note that this is not “structural” syntactic innovation
  - The structural features of C haven’t changed - just which phonological form is associated with the node
  - The parallel innovation of separate phonological forms to fill the C node, i.e. *complementizers*, cues us in to the shared structural syntactic reality
  - CP was also left-headed in their ancestor, Proto-Indo-European
1. Reanalysis, as defined by Langacker (1977), is “change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation”
   - Examples of complementizer development discussed above are spec-head reanalyses
   - The reanalyzed elements must be surface-adjacent to the null complementizer
     - The null complementizer must be left-headed to get the correct surface order
   - This also ties in nicely with the Complementizer Attraction Universal of Bresnan (1972), which states that the landing site of a Comp attraction transformation (i.e. wh-movement) must be adjacent to C
     - If we reconstruct clause-initial wh-movement for PIE as most do, then we must reconstruct a left-headed CP as well
Corroborating evidence for left-headed CP in PIE

2. Grammatical particles in Wackernagel positions show behavior indicating that they likely head their own projections in the left periphery

- Koller (2013) locates Tocharian A ne (as well as its Tocharian B cognate nai) in the head of FocP since it immediately follows Wh-phrases (which Koller places in spec-FocP) clause-initially

- For Sanskrit, Hale (1996) places Wackernagel clitics in the C head

- Further, Scharf (2015) points out that the Sanskrit question particle api occurs clause-initially, instead of the clause final position we would expect if CP was right-headed (e.g. ka in Japanese)
Now that we have a hammer, let’s find some nails

- Up next is an age-old problem with significantly more disagreement in the literature
  - Reconstructing a left-headed CP for PIE is pretty well supported by the other arguments
  - Now I will show that the method we used to add support to that position is just as useful for reconstruction elsewhere
Delbrück (1893) was the first to reconstruct clausal word order for PIE.

- He concluded that PIE must have been SOV based mainly on Sanskrit word-order evidence.

Sapp (2016) and Krisch (2017) both reconstruct head-finality within the VP domain.

- Again, this is mostly due to general SOV word order across the early IE languages.

There are plenty of ways to derive SOV word order without needing VP head-finality, however.

Let’s see if we can some more evidence to better triangulate the exact location of these verbal elements.
Early IE Auxiliaries

- Just as we saw with complementizers, auxiliaries are ubiquitous in the early IE languages, but their presence cannot be securely reconstructed for PIE
  - We don’t see the Sanskrit periphrastic perfect showing up until the Atharvaveda’s *gamayām cakāra* ‘he went’, constructed with the *do* verb
  - Latin auxiliary constructions, however, initially use the copula and only later develop with the verb *habere* ‘have’
  - The oldest periphrastic constructions in Greek show up in Homer, mostly with the copula
    - There are also a couple of ambiguous examples of proto-auxiliary constructions with *ěχω* ‘have’
  - Hittite uses *ḥark- ‘have’ and the copula
Some have reconstructed prehistoric periphrastic constructions (like the Latin imperfect as described by Weiss 2009), but it is not known if these date back to PIE

We may not be able to securely reconstruct a single auxiliary construction for PIE

But, as with complementizers in early IE languages, separate innovation of auxiliary constructions in the daughter IE languages can give us insight into their inherited syntax

This is especially the case if all of the earliest attested daughter languages agree in the syntax of their separately innovated auxiliary constructions

So, let’s take a look and see just how similar the early IE auxiliary constructions are
Tocharian B

Tocharian is an extinct Indo-European branch spoken on the northern edge of the Tarim Basin in what is now Xinjiang province in northwestern China.

- It is the eastmost ancient IE language, and had at least three dialects, appropriately named A, B, and C.
- Our manuscripts date from the 6th to 8th centuries CE.

Tocharian is important for PIE reconstruction, as many Indo-Europeanists believe that it was the second language to split off from the parent language after the Anatolian languages.

- This means that it can be compared with the reconstructed ancestor of the other eight branches directly to reconstruct the second-oldest layer of PIE.
Tocharian possesses periphrastic perfect, future, necessitive, and potential constructions consisting of a participle/gerund and an inflected copula.

I gathered all examples of these periphrastic constructions from the translated portion of the Comprehensive Edition of Tocharian Manuscripts (CEToM).

I also gathered a few additional examples from Adams (2015).
## Auxiliary constructions in Tocharian B

<table>
<thead>
<tr>
<th>Period</th>
<th>Type</th>
<th>Part-Aux</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaic</td>
<td>Verse</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Classical</td>
<td>Verse</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Classical</td>
<td>Prose</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>Late</td>
<td>Verse</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Late</td>
<td>Prose</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>65</td>
<td>23</td>
</tr>
</tbody>
</table>

- Tocharian B overwhelmingly (74%) prefers to end auxiliary clauses with a participle followed immediately by the inflected copula.
- In the entire corpus there are no examples of prose sentences ending any other way.
Per Adams (2015), Tocharian B’s “neutral” word order is SOV, comparable to the word order reconstructed for PIE.

This auxiliary data doesn’t fit very well with the head-final VP analysis of SOV word order in PIE, however:

- Assuming the inflected copula would be sitting in a left-headed T above the right-headed VP.
- Most of the periphrastic sentences in the corpus would have to move everything into the left periphery.

I think it more likely that Tocharian B is **right-headed within its TP domain**.

- This accounts for the auxiliary order within Tocharian without resorting to the left periphery for the majority of sentences.
Corroborating evidence: Negation in Tocharian

- According to Adams (2015), *mā* is the most common clausal negator and prohibitive, by itself accounting for 87% of all negated sentences.
  - *mā* may occur either clause-initially or immediately before the inflected verb much lower in the clause.
- I was able to find one instance of *mā* collocated with a verbal auxiliary complex:

(1) \[ \text{tem yiknesa weweꜜnu mā tākaꜜm} \]
\[ \text{this way spoken not be.3SG.SUBJ} \]
\[ "(If) he has not spoken in this way" (331b3/4^L_, Adams) \]
Corroborating evidence: Negation in Tocharian

- Note how the negation appears precisely between the participle and the copula.

- With our posited right-headed TP domain, we would expect a right-headed NegP located between the TP and vP layers.

- And, in the one example we have, that’s exactly where we find it.
Corroborating evidence: Negation in Tocharian

(2)

```
CP
  C
  TP
    T'
      NegP
        vP
          v'
            v
            Neg
              mā
              has
              tākaṃ
              this one
              DP
                tem
                yiknesa weweñu
                in this way spoken
```
I looked at the examples of auxiliary usage in Homer, collected by Bentein (2016). All were periphrastic perfects.

<table>
<thead>
<tr>
<th>Work</th>
<th>Part-Aux</th>
<th>Part-Aux-NP</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Iliad</em></td>
<td>22</td>
<td>4</td>
<td>3</td>
<td>29</td>
</tr>
<tr>
<td><em>Odyssey</em></td>
<td>18</td>
<td>3</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td><em>Hymns</em></td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>44</strong></td>
<td><strong>7</strong></td>
<td><strong>7</strong></td>
<td><strong>58</strong></td>
</tr>
</tbody>
</table>

44 of the 58 Homeric examples (76%) place the auxiliary immediately following the participle clause-finally.

An additional 7 place the auxiliary immediately following the participle clause-finally, except for a single postposed NP or piece of an NP.
Auxiliary behavior in early Greek

- We see that the Greek auxiliary facts closely resemble the Tocharian B data just discussed
  - SOV, but with the inflected copula following the participle clause-finally

- Here too, we see a strong case for right-headedness in TP to best account for this word order, specifically due to the relationship between the participle and inflected copula

- If I can find a solution (perhaps prosodic?) to the postposed NPs and split NPs, the case for right-headedness in TP in the earliest Greek would be even stronger
Auxiliary behavior in later Greek

- It is worth noting that, as pointed out by Ceglia (1998), by the time of Herodotus the participle generally follows the copula in the sentence.

- This mirrors the observation of Taylor (1994) that Homeric is primarily OV, with the younger Greek dialects developing more frequent VO word order.

- I am suspicious that these facts constitute a shift in TP headedness.
  - I think this shift in headedness will be central to any eventual syntactic solution to the “unique degree of word order variation” seen in Classical Greek, as addressed in Goldstein (2015).
Auxiliaries and word order in the other old IE languages

With Greek and Tocharian showing such striking similarities in auxiliary behavior, let’s turn to the existing literature on the other old IE languages, and see if we can find some more.
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) specifically argues instead 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.
Most importantly, he also notes the behavior of the auxiliaries ḫark- ‘have’ and ēš- ‘be’, which always follow the participle, as seen below:

(3) [(našma)] ĖSAG kuiš ZI-it
or granary somebody.NOM.SG.C by.his.will
kīnu-an ḫar-z[(i)]
break-PRTC.NOM.SG.N have-3SG.PRS

“Or somebody has broken open a granary by his own will”

(MH/MS (CTH 261.3) KUB 13.1(+) rev. iv 20’-23’)
Sidetsev also concludes that the only reasonable syntactic structure that can account for these auxiliary word order facts is a left-headed CP and a right-headed TP.

This nicely mirrors what we saw from Tocharian and Greek earlier.

More work remains to be done on the other Anatolian languages to determine the extent to which their auxiliary facts reflect those of Hittite.
Once periphrastic constructions like the aforementioned *gamayām cakāra* from the Atharvaveda start showing up, they usually occur clause-finally with the auxiliary following the verbal element, mirroring the behavior of the other old IE languages.

Schaufele (1991), one of the most complete analyses of Sanskrit word order, follows most of western scholarship in assuming base SOV word order, and claims that the majority of phrases are head-final.

Similarly, Hock (1984) notes that 97% of Vedic prose texts are verb-final, compared to 65% of poetic texts.

For our purposes, these tendencies are telling but not yet conclusive. I plan to do a more in-depth analysis of periphrasis in Vedic soon.
The most thorough work on phrasal headedness in Italic is undoubtedly Ledgeway (2012). He describes in detail the gradual change from head-final to head-initial exhibited throughout Latin to the modern Romance languages.

The argument seems to be that both TP and CP emerged over the (pre-)history of Latin and Romance. The CP argument originates in the idea that PIE lacked clausal embedding; see Probert (2014) for evidence to the contrary. This argument also seems odd since Ledgeway uses the left periphery to account for much of Latin’s free word order, which is mirrored by other IE languages. Also note that we do see complementizers already in the Latin data, and that when they appear, they show up clause-initially.
For TP, the claim is that the development of TP corresponds to the rise of the left-headed auxiliary constructions in later Romance.

But, clause-final auxiliary constructions are already ubiquitous in Latin itself, both with the copula and later with habere.

(4) cum cognitum habeas [...] when known you have [...] “When you realize [...]” (Cic. Fin. 4.11, Ledgeway (2012))

I would argue that the major innovation from Latin to Romance was not the development of TP, but was more likely the switch of TP-headedness from clause-final to clause-initial.
Sapp (2016) presents a detailed argument for base SOV word order and head-final VPs in Old High German.

He derives surface V2 word order in Germanic through raising of the verb.

He mentions that his analysis is compatible with that of Lenerz (1984), who had earlier posited head-final TP structure for OHG.

Weiß (2007), on the other hand, argues for head-initial TP, and maintains that surface V2 word order is derived through movement of the finite verb into T itself.
And then, of course, there’s Modern German, which many would consider the *Paradebeispiel* for left-headed CP/right-headed TP languages, especially in embedded clauses.

For our purposes, the main syntactic innovation of Germanic from PIE would be V2 word order through obligatory T-to-C movement.
We’ve seen striking similarities in auxiliary behavior across the early IE languages

Once the daughter languages develop auxiliaries, they overwhelmingly come clause-finally, usually immediately after the participle

The daughter languages all point toward the same synchronic structural relationship between inflection (T) and the rest of the verbal domain

The syntactic parameter that fits this word order best is right-headedness within TP
Wrapping up

- With our structural correspondence set in agreement, we can now reconstruct this structural relationship for the parent language as well.

- These parallel innovations of clause-final auxiliaries show us the relationship between the head-final TP and VP inherited from PIE.
Alternatives?

1. Disharmonic headedness was an areal feature that spread throughout the Indo-European languages
   ▶ The IE languages are so spread out that this occurred either early enough as to be indistinguishable from PIE, or across an infeasibly broad geographic area

2. These auxiliary construction innovations conspired to produce the disharmonic headedness in each of these daughter languages independently
   ▶ The likelihood of all of the daughter languages agreeing this closely by chance is, in my opinion, prohibitively low

3. Disharmonic headedness was innovated in one (or more) Indo-European language, and spread to the others through borrowing
   ▶ Similar to the areal feature hypothesis, for this feature to have been borrowed into all of the earliest IE languages it would have to either travel unreasonably far, or happen early enough to be indistinguishable from PIE
This idea of parallel syntactic innovation revealing underlying syntactic similarities inherited from the parent language constitutes a new tool available for syntactic reconstruction.

- It provides a new argument not only in favor of reconstructing SOV word order for PIE, but of reconstructing a specific corresponding underlying structure.

- Combined with the complementizer data discussed earlier, it provides evidence for reconstructing a left-headed CP and right-headed TP for PIE.

- This structure was then inherited and made explicit through the innovation of auxiliary constructions in the earliest IE daughter languages.
Why is all of this important?

- This project illustrates just how valuable feature-based syntactic reconstruction is
  - It takes the idea a step further and argues that in some cases we can even reconstruct syntactic information where we cannot reconstruct phonological or morphological information
- It adds syntax to the list of innovations useful for subgrouping
  - As Kim (2018) states, it is currently assumed that the innovations useful for subgrouping “may be phonological, morphological, or lexical”
- It adds to our knowledge about PIE!
  - A better understanding of PIE syntax will better inform typological generalizations and can even help inform PIE phonological and morphological reconstruction
Other takeaways: IE and the FOFC

▶ If PIE really was left-headed above TP and right-headed below TP, and the IE daughter languages inherited the same syntactic structure

▶ Then at no time during IE’s reconstructible history did a right-headed projection dominate a left-headed one

▶ The Final-over-final Constraint (Holmberg, 2000), which states that a right-headed projection may not dominate a left-headed one, seems to be borne out by the IE data, as predicted by Biberauer et al. (2014)

▶ Both synchronically by the early IE data, and diachronically by reconstruction
What I’m doing now

- I’m gathering each auxiliary example from the Vedic poetic and prose texts to see what structural insights the word order variations can give us.

- I’m very interested in the word salad that is post-Homeric ancient Greek.
  - I will see if treating Greek as a language in the process of switching TP headedness from right to left gives us any insight into its striking word order variations.
Thank you for your attention!
References I


Delbrück, Berthold. 1893. Vergleichende Syntax der indogermanischen Sprachen. 3.


References III


References IV


