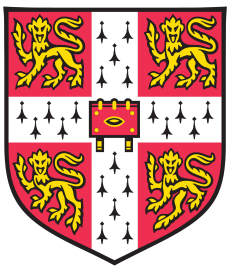


University of Cambridge
Faculty of Classics

The Syntax of Relative Clauses and Related Phenomena in Proto-Indo-European

Krishnan J. Ram-Prasad

Peterhouse

A thesis submitted for the degree of

Doctor of Philosophy

March, 2022

Declaration of Originality

I, Krishnan Jesse Ram-Prasad, confirm that this thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

This thesis is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

The length of this thesis does not exceed the limit of 80,000 words prescribed by the Faculty of Classics Degree Committee.

Krishnan J. Ram-Prasad
March, 2022

Abstract

In this thesis, I reconstruct the syntax of relative clauses in Proto-Indo-European (PIE).

Syntactic reconstruction, particularly in the case of PIE, has presented a perennial challenge to Comparative Philologists. I demonstrate that Minimalist reconstruction provides a viable methodology through which we may address this challenge, integrating the specific task of reconstruction into the broader field of historical syntax. The methodology necessarily demands a discussion of Minimalist theories of relative clause structures, which I provide in relation to the synchronic analysis of ancient Indo-European (IE) languages.

I then undertake a philological survey of descendants of the putative relative pronouns in PIE: **yó-* and **kʷí-*/**kʷó-*. The debate on which (if either) of these was the "original" relative pronoun in PIE has lasted for over a century. I argue that neither form can be excluded as a relative pronoun in PIE, and that together they reflect what was a unitary functional category: the relative pronoun, *REL.

Relative clauses across the ancient IE languages exhibit grammaticalised fronting of the relative pronoun, traditionally referred to as "*wh*-movement". To analyse the nature of this movement, I provide a detailed reconstruction of the PIE left periphery. Because of the additional role of discourse-driven movement (Topicalisation and Focalisation) and clisis ("Wackernagel's Law") in the left periphery, my reconstruction has implications beyond relative clauses, shaping our understanding of the significant role pragmatics plays in the ordering of constituents in PIE, as well as the syntax-phonology interface.

On the basis of all the above, I reconstruct the syntax of relative clauses in PIE. I argue that PIE had what I refer to as an "anaphoric" relative clause, that could generate a variety of surface forms, including correlatives and postnominal relatives. The "anaphoric" relative clause was adjoined to the matrix clause at the CP level, and could either precede or follow the matrix clause. I argue further that we cannot exclude embedded relative clauses from our reconstruction of PIE, but that it is possible to derive them diachronically from an earlier "anaphoric" relative clause.

This thesis thus demonstrates that Minimalist reconstruction is not only a viable methodology, but a fruitful one, allowing us to establish concrete conclusions about PIE syntax.

Acknowledgements

If I've learned anything in the last four years, it's that you can't write a PhD thesis on your own. Thanks are due to many people who've supported me during my time as a graduate student. Here are some of the most important.

First, I'd like to thank my supervisor, Rupert Thompson. In his prodigiously good-natured and entertaining way, Rupert always manages to help me see data and arguments from perspectives I hadn't (and wouldn't have) considered. The combination of this argumentative creativity with his philological rigour have kept my feet firmly on the ground while addressing the rather lofty and daunting topic of PIE syntax. I'd also like to thank Rupert for his pastoral support, especially through the wilderness of the year 2020, and later in my negotiation of that other great wilderness, the academic job market.

My second supervisor, James Clackson, saw me through my registration and annual reviews, and provided detailed and helpful feedback at various stages of the writing process. His encyclopaedic knowledge of the literature has been invaluable in bringing together theoretical and philological research to address the topic of syntax in ancient languages.

The reader of this thesis will note the prominence of Vedic Sanskrit in my analysis. I owe my knowledge of this language to Vincenzo Vergiani, who's been my Sanskrit teacher for almost eight years now. Indeed, it was probably in one of his classes back in 2015 that I encountered my very first correlative clause. Although the Faculty of Classics has always supported my work on Sanskrit under their *aegis*, Vincenzo, as external supervisor for both my MPhil and PhD, has been instrumental in facilitating my study of Sanskrit as a Classicist.

On the syntactic side of things, I've benefited greatly from extensive discussions with Ian Roberts, who's read and commented upon various parts of this thesis over the last year. Ian's theoretical work on Minimalist reconstruction is part of what inspired me to take on the challenge of reconstructing PIE syntax; I am very fortunate to have had his input as my theories on relative clauses and the left periphery crystallised.

Thanks are also due to the members of the Philology Caucus at large, who have so enriched my time as a graduate student, both as colleagues and as friends. Among these, Mark Darling merits a special mention for his extremely helpful guidance through the Celtic data. The one other person in my cohort who had the temerity to undertake a PhD in Comparative Philology, Edoardo Chiattelli, has been an exceptional companion from day one; it has been a privilege to work alongside him.

Indeed, all the academic support in the world would be worth little without the camaraderie of my fellow graduate students, and my dear friends. In addition to those already mentioned, I'd like to thank Zoë Audra, Flo Kipps, Konstantinos Lygouris, Sarah Sheard, Stephan Stephanides and Nir Stern.

This brings me to my family, who've supported me throughout my education. From varying distances, my grandparents have maintained a keen interest in what I'm doing, and continue to show unerring optimism for what's in store in the future. My mother and father have far surpassed the threshold of

parental duty in their commitment to reading and discussing my research with me, despite it being—at times—lost on them. And it's been an utter pleasure to have shared my last two years in Cambridge with my sister, Nalini. Her humour and faith have been a tonic for the soul.

Finally, and foremost, I thank Imogen Russell: my partner in this life, my best friend, and the one who has shared every low and every high with me during the long years of this degree. Imo, you have supported me and continue to support me in ways no one else ever could. I will be forever grateful to you, and for you.

To those I've missed from these acknowledgements, I can only apologise: assume that if it's not for reasons of space, I've simply been absent-minded. To be fair, my mind is due a holiday.

***

This research was funded by by the Arts & Humanities Research Council and the Ibn Battutah Fund. Grant Reference AH/L503897/1.

For Pāṭṭi

அன்போடு இயைந்த வழக்கென்ப ஆருயிர்க்கு
என்போடு இயைந்த தொடர்பு

a̲npōṭu iyaintu va̲la̲kke̲npa āruyirkku
e̲npōṭu iyaintu toṭarpu

‘When breath and bones come together, they create
life lived with love’

(Tirukku̲ra̲ḷ 8.3, trans. Shulman)

Contents

1 Introduction 1

1.1 Background . . . 2
1.2 On the viability of syntactic reconstruction . . . 3
1.3 Minimalist terminology and theoretical assumptions . . . 4

2 Methodological and Theoretical Foundations 8

2.1 Minimalist reconstruction . . . 8
2.1.1 Syntactic change . . . 8
2.1.2 Syntactic comparison . . . 12
2.1.3 Functional categories . . . 15
2.1.4 Uniting old and new . . . 17
2.2 Relative clauses . . . 18
2.2.1 Semantic types . . . 18
2.2.2 Syntactic types . . . 21
2.2.2.1 'Plain' relatives . . . 21
2.2.2.2 Correlative relatives . . . 33
2.3 The left periphery . . . 39

3 The Relative Pronoun(s) of Proto-Indo-European 46

3.1 A brief survey . . . 46
3.2 **yó-* . . . 47
3.2.1 Relative uses of **yó-* . . . 47
3.2.2 Non-relative uses of **yó-* . . . 53
3.2.2.1 Demonstrative . . . 53
3.2.2.2 Sentential connective . . . 56
3.3 **kʷí-/*kʷó-* . . . 59
3.3.1 Vocalism . . . 59
3.3.2 Relative uses of **kʷó-* . . . 61
3.3.2.1 'Determinate' and 'indeterminate' relatives in Hittite . . . 65
3.3.3 Non-relative uses of **kʷó-* . . . 72
3.3.3.1 Interrogative . . . 72

3.3.3.2 Indefinite . . . 74
3.4 Which "which" is which? . . . 78

4 Clisis and the Left Periphery in Proto-Indo-European 81
4.1 Onwards and upwards: The "front" of the clause . . . 81
4.2 Clisis . . . 84
4.2.1 Clitic characteristics . . . 84
4.2.2 "Second position": Wackernagel's legacy . . . 86
4.3 The Vedic initial string: A case study . . . 89
4.3.1 Prosody-dominant approaches to Vedic clisis . . . 90
4.3.2 Syntax-dominant approaches to Vedic clisis . . . 92
4.3.3 Vedic clisis reconsidered . . . 99
4.4 Left-peripheral clitics in PIE . . . 101
4.4.1 Ancient Greek . . . 102
4.4.2 Latin . . . 110
4.4.3 Hittite . . . 115
4.4.4 Summary . . . 122

5 Relative Clauses in Proto-Indo-European 125
5.1 Position of the relative pronoun . . . 125
5.1.1 Relative clisis in Proto-Indo-European: A negative appraisal . . . 126
5.1.1.1 Sanskrit . . . 126
5.1.1.2 Celtic . . . 128
5.1.1.3 Hittite . . . 134
5.1.2 Fronted *REL . . . 138
5.2 Correlatives . . . 141
5.2.1 CP vs. DP? . . . 141
5.2.2 Adjunction level . . . 143
5.2.3 Correlative pronouns . . . 149
5.3 Position of the head noun(s) . . . 152
5.4 'Plain' relatives . . . 155
5.5 "Anaphoric" relatives: The crossroads . . . 165
5.6 Questions of semantics . . . 167
5.7 Summary . . . 171

6 Conclusions and Outlook 173
6.1 Conclusions . . . 173
6.2 Outlook . . . 175

References 178

List of Abbreviations

Glosses

1 . . . . . First person

2 . . . . . Second person

3 . . . . . Third person

ABL . . . . . Ablative

ACC . . . . . Accusative

AOR . . . . . Aorist

C . . . . . Common

CL . . . . . Clitic

CND . . . . . Conditional

CONN . . . . . Sentence connective

DAT . . . . . Dative

DEF . . . . . Definite

DEM . . . . . Demonstrative pronoun

DET . . . . . Determinate

DU . . . . . Dual

EMPH . . . . . Emphatic particle

ERG . . . . . Ergative

F . . . . . Feminine

FUT . . . . . Future

GEN . . . . . Genitive

ILL . . . . . Illative

IMPF . . . . . Imperfect

IMPV . . . . . Imperative

INDEF . . . . . Indefinite pronoun

INDET . . . . . Indeterminate

INJ . . . . . Injunctive

INS . . . . . Instrumental

INT . . . . . Interrogative pronoun

LOC . . . . . Locative

LP . . . . . Local particle

M . . . . . Masculine

MOD . . . . . Modal particle

N . . . . . Neuter

NEG . . . . . Negator

NOM . . . . . Nominative

OPT . . . . . Optative

PERF . . . . . Perfect

PERFP . . . . . Perfect participle

PL . . . . . Plural

PPL . . . . . Participle

PPP . . . . . Perfect passive participle

PRES . . . . . Present

PRO . . . . . Personal pronoun

PTC . . . . . Particle

PW . . . . . P-word

QUOT . . . . . Quotative particle

REL . . . . . Relative pronoun

SG . . . . . Singular

SUBJ . . . . . Subjunctive

VOC . . . . . Vocative

Syntactic phrases

AdvP . . . . . Adverb Phrase

AP . . . . . Adjective Phrase

ConjP . . . . . Conjunction Phrase

ContrFocP . . . . . Contrastive Focus Phrase

CP . . . . . Complementiser Phrase

DemP . . . . . Demonstrative Phrase

DisjP . . . . . Disjunction Phrase

dP . . . . . 'Weak' Determiner Phrase
DP . . . . . Determiner Phrase
FinP . . . . . Finite Phrase
FocP . . . . . Focus Phrase
ForceP . . . . . Force Phrase
FP . . . . . Functional Phrase
IntP . . . . . Interrogative Phrase
IP . . . . . Inflection Phrase
NegP . . . . . Negator Phrase
NP . . . . . Noun Phrase
PP . . . . . Preposition Phrase
QP . . . . . Quantifier Phrase
TopP . . . . . Topic Phrase
SAP . . . . . Speech Act Phrase
TP . . . . . Tense Phrase
*v*P . . . . . Functional Verb Phrase
VP . . . . . Verb Phrase

Languages

AGk. . . . . Ancient Greek
Av. . . . . Avestan
Cb. . . . . Celtiberian
Eng. . . . . English
Gaul. . . . . Gaulish
Got. . . . . Gothic
Hit. . . . . Hittite
IE . . . . . Indo-European
Lat. . . . . Latin
Lit. . . . . Lithuanian
ME . . . . . Middle English
Myc. . . . . Mycenaean
NIE . . . . . 'Nuclear' Indo-European
OCS . . . . . Old Church Slavonic
OE . . . . . Old English
OIr. . . . . Old Irish
PDE . . . . . Present Day English
PIE . . . . . Proto-Indo-European
Skt. . . . . Sanskrit
Toch. . . . . Tocharian

Miscellaneous

2P . . . . . Second Position
ARC . . . . . Appositive Relative Clause
DM . . . . . Distributed Morphology
EXREF . . . . . External Reference
FR . . . . . Free Relative
HEA . . . . . Head External Analysis
HN . . . . . Head Noun
HRA . . . . . Head Raising Analysis
IAST . . . . . Intl. Alphabet of Sanskrit Transliteration
LF . . . . . Logical Form
LHNR . . . . . Leftmost Head Noun Rule
MA . . . . . Matching Analysis
MC . . . . . Matrix Clause
OC . . . . . Omission of the Copula
OP . . . . . Operator
P&P . . . . . Principles and Parameters
PF . . . . . Phonetic Form
PI . . . . . Prosodic Inversion
PLD . . . . . Primary Linguistic Data
RC . . . . . Relative Clause
RRC . . . . . Restrictive Relative Clause
V2 . . . . . Verb Second
WL1 . . . . . Wackernagel 1 (pronominal clitics)
WL2 . . . . . Wackernagel 2 (clausal clitics)
WLH . . . . . Wackernagel Hittite (clitics)
φ . . . . . Phonological phrase
ω . . . . . Phonological word

Chapter 1

Introduction

This thesis concerns itself with the reconstruction of relative clause syntax in Proto-Indo-European (PIE). It is structured as follows:

§1 **Introduction**. I provide some background to the importance and relevance of relative clauses in PIE syntax. I also introduce some of the issues surrounding syntactic reconstruction, and introduce Minimalist reconstruction as a methodology to address these issues. I finish the introduction with a brief overview of some basic tenets of Minimalist syntax, and some assumptions I make for the purposes of the thesis.

§2 **Methodological and Theoretical Foundations**. I present in detail the methodological and theoretical context for this thesis. I discuss how Minimalist reconstruction fits into a broader Minimalist theory of syntactic change, and how this shapes the way we go about syntactic reconstruction. I provide an overview of relative clause types and the Minimalist literature on relative clause structure. I conclude with a short exposition on the left periphery, which constitutes an important part of this structure.

§3 **The Relative Pronoun(s) of Proto-Indo-European.** I evaluate the syntactic behaviour of the descendants of the two possible forms of the relative pronoun in PIE: **yó-* and **kʷí-/*kʷó-*. This includes a survey of their descendants as attested in the ancient IE languages, as well as an analysis of their functions in both relative and non-relative uses.

§4 **Clisis and the Left Periphery in Proto-Indo-European.** I offer a detailed reconstruction of two intertwined aspects of PIE syntax: "second position" clitics and the left periphery. As noted above, the left periphery plays an important part in understanding the syntax of relative pronouns and relative clause structure more generally, constituting a key component of my overall theory. "Second position" clitics exhibit a special distribution in that left periphery, which places them at the centre of the discussion.

§5 **Relative Clauses in Proto-Indo-European**. I build a reconstruction of relative clause structure in PIE. I discuss the syntax of the relative pronoun, including its position within the left periphery

and its syntactic features. I also address larger questions of relative clause structure in PIE, including the categorical status of the relative clause; its syntactic relation to the main clause; the position of the head noun; the diachronic pathways between different types of relative clause; the interaction of syntax and semantics in PIE relative clauses.

§6 **Conclusions and Outlook.** I summarise my findings, noting the questions I have addressed and those which remain for future research.

1.1 Background

A relative clause, broadly defined, is a subordinate clause that modifies a given noun phrase. The noun phrase in question is traditionally described as the *antecedent*; in more contemporary terms, it is referred to as the *head noun*. Typically, a relative clause is introduced by a relative pronoun. Consider the following example:

(1.1) The tiger [**who** came to tea] (Kerr, 1968)

In this example, the head noun ('tiger') is underlined, the relative pronoun ('who') is in **bold**, and the relative clause is [bracketed]. While the head noun is modified by the relative clause, it typically acts as an argument in the main clause:

(1.2) The tiger [**who** came to tea] ate all the biscuits.

Relative clauses therefore yield a form of complex nominal, making them a point of special syntactic interest. This is compounded by the fact that relative clauses appear to be attested across a wide range of languages, in various syntactic forms. For example, the position of the head noun, the relative pronoun (if present), and indeed the relative clause itself are all variable, sometimes even within one language. As such, relative clauses have been and continue to be the subject of much enquiry within the field of syntactic theory.

Relative clauses are particularly well attested within the ancient Indo-European languages. As such, relative clauses in PIE have concerned numerous linguists over the last three centuries. In the 19th century, much of the work on the topic surrounded the 'original' form and function of the relative pronoun in PIE; from the mid-20th century onward, scholars have additionally concerned themselves with the more general syntactic structure of relative clauses in PIE. Indeed, the very possibility of syntactic reconstruction here makes relative clauses a central point of interest in our understanding of the language; through a conspiracy of methodological difficulties (see §1.2 below), the syntax of PIE remains elusive. A central task in contemporary Indo-European linguistics, therefore, is to draw back the curtain on its syntax: and relative clauses may help us do this.

This thesis, which constitutes a study in the syntax of relative clauses in PIE, is therefore motivated in (at least) two ways. First, relative clauses are a very interesting part of syntax; second, syntax is a very interesting part of PIE, and relative clauses, because they are well-attested **and** syntactically significant, represent something of a keystone for the reconstruction of PIE syntax.

1.2 On the viability of syntactic reconstruction

It is a common feature of works on syntactic reconstruction to include a spirited defence of the whole endeavour. Almost all such publications frame the defence as a response to the most detailed, recent and categorically negative appraisal of syntactic reconstruction: Lightfoot (2002). Building on earlier sceptical discussion of syntactic reconstruction (e.g., Watkins, 1976; Lightfoot, 1979; Winter, 1984) Lightfoot (2002) presents three central problems for syntactic reconstruction that supposedly do not apply to phonological reconstruction, which is often taken to represent the "gold standard" for Comparative Linguistics. Willis (2011) codifies Lightfoot's (2002) arguments in the following way:

- Correspondence-set Problem. Whereas in phonology one can form correspondence-sets—long lists of cognate words where one phoneme *corresponds* to another in a different language—such correspondence is not possible in syntax. This follows from the generativist notion that grammar is a system that generates sentences, which is to say sentences are not 'inherited' in the same way words are.

- Directionality Problem. We have no theory of universal *direction* in syntactic reconstruction. Whereas in phonology, we may draw upon ideas of "natural" change, e.g. voicing of consonants between vowels, there is no analogue for this in syntax.

- Radical-reanalysis Problem. Since syntactic change occurs when acquirers abduce "false" conclusions from the Primary Linguistic Data, and hence posit rules that are not present in the previous generation's grammar, change is random and unconditioned. Thus language users may make *radical reanalyses* and over a generation change may happen catastrophically, unpredictably and, crucially, in such a way that the previous state of the grammar cannot be reconstructed.

These problems have received extensive treatment in the last two decades (see, e.g., Roberts, 2007; Pires & Thomason, 2008; Willis, 2011; Walkden, 2013, 2014; Vázquez-González & Barðdal, 2019; Gildea *et al.*, 2020; Windhearn, 2021). The problems present a crucible in which any proposed methodology for syntactic reconstruction must be tested. Attempts at syntactic reconstruction that fare poorly in Lightfoot (2002) include the typological approaches of the 1970s (e.g., W. P. Lehmann, 1974; Friedrich, 1975; Miller, 1975), and pattern-based approaches such as Harris & Campbell (1995).[1] A methodology which seems more apt to address Lightfoot's problems—and the one I adopt in this thesis—is **Minimalist reconstruction**, which I discuss in detail in §2.1. This methodology sits within the theoretical framework of **Minimalism** (Chomsky, 1995), to which I give a brief introduction below. In Minimalist reconstruction, we see the convergence of what have become two somewhat divergent research areas: the formal study of syntactic change, and Comparative Philology. Clackson (2017: 202) suggests this division may arise from 'a curious accident of the history of linguistics[, namely] that the most influential work on the C[omparative] M[ethod] appeared just too early to take account of the Chomskyan revolution in

[1] See *inter alia* Watkins (1976), Lightfoot (2002), Willis (2011), Walkden (2013), Clackson (2017), Gildea *et al.* (2020) for discussions of these methods and their shortcomings.

syntax.' In this context, he suggests that '[h]istorical syntax has been playing catch-up ever since...'. An alternative framing might be to suggest that it is Comparative Philology that has to play catch-up with the advancements in syntactic theory that have been made since those most influential comparative works appeared. However we draw the metaphor, I aim for the use of Minimalist reconstruction in this thesis to be part of the 'catching up': of historical syntax with comparative methodology, and Comparative Philology with syntactic theory.

It is worth echoing here a sentiment expressed in some form in Walkden (2013: 109) and Windhearn (2020: 6): my aim is not to prove that Minimalism is the **only** formalism within which syntactic reconstruction is feasible. Indeed, in recent years syntactic reconstruction has also been undertaken within the framework of Construction Grammar (e.g., Barðdal *et al.*, 2013; Vázquez-González & Barðdal, 2019), which provides an alternative approach. Nevertheless, aside from its power in a reconstructive context (see §2.1), what Minimalism facilitates is the integration of syntactic reconstruction into a larger theory of syntax and syntactic change. Relative clauses especially have received such extensive treatment in both Generative syntax (see §2.2 for bibliography) and IE Philology (see §3.4 for bibliography) that the two disciplines present something of a natural pairing when addressing the question of PIE relative clause syntax. A different research question, still within the domain of syntactic reconstruction, may merit a different formalism—but for the purposes of this thesis, Minimalism provides precisely the necessary theoretical machinery for the task.

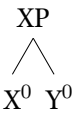
1.3 Minimalist terminology and theoretical assumptions

A PhD thesis is no place for a comprehensive introduction to Minimalism: for such purposes, I refer the reader to textbooks such as Radford (1997, 2004) and Boeckx (2007). I provide here only the briefest of overviews to account for some of the terminology used, and the basic concepts employed in this thesis. I also note at the outset that a substantial portion of the terminology and technical representations I use are not strictly Minimalist, but a represent mixture of Minimalism and earlier Generativist syntax, including Government & Binding (GB) theory and Principles & Parameters (P&P) (Chomsky, 1981; Chomsky & Lasnik, 1993). I highlight specific instances in the footnotes: where I rely on non-Minimalist theory, this is primarily for notational convenience, and to facilitate my engagement with previous syntactic work on ancient IE languages, very little of which follows 'strict' Minimalist theory.

The Minimalist program, as established in Chomsky (1995), is the latest iteration of Generativist syntax, a hugely influential approach to grammar dating back to Chomsky (1957). Minimalism inherits some core assumptions and terminology from earlier stages of Generativism, including Transformational grammar (Chomsky, 1957, 1965), and GB theory (Chomsky, 1981). Generativism is a nativist theory that presupposes the existence of an innate, domain-specific language faculty in the human mind: under these assumptions, the Minimalist program is built on the additional premise that this innate grammatical component is as small/efficient/**minimal** as possible. The structure-building operation that is unique to the grammar is MERGE. MERGE takes two items from the mental lexicon and *merges* them into a hierarchical structure in which one item *projects*, and the other does not. The items merged from the lexicon

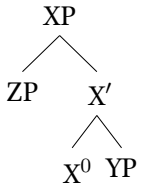
are conventionally referred to as *heads*, while the resulting projection is conventionally referred to as a *phrase*.[2]

(1.3)



Where a head projects up more than one level, the intermediate projections are conventionally labelled as barred (X-bar schema), leading to the following structure:

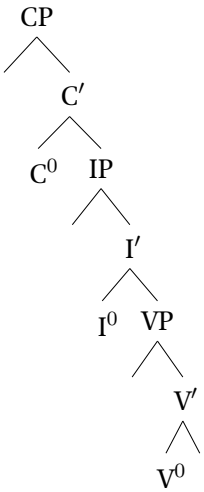
(1.4)



Additionally, each head may admit a *specifier* which precedes it and a *complement* which follows it. These positions are referred to as [Spec, XP] and [Comp, XP] respectively. In the diagram above, ZP is the specifier of X° ([Spec, XP]) and YP is the complement of X° ([Comp, XP]). Note that since each of YP and ZP has the same basic structure as XP, this structure is recursive. Within this schema, we may talk of the *domain* of a head: this constitutes anything contained within its complement; the domain of X° includes YP, [Spec, YP] and [Comp, YP].

With these basic building blocks, we can model a clause in the following way:

(1.5)



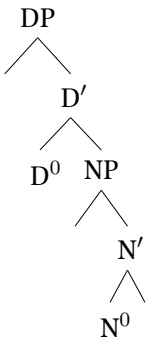
Each clause contains a CP-layer, an IP-layer and a VP-layer,[3] in which different types of linguistic information are encoded. In general terms, the VP-layer is responsible for lexical information and argument structure; the IP-layer is responsible for grammatical information such as tense/mood/aspect;

[2]These terms are not strictly part of Minimalism, but are inherited from transformational grammar; I use them over the preferred Minimalist terminology in terms of minimal/maximal projections for the sake of convenience, as noted above. I also use categorical labels in my diagrams for clarity, without any strong commitment for or against Bare Phrase Structure (Chomsky, 1995: 222–313).

[3]I collapse here intermediate functional projections such as *v*P for the sake of space—they are generally irrelevant for the purposes of this thesis.

and the CP-layer is responsible for clause-level information such as illocutionary force. As for nominals, I assume the DP-hypothesis (Abney, 1987):

(1.6)

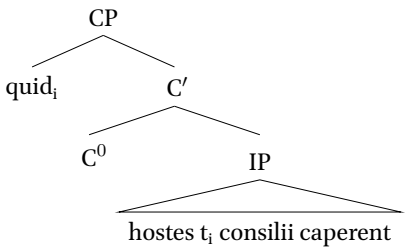


Under the DP-hypothesis, nominals are similarly divided into different 'layers': the DP-layer hosts functional items (e.g., articles), while NP hosts lexical nouns. Functional heads such as D° (and I° and C°) can be phonetically null, which is to say they are not visible in surface strings, but their presence is inferred by the behaviour of overt lexical items.

Returning to the basic structure-building operation, MERGE, Minimalism allows for two possible types: EXTERNAL MERGE and INTERNAL MERGE. The former takes new items from the mental lexicon and merges them into the derivation. The latter takes items that are already contained within the derivation, and merges them into a different location. This phenomenon was referred to as *movement* in earlier theories: in keeping with popular convention, I retain this terminology, while maintaining the assumption that movement is not a distinct process from MERGE, but rather a version of it. When something is moved, it leaves a *copy* in its base-generated position, which is subsequently deleted at the phonetic interface, leaving only the top-most copy to be pronounced. In conventional terms, the deleted copy is referred to as a *trace*, conventionally written as *t*. A simple example of INTERNAL MERGE (henceforth 'movement') is the fronting of interrogative pronouns in questions ("*wh*-movement"):

(1.7) [***quid***$_{i}$ *hostes* t_{i} *consili* *caperent*] *exspectabat*
INT.ACC.SG.N enemy.NOM.PL.M t_{i} plan.GEN.SG.N take.IMPF.SUBJ.3PL wait.IMPF.3SG
'He waited to see [**what**$_{i}$ plan the enemy would devise t_{i}]' (Caes. *Gal.* 3.22.1)

(1.8)



Here, the interrogative pronoun is fronted out of its base-generated position within VP to [Spec, CP].[4] To motivate this behaviour, Minimalism introduces the **feature theory** of movement. This means

[4]When I make no strong claims about the internal structure of a phrase or set of phrases (e.g., the IP, VP and DP in this example) I follow the convention of using a triangle in tree diagrams for shorthand.

that the relevant functional head (e.g., C°) can bear an *uninterpretable* feature, e.g, [*u*WH], which in this case types the clause as a *wh*-question. If an uninterpretable feature reaches the interfaces (see below), the derivation crashes. C° therefore acts as a *probe*, looking for a *goal* that has the relevant *interpretable* feature, [*i*WH], which can eliminate [*u*WH]. Interrogative pronouns bear just such an interpretable feature. So when C° probes into its domain and 'finds' *quid*, its [*u*WH] feature is eliminated, and the derivation is successful. If the [*u*WH] feature on C° is *strong* (per Chomsky, 1995), it triggers movement of the goal to its specifier; if [*u*WH] is weak, there is no overt movement.

The operations described above take place at what is often described as the **narrow syntax**, which takes items from the lexicon and builds them into a hierarchical structure. The conventional assumptions of Generativism are that once a given structure has been 'built' in the narrow syntax, it is sent to the interfaces, of which there are two:[5]

- **Phonetic Form** (PF). This is where the lexemes merged into the narrow syntax receive their phonological content, which ultimately produces the articulatory/gestural strings that constitute utterances.

- **Logical Form** (LF). This is where the structure receives its semantic interpretation, and is the locus of logical processes that are not necessarily encoded in the narrow syntax, such as scope and variable binding.

The four modules established so far—lexicon, narrow syntax, PF and LF—constitute what is traditionally be referred to as the (inverted) Y-model of grammar:

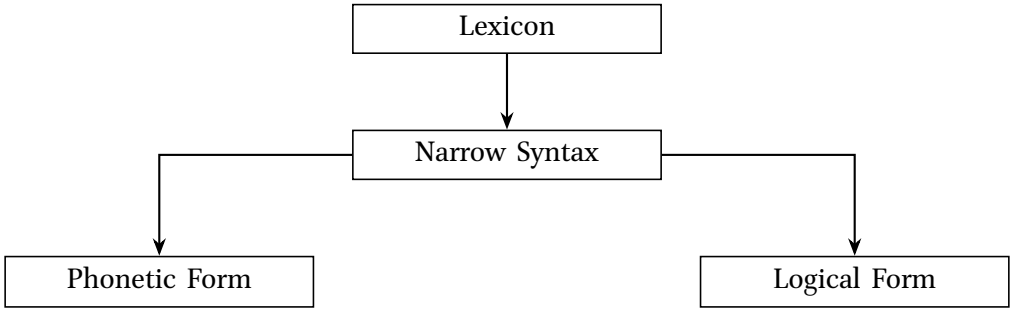


Figure 1.1: The Y-model of grammar.

In this thesis I focus primarily on the features of items in the lexicon and operations that take place in the narrow syntax; there are also various points at which PF is also implicated in surface word orders. For this reason, it is worth stating that once an element leaves the narrow syntax and reaches PF, hierarchical structure is no longer visible; PF is sensitive only to linear order.

This concludes my brief overview of the basic theoretical assumptions in this thesis: other aspects of Minimalist theory will be elaborated upon as necessary in the course of the discussion.

[5]Once again, I rely here on a convenient use of pre-Minimalist theory; in this case, the 'standard' Y-model as put forth in Chomsky & Lasnik (1977) (minus the distinction between 'deep' and 'surface' structure). Minimalist theories of the interfaces may differ substantially, on which see Boeckx (2006: 72–7); nevertheless, the distinction is maintained between syntactic and phonetic representations, which is the relevant part of this model for the purposes of this thesis.

Chapter 2

Methodological and Theoretical Foundations

In this chapter, I establish the foundations of a Minimalist reconstruction of relative clauses in Proto-Indo-European. I first expand upon the methodology of **Minimalist reconstruction**. I then turn more specifically to Minimalist analyses of relative clause structure, within which I will situate my analyses of ancient IE languages. I finish the chapter with a brief exegesis on the fine structure of the left periphery, which plays a crucial role in the word order constraints of ancient IE languages, with special relevance for relative clauses.

2.1 Minimalist reconstruction

The methodology of Minimalist reconstruction emerges from the broader picture of Minimalist approaches to syntactic change (to be discussed below): it is advocated in some form in Roberts (2007), Pires & Thomason (2008), Willis (2011), Walkden (2014) and Windhearn (2021). In this section, I give a brief overview of the central tenets of Minimalist reconstruction, with a critical discussion of previous studies implementing the methodology such as Willis (2011) and Walkden (2014). I add some clarifications of my own, addressing some methodological queries specific to PIE.

2.1.1 Syntactic change

To a certain extent, any reconstructive methodology requires an underlying theory of language change. In Neo-Grammarian phonological reconstruction, there must be a theory of sound change in order to 'undo' the relevant changes in the descendant languages to find the PIE form. The same is true, therefore, of syntactic reconstruction. Indeed, one of the central criticisms of various approaches to syntactic reconstruction is precisely that they did not incorporate a theory of syntactic change (Balles, 2008). Minimalist reconstruction, on the other hand, places a heavy emphasis on the theory underpinning language change. This is neatly summarised in the following from Pires & Thomason (2008: 42):

> Diachronic change does not exist as a primitive, atomic operation, but rather is just a convenient way...to refer to the acquisition of distinct grammars by different generations and to the fact that the output of a set of grammars 1 can function as the input triggering the acquisition of a (maybe different) set of grammars 2.

This argument must be contextualised within the broader picture of Generativist approaches to syntactic change, dating back to authors such as Andersen (1973), Langacker (1977), Timberlake (1977) and Lightfoot (1979). These authors concentrate on the language acquisition process as the locus of syntactic change. There are four relevant variables in this process: the mental grammar of the parent/caregiver (G_1) and the linguistic output of that grammar (O_1), typically referred to as the **Primary Lingusitic Data** (PLD). We then have the mental grammar of the child/acquirer (G_2), and the linguistic output of the acquirer (O_2). During the acquisition process, the acquirer must build G_2 on the basis of the patterns they observe in O_1 (the PLD); they can use G_2 to generate their own output (O_2). Crucially, however, the acquirer does not have **direct** access to G_1; rather, their connection to G_1 is mediated via the PLD. This can be visualised in the Z-model of language change:

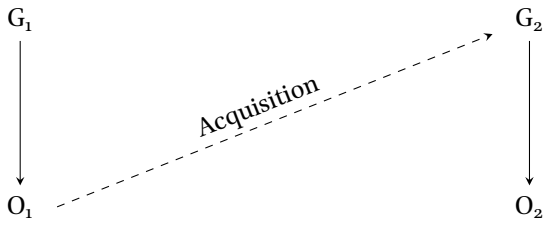


Figure 2.1: The Z-model of language change.

This model incorporates the notion that the output O_1 could be generated by either G_1 **or** G_2, which suggests that the two grammars are very similar—perhaps so similar as to be identical. However, as noted by Willis (2011: 416), '[a] given syntactic pattern may be produced by more than one possible grammar'. In other words, G_1 and G_2 can be distinct and yet produce the same surface strings. This introduces the possibility that a change could occur between G_1 and G_2 which is not visible on the surface. This is essentially what we mean by **reanalysis**: a 'change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation' (Langacker, 1977: 58). To take a textbook example from Harris & Campbell (1995: 62), consider the following string:

(2.1) *...it is bet for me / to sleen myself, than been defouled thus*
'It is better for me to slay myself than to be violated thus' (Chauc. FranT. 714–5)

In ME, *for* is strictly a preposition governing *me*, as represented in the following bracketing:

(2.2) [It is better [for me]][to slay myself than to be violated thus]

In such as string, however, *for* can be reanalysed not as a preposition governing *me*, but as a complementiser introducing an infinitival clause *me to slay myself*:

(2.3) [It is better][for me to slay myself than be violated thus]

We know this analysis holds for PDE because we can generate an alternative string of the form:

(2.4) [For me to slay myself] [would be better than to be violated thus]

Thus we may consider the original string to be **structurally ambiguous**. It could be generated by two distinct grammars: one in which *for* was strictly a preposition (G_1), and one in which it also functions as a complementiser (G_2). G_2, however, is also capable of generating a different version of the string (ex. 2.4), which is no longer structurally ambiguous, but would be ungrammatical in G_1. This is what we call **actualisation**: the reanalysis that occurs between exx. (2.2) and (2.3), originally 'invisible', is actualised in ex. (2.4).

With these terms established, we may theorise syntactic change in the following way:

1. G_1 generates a set of strings in O_1, of which a relevant subset are structurally ambiguous
2. The acquirer ascribes a different underlying structure to the subset of structurally ambiguous strings in O_1 from the one that generated them: G_2 is distinct from G_1 (**reanalysis**)
3. On the basis of the reanalysis, G_2 generates structures manifest in O_2 that could not have been generated by G_1 (**actualisation**)

This understanding of syntactic change in this way is crucial to Minimalist reconstruction. Pires & Thomason (2008: 45) argue that for our reconstruction to be sufficient in this theoretical framework for syntactic change, we must provide:

(i) hypotheses about the properties of the mental grammars that could generate the outputs possible in the daughter languages; [i.e., G_2]

(ii) hypotheses about the properties of the mental grammars that could have been internalized by speakers of the ancestral language (the grammar of the proto-language); [i.e., G_1]

(iii) hypotheses about how the different grammars of the daughter languages could have developed from the exposure of earlier generations to the output of the proposed ancestral grammar. [i.e., reanalyses]

In order to go about Minimalist reconstruction, each of these sets of hypotheses must be established. Except in cases of identity—in which case no change is hypothesised to have taken place— it is not enough to investigate a syntactic pattern in the attested languages and reconstruct by an axiom like 'majority rules'; we must provide a plausible diachronic pathway for any posited change. We may consider again the standard 'Z'-model of syntactic change, and apply it specifically to reconstruction:

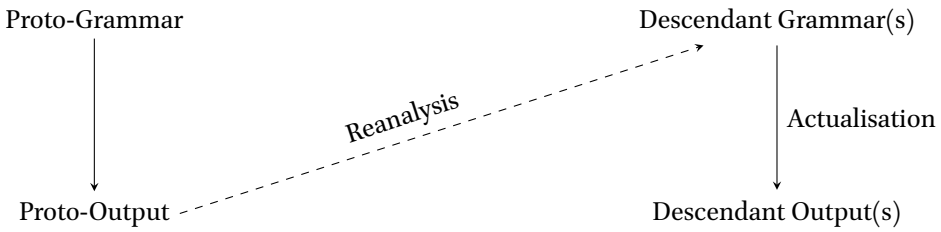


Figure 2.2: The Z-model of reconstruction (adapted from Willis, 2011: 417).

Since the syntactic change itself does not 'hop' from the Proto-Grammar to the Descendant Grammar(s), but is mediated through language acquisition and the mechanisms therein, syntactic reconstruction should not 'hop' either. For any given syntactic feature in the descendant language(s), we

should aim to trace the path of reanalys(i/e)s and extension(s) that would account for the change(s) from the proto-language to the descendant language(s). In doing so, this methodology addresses the RADICAL-REANALYSIS PROBLEM (see §1.2). Radical though any given reanalysis might be, syntactic change is by no means **random**. Willis (2011: 13, emphasis mine) makes this point succinctly: 'The output of the grammar after reanalysis approximates to that of the grammar before reanalysis, and [we must] posit **plausible and motivated reanalyses**' between the two grammars. In short: Minimalist reconstruction requires our hypotheses about the proto-language to be systematically informed by a theory of syntactic change, with an emphasis on positing plausible reanalyses, actualised in the descendant languages.

Another important component of a theory of syntactic change is **grammaticalisation**. The term dates back to Meillet (1912) and has played a key part in studies of language change across a variety of theoretical frameworks. I will adopt the definition of grammaticalisation, as phrased in Kuteva *et al.* (2019: 3), as 'the development from lexical to grammatical forms and from grammatical to even more grammatical forms.' Kuteva *et al.* (2019: 3) note that 'grammatical form' in this definition ultimately refers to a **functional category** (see §2.1.3 below). As for what makes something 'even more' grammatical means in this context, we may refer back to Hopper & Traugott (1993: 7) who suggest the possibility of a 'cline' from least to most grammatical of the following form:

(2.5) content item > grammatical word > clitic > inflectional affix

Grammaticalisation, then, is visualised as the movement along this cline. The reason that grammaticalisation is so important to theories of language change (and hence, reconstruction) is the hypothesis that it is **undirectional**. That is to say, e.g., lexical items routinely become functional, but not *vice versa*. It is certainly the case that grammaticalisation is overwhelmingly common: Kuteva *et al.* (2019) list 544 grammaticalisation pathways, attested across over 1000 language varieties. There are, however, a limited number of exceptions, typically referred to as *degrammaticalisation* (see, e.g., Joseph & Janda, 1988; Ramat, 1992; Newmeyer, 1998; Norde, 2009). For the purposes of this thesis, therefore, I do not subscribe to the strong thesis of unidirectionality, but will hold that since grammaticalisation is overwhelmingly more common than degrammaticalisation, it should factor into our reconstructive methodology.

Within the Minimalist framework, grammaticalisation receives extensive treatment in Roberts & Roussou (2003). Their definition of grammaticalisation explicitly calls upon the notion of reanalysis that is so central to Minimalist theories of language change (2003: 2, emphasis mine):

> ...grammaticalization involves the creation of new functional material, either through the **reanalysis of existing functional material** or through the **reanalysis of lexical material**.

From a theoretical perspective, therefore, grammaticalisation does not require us to posit a novel process of syntactic change to account for grammaticalisation; we are still adequately furnished with reanalysis and actualisation. The question for Roberts & Roussou (2003) is how to square the apparent unremarkableness of grammaticalisation at a theoretical level with its remarkable preponderance in attested languages. Clearly, this lies beyond the scope of the current thesis; for our purposes, it is sufficient to note that we can conceive of grammaticalisation in the same way as other forms of syntactic change,

with the added information that all else being equal, a reconstruction in which a given element was originally 'less' grammatical and become 'more' grammatical is preferable to one in which the reverse is posited.

Since grammaticalisation is not strictly unidirectional, when we use it as a reconstructional tool we are operating with probabilities. As it happens, there is an apt parallel with phonological reconstruction here, remarked upon by Willis (2011: 421):

> Universal directionality, then, is a useful tool in syntactic reconstruction, but not an infallible one. The same in fact applies in phonology: universal pathways of sound change vary according to how unidirectional they are, but we must always keep open the possibility that a particular constellation of circumstances led to an unexpected direction of change in a particular instance.

In this way Willis (2011) presents grammaticalisation as a rebuttal to Lightfoot's DIRECTIONALITY PROBLEM. This is coupled with the solution to the RADICAL-REANALYSIS problem by fitting into the broader requirement that all hypothesised syntactic changes must be supported by a detailed account about the process(es) of reanalysis involved. The question that remains, however, is: how does Minimalist reconstruction work with the Comparative Method?

2.1.2 Syntactic comparison

As noted by Watkins (1991: 178), cited in Clackson (2017: 191), 'the first law of the comparative method [is that] you have to know what to compare.' In the case of phonology, this comparison has taken the form of **correspondence sets**, lists of lexical items that are hypothesised to be etymologically cognate, on the basis of which sound changes can be posited which, when undone, unearth the proto-form. Much ink has been spilled attempting to establish an analogous correspondence set for syntax; the nature of what one considers to be a viable syntactic correspondence clearly depends on one's theoretical assumptions about the nature of syntax (and syntactic change). Outside the Minimalist program, therefore, correspondence sets have been constructed in various forms including cognate poetic formulae (e.g., Watkins, 1976, 1995), cognate sentence patterns (e.g., Harris & Campbell, 1995; Harris, 2008) and cognate argument-structure constructions (e.g., Barðdal *et al.*, 2013; Vázquez-González & Barðdal, 2019).[6] Within Minimalism, the issue has been addressed by Roberts (2007), Pires & Thomason (2008), Willis (2011), Walkden (2013, 2014), and Windhearn (2020); the solution lies in the notion of syntactic **parameters**.

Parametric comparison follows from the Generativist theory of **Principles & Parameters** (P&P). Beginning with Chomsky (1981), the central tenet of P&P is that all languages share underlying, invariant *principles*; cross-linguistic variation, therefore, is captured by a finite set of variable *parameters*. A traditional example of such a parameter is the '*pro*-drop' or 'null subject' parameter, which allows a finite verb to occur without an overt subject. This parameter is 'on' in e.g., Latin, but 'off' in modern French:

(2.6)

arma	*virum-que*	*cano*	*Troiae*	*qui*	
arms.ACC.PL.N	man.ACC.SG.M-and	sing.PRES.1SG	Troy.GEN.SG.F	REL.NOM.SG.M	
primus	*ab*	*oris*			
first.NOM.SG.M	from	shore.ABL.PL.F			

[6] For an overview of these different types of approach, see e.g. Clackson (2017), Gildea *et al.* (2020).

'I sing of arms and the man, who first from the shores of Troy...' Virg. *Aen.* 1.1

(2.7) ***Je*** *chante les héros dont Esope est le père*
I sing the heros of-whom Aesop is the father
'I sing of the heroes whose father is Aesop...' La Fontaine, *Fables* (1668)

Perhaps the primary benefit of this theory in Chomskyan terms is that it reduces the amount of information a language learner must deduce from the PLD to a set of binary choices, which are set when the speaker is exposed to the appropriate trigger environment. Moreover, the locus for cross-linguistic variation is constrained to the parameters. In the context of diachronic syntax, therefore, syntactic changes are captured by a change in parameter setting between generations; in the examples above, one would argue that at some point between Vergil and La Fontaine, the '*pro*-drop' parameter was 'switched' from [+] to [-].

The precise nature of parameters, however, is a point of contention. In the early stages of P&P, parameters were generally thought to be large in scale (so-called **macro-parameters**), taking scope over the whole language (e.g., '*pro*-drop'), and forming a distinct component of the grammar. Reasonably early on, however, some authors proposed that parameters should not be thought of as constituting an independent part of the grammar, but should instead be associated with lexical items, in the form of grammatical **features** (see §1.3). This was first suggested in Borer (1984), finding subsequent support in Wexler & Manzini (1987), eventually making its way into mainstream generativist syntax in Chomsky (1995). This association between parameters and lexical items is formulated in what is often referred to as the **Borer-Chomsky Conjecture**, defined by Baker (2008: 353) in the following way:

> Borer-Chomsky Conjecture: All parameters of variation are attributable to the features of particular items (e.g., the functional heads) in the lexicon.

This theory then breaks down the earlier macro-parameters, affecting the whole grammar, to **micro-parameters**, which affect only a subset of or individual lexical items. Moreover, rather than there being a distinct parametric component of the grammar, parameters are located within the lexicon, on the relevant functional heads. The old '*pro*-drop' parameter is therefore explained by the features of the relevant I° or D° functional head(s), which are found in the lexicon. Thus, according to Pires & Thomason (2008: 41):

> The task of syntactic reconstruction can then be restricted to identifying variation in the feature specification of (functional) lexical items that determine syntactic structure and syntactic variation.

A complementary set of proposals is found in 'Emergentist' parametric theory, which attempts to reconcile these approaches in order to keep the helpful generalisations of the macro-parameters with the flexibility of the micro-parametric approach. Starting with Biberauer (2011), continued in Biberauer & Roberts (2012a,b), Biberauer & Roberts (2017) etc., the 'Emergentist' view of parameters introduces an 'informal' taxonomy of the following description (Biberauer *et al.*, 2014: 109):

> For a given value v_i of a parametrically variant feature F:
>
> a. **Macroparameters**: all functional heads of the relevant type share v_i

b. **Mesoparameters**: all functional heads of a given naturally definable class, e.g. [+V], share v_i
c. **Microparameters**: a small subclass of functional heads (e.g,. modal auxiliaries, pronouns) shows v_i
d. **Nanoparameters**: one or more individual lexical items is/are specified for v_i

This taxonomy opens up various possibilities for the application of parametric comparison to syntactic reconstruction. One may build a correspondence set based on a macro-parameter (e.g., head-directionality, Roberts, 2007; Windhearn, 2021), or an extensive set of meso- and micro-parameters (see, e.g., Ram-Prasad & Roberts, 2022). As we work our way down the hierarchy, parametric correspondence starts to approximate phonological correspondence: the point of comparison is once again, lexical items. The difference is that in phonological correspondence it is the phonetic content of lexical items that is to be compared, whereas in syntactic reconstruction it is the grammatical features of those lexical items that are to be compared. This is the approach taken by Willis (2011) for Common Brythonic *bynnag* in free relatives and by Walkden (2013) for the early Norse 'middle' suffix *-sk*; it is also the approach I take to the reconstruction of the PIE relative pronoun *REL in this thesis in §3 and §5.

While there is an elegance in collapsing syntactic and phonological correspondence sets as simply the behaviour of lexical items, it is not a perfect solution. Walkden (2013) suggests that unlike phonology, syntax fails to meet what he calls the 'Double Cognacy Condition' (2013: 101):

> Double Cognacy Condition: In order to form a correspondence set, the contexts in which postulated sounds occur must themselves be cognate.

Walkden notes that we can proceed with phonological reconstruction only by comparing cognate phonemes within cognate words. This means there are two 'levels' of correspondence: the 'abstract' phoneme and the 'concrete' word. While the comparison of parameter values, *qua* feature specifications of lexical items, may give 'abstract' correspondence, there is no 'concrete' level of correspondence; sentences, unlike words, are not inherited, and so cannot be cognate. In other words, what we typically refer to as correspondence sets in phonology—that is, cognate word lists—cannot be formed in syntax (at least, not in precisely the same sense of the term): there will necessarily be layer of abstraction in syntactic comparison that is not present in phonological comparison. This presents a categorical difference between the Comparative Method as applied to phonology and its application to syntax. Instead of looking for 'concrete' correspondence in syntax, Walkden (2013: 111) argues that '[e]vidence must instead be adduced from distributional patterns of the lexical items in question, i.e. the syntactic environments in which they can be found in the daughter languages.'

In the context of this thesis, therefore, we must establish the functional category which we consider to be cognate (e.g., **yó-* or **kʷí-/*kʷó-* in their function as a relative pronoun); we must then examine its distributional patterns within the descendant languages; on the basis of our observations, following the set of requirements established by Pires & Thomason (2008) in §2.1.1, we reconstruct the grammatical features of this functional category in PIE. In this way, we are not building a correspondence 'set', akin to those in phonological reconstruction. Rather, we are building upon a categorically different type of correspondence. I therefore concur with Clackson (2017: 203) that 'syntactic reconstruction is a

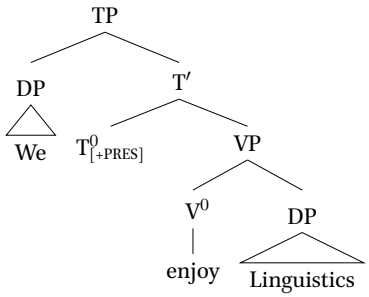
different type of enterprise from phonological reconstruction'. The methodology I adopt in this thesis is one in which parameters, located on lexical items, as the theoretical locus of syntactic variation, represent the point of comparison; they are thus a solution to the CORRESPONDENCE-SET PROBLEM, though I do not treat them as a correspondence 'set'.

2.1.3 Functional categories

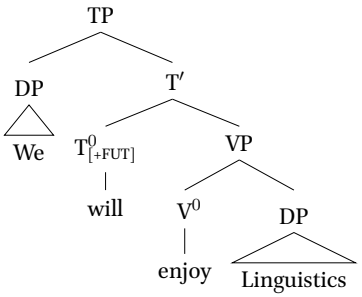
Although in the main I accept the methodology of Minimalist reconstruction as outlined above, I believe a more precise explanation of the role 'Functional categories' should play in our reconstruction is needed. I agree that they should be the focus of our reconstructions: Functional categories, and their behaviour, are effectively what grammar is, in Minimalism. Nevertheless, though Functional categories may be the end of syntactic reconstruction, they may not necessarily be the beginning.

In Minimalist theory, the lexicon is divided into 'Lexical' and 'Functional' categories.[7] The former includes nouns, lexical verbs, adjectives and adverbs; the latter, determiners, auxiliary verbs, prepositions and complementisers. Functional categories can be phonetically realised, but they may also be phonetically null. Consider, e.g., tense markers in English:

(2.8)



(2.9)



Accordingly, items from the lexicon fall into three categories: Lexical, Functional (overt), and Functional (silent). Of these, Functional (overt) items present the golden combination for syntactic reconstruction; their phonetic content can be reconstructed, to which we may add our reconstruction of their grammatical features. Common Brythonic *bynnag* (Willis, 2011) and early Norse **-sk* (Walkden, 2013) fall into this ideal category. Yet such ideal candidates for feature-based reconstruction, to the extent that they can be found for PIE at all, present only a very limited picture of its syntax. To present as full a picture as possible of the proto-grammar, therefore, we must also aim to reconstruct the syntactic features of Functional (silent) items.

To do this, we must look at the syntactic behaviour of Lexical items. For example, in §4, I will provide extensive discussion of the left periphery of PIE. The functional heads responsible for, e.g., Topicalisation (Top°) and Focalisation (Foc°), are not (always) themselves visible. Rather, their effects are visible because they attract overt lexical elements to their specifiers (i.e., they bear 'strong' features, see

[7] In this subsection I use capitalised forms of these words to indicate their meaning within Minimalism rather than their general sense.

§1.3). Take, for example, the category of 'local particles', or 'P-words'.[8] Though the exact syntactic categorisation of P-words is somewhat obscure, in the earliest stages of these languages they bear most similarity to adverbs, making them more Lexical than Functional. One syntactic pattern associated with P-words is fronting to sentence-initial position, as in the following examples:[9]

(2.10) ἐν γὰρ Πάτροκλος φόβον ἧκεν ἅπασιν
in.PW for Patroclus.NOM.SG.M fear.ACC.SG.M send.AOR.3SG all.DAT.PL
'For Patroclus had set fear **into** them all' (Hom. *Il.* 16.291)

(2.11) ***prá*** *yáḥ* *vām* *mitrā* *varuṇā* *jūráḥ*
forth.PW REL.NOM.SG.M you.ACC.DU mitra.VOC varuna.VOC quick.NOM.SG.M
dūtáḥ *ádravat*
messenger.NOM.SG.M run.IMPF.3SG
'The quick messenger who ran **forth** to you, O Mitra and Varuna' (RV 8.101.3ab)

I have previously argued that in Sanskrit, P-words are base-generated within the *v*P, and subsequently fronted to the left periphery (Ram-Prasad, 2018). In §4.2 of this thesis, I argue specifically that in Vedic, the functional head responsible for this movement is Foc°. In this case, the Functional category (Foc°) itself is not manifest, but its presence can be inferred from the behaviour of Lexical items (P-words). If we were to hypothesise that the same process is at play in other IE languages with fronted P-words, we may reconstruct the silent Functional category Foc° for PIE, despite no direct evidence for its existence.

That we can and should reconstruct silent grammatical heads on this basis is implicit in Walkden (2014), especially in his account of V2 and V3 structures in Proto-Germanic.[10] The issue is treated more explicitly in Windhearn (2020: 11–18), who argues for the possibility of reconstructing the features of C° in PIE even in the absence of a single complementiser that can be securely reconstructed phonologically; his point is rather different, however, as he argues that this reconstruction is possible because overt complementisers are still attested in the descendant languages. This is not the case for other functional heads such as Top° and Foc°, which may also lack phonetic content in the attested languages.

At a theoretical level, within Minimalism the distinction may actually seem quite trivial; silent heads are *de facto* part of what syntax is. Setting aside theoretical prejudice, however, one may argue that for syntactic reconstruction to be as reliable as phonological reconstruction, we must stick with the lexical (small-l) items we have already established as cognates; necessarily, therefore, we must work with those items that have reconstructable phonetic content. Such objections, though they may seem on the face of it to address syntactic reconstruction, are in fact tantamount to criticisms of silent heads *per se* (and, by extension, the Minimalist framework). Even in contemporary attested languages, the presence

[8] I take this terminology from Bertrand (2014).

[9] In a break with convention, sandhi is dissolved in all Sanskrit examples except where expressly noted, to facilitate glossing. I cannot provide the original form of the text owing to issues of space.

[10] See in particular Walkden (2014: 89), where he 'assume[s]...the relevant left-peripheral heads—Shift°, Foc°, and Fam° in particular—are all cognate [in the relevant languages]...on grounds of formal similarity alone.' On the other hand, for many working in the Minimalist framework, the inventory of functional heads is universal (see Ramchand & Svenonius, 2014 for discussion), so the question of whether to reconstruct a specific head in a specific language is vacuous.

of silent heads can only be inferred by the behaviour of phonologically overt content. I have already noted that there is a necessary level of abstraction for syntactic comparison: in fact, I would argue that abstraction is necessary for a syntactic analysis of any sort. We should not be troubled, then, when such abstraction turns up in syntactic reconstruction; it is simply part of what syntax in the Minimalist framework looks like.

In light of this acceptance, we must revisit the role Functional categories play in our methodology of syntactic reconstruction: it becomes clear that we need not limit ourselves to working with Functional morphemes as our starting point. Furthermore, insofar as their presence might be inferred from the behaviour of other lexical items, silent Functional heads can and should be reconstructed. There are ultimately two intertwined implications from this:

- We can start syntactic reconstruction by examining the syntactic behaviour of Lexical items as well as Functional ones.
- To explain the syntactic behaviour of Lexical items, we may allow ourselves to posit silent Functional heads

These are my two theory-internal adaptations of the Minimalist reconstructive method; I do not believe these are particularly problematic, but points of methodological importance that should be addressed directly within this framework.

2.1.4 Uniting old and new

There remains a final point of methodology to make explicit, which is that despite the various allowances we must make for syntactic reconstruction that are not required for phonological reconstruction, Minimalist reconstruction can complement, rather than contradict, the 'traditional' conception of the Comparative Method.

As noted in §2.1.2 above, the Borer-Chomsky conjecture effectively locates a language's grammar entirely in the lexicon. The response to this conjecture in Minimalism is to ascribe a set of 'features' to each lexical item. The value and 'strength' of these features are what motivates syntactic movement, and hence word order (see §1.3). In entirely theoretical terms then, Minimalism ascribes a great deal of syntactic power to lexical items. In particular, functional heads, such as determiners, tense markers and complementisers, are the primary driving force behind word order.

However, coming from "the other side"—i.e., Neo-Grammarian phonological reconstruction—lexical items are often not the beginning, but rather, the end of reconstruction. The Comparative Method, as it is used for phonology, allows us to posit proto-morphemes.[11] In the case of functional morphemes (e.g., inflectional endings), by observing the use of their descendants in the daughter languages, we may hypothesise some of their grammatical functions in the proto-language, e.g. the morpheme *-*mi* marks

[11] I include both lexical roots and bound morphemes in this descriptor.

the first person singular present tense on athematic verbs. Such a hypothesis can be 'translated' in a Minimalist feature matrix applied to this morpheme, e.g.:[12]

(2.12) *-*mi* = tense:present, *i*Num:sg, *i*Pers:1, etc.

Equipped with lexical items and their feature matrices, we now find ourselves at the door of Minimalist reconstruction. The narrow syntax, i.e. the operation MERGE, takes these lexical items and combines them into hierarchical structures. Such structures—established using the methodology of Minimalist reconstruction detailed above—form our syntactic reconstruction. Of course, the difficulty lies precisely in securing the reconstructed feature values for various lexical items, especially when there may be 'strong' features that drive syntactic movement. But the point stands that, owing to Minimalism's focus on the behaviour of lexical items as the locus of of syntactic features, Minimalist reconstruction represents a knitting-together of philological and theoretical approaches to proto-languages.

This methodological consilience is particularly apt for the purpose of this thesis: reconstructing the relative syntax of PIE. An old and quite persistent debate concerning PIE relative clauses surrounds the status of the two possible relative pronouns, **yó-* and **kʷí-/*kʷó-* (see §3). In the descendants of these forms, therefore, we are presented with a functional category (the relative pronoun) whose syntactic behaviour can be compared. In other words, while phonological comparison can help us establish **yó-* and **kʷí-/*kʷó-* as possible functional categories in PIE, Minimalist reconstruction allows us to posit the syntactic features associated with these lexical items, which are to be compared. In fact, as I will argue in §3, I believe that when used as a relative pronoun, **yó-* and **kʷí-/*kʷó-* can be treated as representing the same functional category (viz. the relative pronoun) in PIE. I reach this conclusion precisely by an extensive comparison of the descendants of **yó-* and **kʷí-/*kʷó-*, which mirrors the sort of lexically-driven correspondence that characterises phonological reconstruction.

2.2 Relative clauses

In this section, I discuss the various Minimalist accounts of relative clause (RC) syntax. I start with a brief summary of the different semantic types of RC, which may also correspond to different syntactic types. I then turn to the different syntactic models of RCs in Minimalist theory, including both 'plain' relatives (those that are not part of a relative-correlative structure) and correlatives (those that are).

2.2.1 Semantic types

A basic distinction that has long been observed is that RCs can be either **restrictive** or **non-restrictive**—the latter type is also referred to as **appositive**. Restrictive relative clauses (RRCs) are so-called because the RC 'restricts' the set of entities evoked by the antecedent. For example, in the sentence:

[12] Ascribing the appearance of inflectional morphemes to the syntax, rather than the lexicon, is the central thesis of Distributed Morphology (DM) as proposed by Halle & Marantz (1993, 1994) (see Siddiqi (2010) for a recent overview). Although I have done so in this example, in the remainder of this thesis I do not commit myself explicitly to DM, since morphological derivations are not my research focus.

(2.13) The books [that I like] are novels

The antecedent denotes the set of books, while the [RC] restricts this set to only the ones liked by the speaker.[13]

For ARCs, however, the identity of the subject is already known, and the material in the RC provides only extra descriptive content, rather than a restriction. For example, in the sentence:

(2.14) *A Tale of Two Cities*, [**which** I like], is a novel.

Here, the antecedent already tells us exactly which book is being referred to—there is no 'set' of books called *A Tale of Two Cities*, to be restricted by the speaker's preference. Rather, the [RC] here does not impose any set-based restriction, but provides extra information, non-essential for the purposes of identifying the referent.

In Present-Day English, there are two major strategies for identifying RRCs vs ARCs:

- ARCs can be preceded and followed by intonational gaps, denoted by the commas in ex. (2.14).
- ARCs can only use *who/which* as a relativiser, while RRCs can use *who/which*, *that* or Ø.

Such diagnostics vary cross-linguistically, and can be absent altogether, as in, e.g. Japanese (Andrews, 2007: 207). This is worth bearing in mind when we reconstruct these categories in PIE.

Further to this distinction, Grosu & Landman (1998) identify RCs of a 'third kind', 'characterized by an interpretive operation of **maximalisation**', and are referred to as **Maximalising RCs** (MaxRCs) (Alexiadou *et al.*, 2000: 21). What is meant by 'maximalising' in this context is described by Probert (2015: 77) as 'pick[ing] out a unique entity, everything in a set, or a complete lot of stuff.' Consider, for example, the following:

(2.15) **Whoever** thinks that is wrong.

(2.16) I gave him **what** coins I had.

(2.17) I drank **what** water was in the glass.

These sentences could be rephrased in the following way:

(2.18) **The** person who thinks that is wrong.

(2.19) I gave him (all) **the** coins I had.

(2.20) I drank (all) **the** water that was in the glass.

But not, e.g., in the following way:

(2.21) **Some** people who think that are wrong.

(2.22) I gave him **some** of the coins I had.

[13] This conception of RRCs as denoting a set which combines (i.e. intersects) semantically with the head noun is detailed in Partee (1973).

(2.23) I drank **some** of the water in the glass.

In other words, MaxRCs have 'built into them...a meaning similar to that of the definite article' (Probert, 2015: 73); it is this meaning that Grosu & Landman class as 'maximalising'. MaxRCs always give a reading that selects the maximal value for the entity to which they refer: if there is is one entity, it is unique (cf. **the** person); if there are multiple entites, the MaxRC denotes the entire set (cf. **all** the coins); if it is a continuous variable, the MaxRC denotes the 'complete lot of stuff' (e.g. **all** the water).

Within the category of MaxRCs, there are three subtypes (according to Grosu & Landman, 1998). Two of these we have already encountered:

- Free relatives (FRs): Traditionally described as having an 'implied antecedent', FRs do not have an overt external head noun, only the relative pronoun, cf. ex. (2.15).
- Degree relatives (DegRCs): First discussed in Carlson (1977) and Heim (1987), DegRCs 'invole abstraction over degrees or amounts, rather than individuals' (Alexiadou *et al.*, 2000: 28). Cf. exx. (2.16–7)

The third type of MaxRC (according to Grosu & Landman, 1998), **correlatives** (discussed extensively in §2.2.2.2 below) are an unmarked relativisation strategy in Hindi and other modern Indo-Aryan languages. For correlatives, in addition to the relative pronoun, there is a correlative (demonstrative) pronoun that occurs in the Matrix Clause (MC), such that the RC itself does not occupy a DP position within the MC. Correlatives play a crucial part in my reconstruction of RC structure in PIE (§5.2), since they are very common in ancient IE languages, e.g.:

(2.24) *ex* *malis* *multis* *malum* ***quod*** *minimum-st,*
from evil.ABL.PL.N many.ABL.PL.N evil.NOM.SG.N REL.NOM.SG.N least.NOM.SG.N-is
id *minim-est* *malum*
DEM.NOM.SG.N least.NOM.SG.N-is evil.NOM.SG.N
'**The thing which** is least evil from among the many evils, **that thing** is the least (like) an evil.'
(Plaut. *Stich.* 120)

(2.25) *ágne* ***yám*** *yajñám* *adhvarám* *viśvátaḥ*
Agni.VOC REL.ACC.SG.M worship.ACC.SG.M sacrifice.ACC.SG.M everywhere
paribhū́ḥ *ási* ***sáḥ*** *ít* *devéṣu* *gacchati*
surrounding.NOM.SG.M be.PRES.2SG DEM.NOM.SG.M EMPH god.LOC.PL.M go.PRES.3SG
'Agni, the worship and sacrifice **which** you surround on every side, **that** indeed goes to the gods.'
(RV 1.1.4)

(2.26) ***hὸς*** νῦν ὀρχēστῶν πάντῶν ἀταλό́τατα παίζει
REL.NOM.SG.M now dancer.GEN.PL.M all.GEN.PL.M most-friskily dance.PRES.3SG
τō τοδε κ[
DEM.GEN.SG.M DEM.NOM/ACC.SG.N ?
'**Whichever** of all the dancers now dances most friskily, **of him** this...' (*IG* I² 919)

I will contrast correlatives with **'plain' relatives**, which includes those which do not contain a correlative demonstrative in the MC. As I will demonstrate in §2.2, the presence of a correlative demonstrative is only one of several syntactic features that distinguish correlatives from 'plain' relatives.

In the Grosu-Landman typology, correlatives are classed as MaxRCs; they argue that 'correlates [are] restricted to definites and universals, and other effects' (Grosu & Landman, 1998: 168). While this appears to be the case in Hindi, taken synchronically as the canonical correlative language, it is not the case in older European languages; indeed, appositive correlatives are reasonably common in Vedic Sanskrit (Hettrich, 1988).

(2.27) ***yáḥ*** *rāyáḥ* *avániḥ* *mahā́n* *supāráḥ*
REL.NOM.SG wealth.NOM.PL stream.NOM.SG.F great.NOM.SG.M well-crossing.NOM.SG.M
sunvatáḥ *sákhā* ***tásmai*** *índrāya*
one-who-presses-out-soma.GEN.SG.M friend.NOM.SG.M DEM.DAT.SG.M Indra.DAT.SG.M
gāyata
sing.IMPV.2PL
'The one **who** [is] the great stream of riches, easy to cross, and the friend of the one who presses the Soma, to **that** Indra, sing!' (RV 1.4.10)

This is accounted for by Davison (2009: 153) by positing semantic feature [PRED] on C°, present in Hindi, that requires any given correlative clause to be a MaxRC. This feature is apparently absent in Sanskrit,[14] and so for the purposes of this study, therefore, I will amend the Grosu-Landman model. Rather than classing correlatives as a subset of MaxRCs, I will treat them as a separate type of RC altogether, without any strong commitment to their semantic value.

To summarise: RCs can be restrictive, maximalising, or appositive. At this stage, I do not commit to tying any particular semantic type to a particular syntactic formation, though as I will discuss below and return to in §5.4, the relationship between semantic type and syntactic structure is not arbitrary.

2.2.2 Syntactic types

In this section, I evaluate the competing Minimalist analyses of 'plain' relatives and correlatives. The first category includes quite a large set of different syntactic types, including internally-headed vs. externally-headed RCs, as well as pre- vs. post- vs. circum-nominal RCs. See, e.g., Lehmann (1984), de Vries (2002) and Cinque (2020) for major cross-linguistic overviews of these different types under varying theoretical assumptions. For the purposes of the following discussion, I will focus on what are traditionally referred to as **externally-headed postnominal RCs** as examples of 'plain' relatives, as these are the main type found in IE languages.

2.2.2.1 'Plain' relatives

Generativist theories of RC syntax can be divided into broadly three analyses, which vary concerning the details of the syntactic relationship between (a) The matrix clause (MC) (b) the head noun (antecedent) (c) the relative clause (RC):

[14] There is some doubt here, and I return to the question in §5.6.

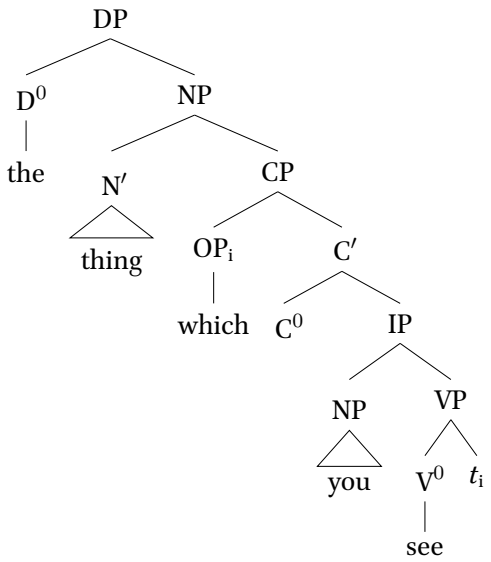
1. **Head External Analysis (HEA)** (Chomsky, 1981):[15]
   (a) *Adjunction hypothesis*: RCs are adjoined to NP or DP
   (b) *Base-generated head hypothesis*: The head noun of an RC is base-generated outside the RC
2. **Head Raising Analysis (HRA)** (Brame, 1968; Schachter, 1973; Vergnaud, 1974; Kayne, 1994; Bianchi, 1999, 2000; de Vries, 2002):[16]
   (a) *Determiner complementation hypothesis*: The RC is the syntactic complement of D°
   (b) *Head-raising hypothesis*: The head noun raises from a structural position within inside the RC
3. **Matching Analysis (MA)** (Lees, 1960, 1961; Sauerland, 1998, 2003; Citko, 2001; Salzmann, 2006, 2017; Douglas, 2016):[17]
   (a) *Base-generated head hypothesis*: The head noun of an RC is base-generated outside the RC
   (b) *Matching hypothesis*: There is a complex representation of the head noun **within** the RC, which is deleted at PF under identity

To illustrate the difference between these theories, consider the following RC:

(2.28) The thing **which** you see

Under the HEA, the structure posited for this would be:

(2.29)



The functional head associated with 'which' in ex. (2.29) is referred to as the relative **operator** (OP); in English RRCs such as ex. (2.29), OP can be null, in which case the declarative complementiser 'that' can be optionally pronounced.

[15] According to Douglas (2016: 10, n.5), 'This was arguably the mainstream approach in the literature until Kayne's (1994) reintroduction of the HRA. Bhatt (2015: 728–9) notes that the HEA was so widespread in the literature that its origins are obscure'.

[16] See Bhatt (2015: 727–8) and Douglas (2016: 10) for further references.

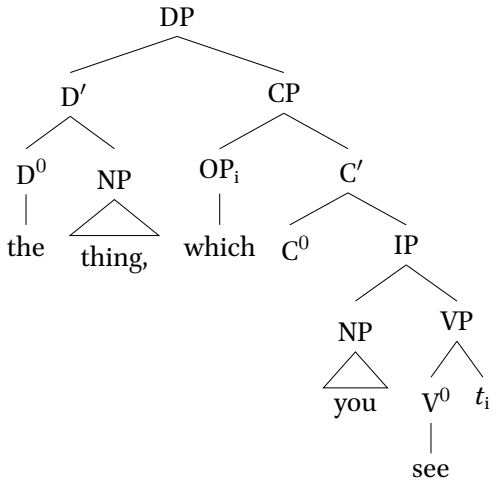
[17] See Bhatt (2015: 729) and Douglas (2016: 12) for further references.

(2.30) The thing OP_i (that) you see t_i

OP originates as a complement of VP, and raises to a position in [Spec, CP]. Here, it turns the CP into a predicative expression.[18]

Under the HEA, the difference between RRCs (and MaxRCs) and ARCs can be encoded in the syntax: namely, by what constituent the RC adjoins to. There is some disagreement between authors on where exactly the adjunction happens (cf. Browning, 1991: 51), but it has been argued that the difference between RRCs and ARCs can be encoded in terms of scope: for RRCs, the external D° takes scope over the relativised noun phrase. For instance, '**the** coins that I have' selects every item in the intersection between the sets COINS and THINGS THAT I HAVE, but '**some** coins that I have' selects only a subset of these items. On the other hand, for ARCs, '[t]he semantic effect of the...article...is applied to the antecedent noun phrase in a way that logically precedes the modification of this noun phrase by an [ARC]' (Probert, 2015: 69). So, in '**the**/**some** coins, which I have' the article does not take scope over the RC, as it has already 'applied' its semantic constraints to the head noun before relativisation. On a theory where scope = c-command,[19] therefore, ARCs should adjoin higher up in the structure such that the external D° does not c-command that RC, and therefore does not take scope over its contents. Thus, while RRCs adjoin to NP, as in ex. (2.29) above, ARCs adjoin at the DP level (Demirdache, 1991), as in (2.31) below.

(2.31)



However, as noted above, the diagnostics that distinguish between RRCs vs ARCs are language-specific. There are languages in which there is no distinction in the surface forms of RRCs and ARCs, including Basque and Japanese (Bianchi, 1999: 136).[20] It is also possible to encode the semantic content of ARCs in other ways, such as treating relative pronouns as E-type pronouns (I discuss this possibility for PIE in §5.6). For this reason, hard-wiring a distinction into the syntactic derivation, in the form of adjunction level, is not the only way of capturing the interaction of semantics and syntax here.

[18]There is, however, much discrepancy on what OP actually is. It could be PRO (Chomsky, 1981), *pro* (Browning, 1991) or something else altogether. For the purposes of this discussion, I have treated it as a generic D°—in this specific example, it is overt in the form of the relative pronoun 'which'.

[19]See Szabolsci (2001) for theoretical background, and Horrocks (2011) for an implementation of this approach to RCs in Latin.

[20]See further del Gobbo (2017) and Cinque (2020) on the difference between 'integrated' and 'non-integrated' ARCs, which adds a further level of complexity here.

Moreover, there are several weaknesses to the HEA. These are addressed in detail across the authors who favour either the HRA or the MA to the HEA (see references above). The earliest arguments against the HEA and for the HRA, as put forward by Brame (1968), Schachter (1973) and Vergnaud (1974), involve **reconstruction** effects,[21] where the relativised head noun needs to be understood as originating within the RC in order to bind an anaphor, here exemplified with the reciprocal pronoun *each other*:

(2.32) The [interest in [each other]$_i$]$_j$ that [John and Mary]$_i$ showed t_j was fleeting
(Schachter, 1973: 43a)

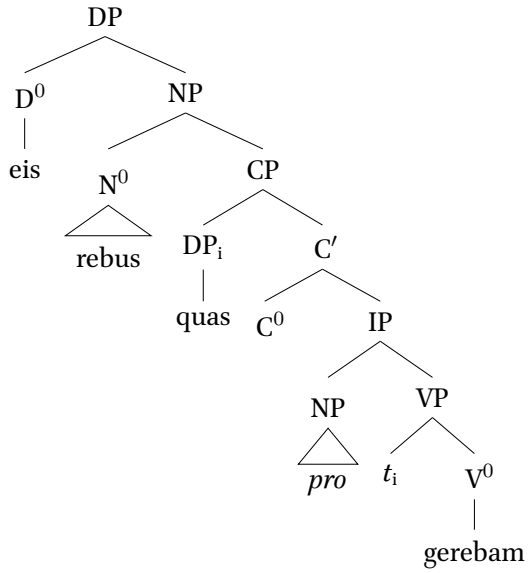
Unfortunately, reconstruction effects of this sort (and other locality constraints) are hard to diagnose in the ancient IE languages.[22] More relevant criticism, for the purposes of this thesis, can be found in Bianchi (2000), who argues against the HEA and in favour of the HRA with reference to ancient IE languages. Her first argument pertains to the status of the relative pronoun itself. Bianchi discusses the following example:

(2.33)

[*animum*	*meum*]	*esse*	*in*	*hoc*	*corpore*	*ex*	*eis*
soul.ACC.SG.M	my.ACC.SG.M	be.INF	in	DEM.ABL.SG.N	body.ABL.SG.N	from	DEM.ABL.PL.F
rebus	*quas*	*gerebam*	*intelligebatis*				
thing.ABL.PL.F	REL.ACC.PL.F	do.IMPF.1SG	understand.IMPF.2PL				

'You understood that my soul was in my body from the deeds I performed.' (Cic. *Sen.* 79)

(2.34)



Here, the relative pronoun *quas* acts as an 'independent pronoun anaphoric to the antecedent NP' *rebus* ('things') (Bianchi, 2000: 56). Bianchi compares this use of the relative pronoun with that found in **correlative** clauses (discussed in detail below). In correlative RCs, rather than acting 'anaphorically', relative *qui* seems to act rather as a determiner that selects the relativised a head noun (Bianchi, 2000: 56):

[21] See, e.g., Douglas (2016: 19–21) for an introduction, and (2016: 22–58) for a detailed discussion of reconstruction effects and their implication in RCs. This meaning is entirely unrelated to the *reconstruction* of a proto-language.

[22] See, e.g., Bertocchi (1989) on Latin and Sevdali (2013) on Ancient Greek. Both authors argue that Binding Theory, as developed to account for English, cannot accommodate the facts from these languages without some modification.

(2.35) *Quibus* *diebus* *Cumae* *liberatae* *sunt* *obsidione,*
REL.ABL.PL.M day.ABL.PL.M Cumae.NOM.PL.F free.PPP.NOM.PL.F be.PRES.3PL siege.ABL.SG.F
isdem *diebus...* *Ti. Sempronius...* *prospere* *pugnat*
same.ABL.PL.M day.ABL.PL.M Titus-Sempronius.NOM.SG.M victoriously fight.PRES.3SG
'In the days in which Cumae was freed from siege, in the same days... Titus Sempronius fought victoriously' (Livy *Ab urbe cond.* 23.37)

In this example, *quibus* fulfils the same syntactic role as *eis* does in (2.33)—i.e., it is a determiner selecting a noun (*diebus*, 'days')—and apparently not the same role as *quas* in ex. (2.34), which Bianchi takes to be an 'independent pronoun'. Bianchi argues that this apparent dual function of the relative pronoun (or OP) is an unresolved issue in HEA, which is more generally inept to account for correlatives (on which see §2.2.2 below).

Second, Bianchi treats issues of case attraction. Case attraction phenomena can lead to either the relative pronoun or the head noun surfacing in a case that appears to be at odds with our expectations (Smyth, 1956: §2522, §2533):

(2.36) ἄξιοι τῆς ἐλευθερίας ἧς κέκτησθε
worthy.NOM.PL.M the.GEN.SG.F freedom.GEN.SG.F REL.GEN.SG.F obtain.PERF.2PL
'Worthy of the freedom which you have obtained' (Xen. *An.* 1.7.3)

Usually, the relative pronoun takes its case from the syntactic role it plays in the RC. In ex. (2.36), as the direct object of κέκτησθε ('you have obtained') we expect the relative pronoun to occur in the accusative case. Instead it appears in the genitive, because it has been "attracted" to its antecedent ἐλευθερίας ('freedom'), which stands in the genitive because it is governed by ἄξιοι ('worthy'). It is worth noting, however, that case attraction phenomena are fairly rare: while certain Classical Greek authors (e.g., Xenophon) exhibit it optionally, it is altogether absent at the earliest stages of the language (Probert, 2015: 167–70); Bianchi (1999, 2000) cites evidence from Classical Latin, but this is highly exceptional (Pinkster, 2021: 489–90).

The opposite, however—so-called "inverse attraction"—is more common.[23] This is where the head noun is "attracted" to the case of the relative pronoun:

(2.37) ἔλεγον ὅτι Λακεδαιμόνιοι ὧν δέονται **πάντων**
say.AOR.3PL that Spartan.NOM.PL.M REL.GEN.PL.N need.PRES.3PL all.GEN.PL.N
πεπραγότες εἶεν
gain.PERFP.NOM.PL.M be.OPT.3PL
'They said that the Spartans had gained everything which they needed.' (Xen. *Hell.* 1.4.2.)

Here, πάντων ('everything') is the direct object of πεπραγότες εἶεν ('(they) gained'), and should stand in the accusative. Instead, it appears in the genitive; it is attracted to the case of the relative pronoun ὧν, which stands in the genitive as the complement of δέονται ('they need').

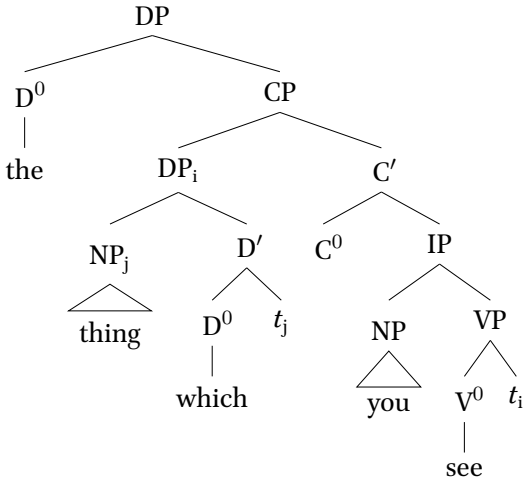
Bianchi (2000: 58-9) argues that both these phenomena are problematic for the HEA. With regard to the 'normal' case attraction in ex. (2.36), it is not clear how the higher D° (τῆς) can assign case

[23] Insofar as it is the regular case pattern for correlative RCs.

to the lower D° (ἧς).[24] The example of 'inverse' attraction in ex. (2.37) is even stranger: how does the QP containing πάντων ('everything') receive its case from a verb δέονται ('they need') supposedly within its own adjunct?

For these reasons, Bianchi argues, we should abandon the HEA, and instead adopt the HRA per Kayne (1994).[25] This is schematised below:

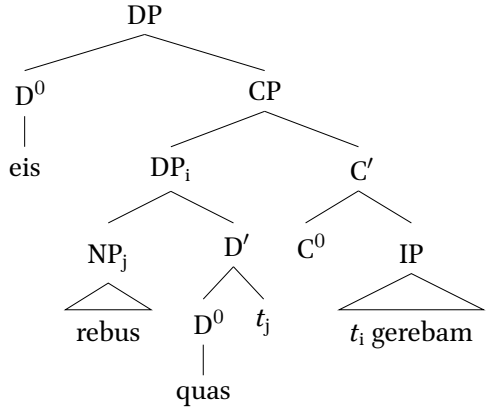
(2.38)



There are two central differences between the HRA and the HEA. First, note that the relative CP, rather than adjoining to the head NP/DP, appears as a complement of the external D° ('the'). The head noun of the RC ('thing'), therefore, belongs to the internal DP, which is base generated in the relative CP. The observed linear order is then derived by two movements: first, the internal DP ('which thing') raises to [Spec, CP] of the RC; then the NP ('thing') to [Spec, DP]. In this account, then, the head noun forms a constituent with the relative pronoun, and not with the external D°. In other words, [the thing] is not a constituent, but [thing which] is.

Recall the aforementioned issue with apparent dual syntactic role of the relative pronoun *qui* in exx. (2.34) and (2.35). Under the HRA, the two are collapsed:

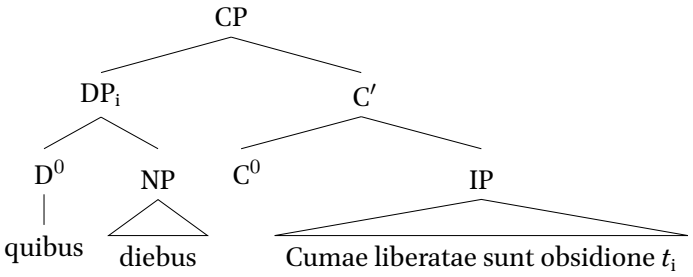
(2.39)



[24] See Bianchi (2000: 59), and further Bianchi (1999: 54-61) for a more detailed discussion of the issues here in each of the various theories underlying agreement phenomena.

[25] Another set of motivations are given against the HEA if one adopts the Linear Correspondence Axiom (Kayne, 1994), discussion of which is beyond the scope of this thesis but see Cinque (1996), Rochemont & Culicover (1997), Haider (2000), Abe (2001), Takano (2003), Ernst (2004) for general discussion.

(2.40)



In both cases, the relative pronoun is a D° that selects an NP complement; the difference in word order is explained by the movement of the NP *rebus* to [Spec, DP] in ex. (2.34). This movement is driven by the external D°, which is not present in correlative RCs such as in ex. (2.40).[26] The apparent lack of agreement in case between *rebus* (ABL) and *quas* (ACC) in ex. (2.34) is also accounted for by this movement: the DP is base generated as *quas res* (ACC), but when the whole phrase raises to [Spec, CP], and the NP raises to [Spec, DP], the external D° assigns ablative case to the relative NP.

I will return to the relation between 'plain' relatives and correlatives in §5. For now, however, note that Bianchi's first issue concerning the status of the relative pronoun has been resolved: it is always a D°. Moreover, "inverse case-attraction" follows logically from the HRA, and is neither "inverse" nor "attraction"; it is simply the retention of the base-generated case, without influence from an external D°. However, the "normal" case pattern, where the head NP and the relative D° can disagree in case requires some further explanation. According to Bianchi (1999, 2000), the NP bears the case of whichever D° it is governed by at the **end** of the syntactic derivation. In other words, morphological case agreement is post-syntactic, and takes place at PF. As Salzmann (2017: 93) notes, this is not impossible, but case agreement does not seem to bear any other distinctive features of a PF process, such as a unique sensitivity to linear order; on the contrary, morphological agreement in Latin (and other ancient IE languages) is expressly **in**sensitive to linear order, as evidenced in the preponderance of discontinuous constituents in structures that often fall under the traditional category of *hyperbaton* (see, e.g., Horrocks, 2011). Bianchi (1999: 94–5) draws support for this possibility by arguing that conventional "case attraction" also appears to show this sort of late stage agreement, but it is not immediately clear that attraction, which is at best optional and often absent altogether, should be accounted for under the same mechanism as the "regular" pattern.

Thus, although the HRA explains several issues that are not dealt with by the HEA, various issues remain. Detailed discussions are given in Sauerland (1998, 2003); Salzmann (2006, 2017); Douglas (2016), *inter alia*. While these criticisms may include case behaviour as discussed above, there are other issues too. Salzmann (2017: 96–102) offers an extensive critical analysis of Bianchi (1999, 2000), identifying five flaws with Bianchi's account (Salzmann, 2017: 101):

1. The lack of reconstruction effects in some RCs[27] suggests the possibility that RCs can merge late

[26] There is some debate as to whether correlative RCs have an external D°—I return to this in §2.2.2.2 below.

[27] E.g., *The picture of John$_i$ that he$_i$ likes*. To quote Douglas (2016: 29): 'If reconstruction occurred, the R-expression [i.e., *John*] in the RC head would be c-commanded and bound by a co-indexed element [i.e., *he*], resulting in a Condition C violation'. In other words, the binding facts suggest the head noun originates outside the RC.

in the derivation. This is more plausible if they are **adjuncts** (as in HEA) rather than **arguments** (as in HRA)

2. Extraction asymmetries: Extraction of the head noun of an RC is marginally acceptable, while extraction of another noun from within the RC is ungrammatical.[28] This suggests the head noun does not originate within the RC, *contra* the HRA
3. Various violations of locality constraints: including, e.g., violations of the Left Branching Constraint and extraction from a PP, etc.
4. Status of the relative pronoun: Relative pronouns don't pattern with determiners, but with demonstratives[29]
5. Case assignment: Under the HRA, the assignment of case by the governing D^{o} appears to take place at PF (once the derivation is complete) rather than in the narrow syntax

A comprehensive discussion of these issues lies beyond the scope of this thesis. Some of Salzmann's criticisms are theory-internal (e.g., late merger), and some do not clearly apply to ancient IE languages. For locality, to take just Latin as an example: the Left Branching Constraint is completely inoperative (Ledgeway, 2018). Again, extraction from a PP falls into the larger pattern of discontinuous constituents, which are perfectly grammatical. Moreover, it is very difficult to diagnose differences between Determiners and Pronouns in these languages: relative pronouns are formally identical whether they appear with an overt head noun or not.[30] As such, they may not be directly applicable for the purposes of the current investigation. Nevertheless, it is worth considering the alternative put forward to remedy these issues: most commonly, this is the Matching Analysis (MA).

The MA represents a mixture of the HEA and the HRA. The theory stipulates that there are two instances of the head noun: one generated externally (per the HEA), and one within the RC (per the HRA). The two are identical, but are not copies in the Minimalist sense: they are formally identical and not linked by a movement chain: they **match**.[31] the RC can be either adjoined to the externally generated head noun (per the HEA), or a complement (per the HRA). The former is demonstrated below:

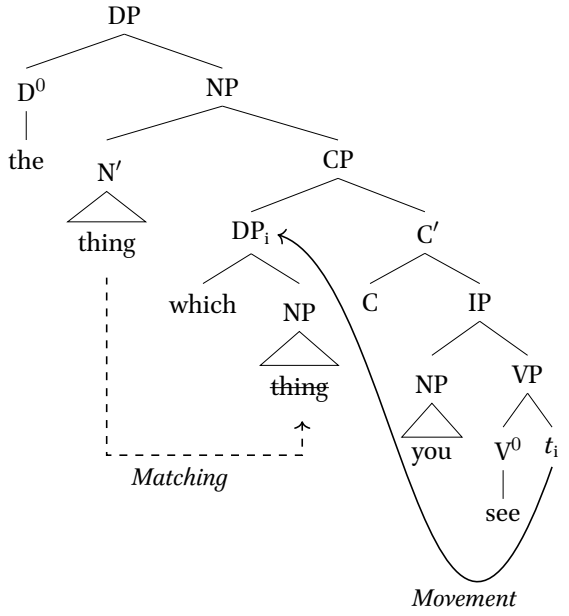
[28] See Salzmann (2017: 89–90) and Bianchi (1999: 54–61). Salzmann (2017: 87–8) also bolsters this claim by discussing evidence from clitic placement in Croatian, which, he argues, suggests the head noun is not part of the RC.

[29] Cf. English *The boy who* vs **Who boy* (Salzmann, 2017: 93–4).

[30] See further §3.3.1 on this point.

[31] See Citko (2001: 137) for an expansion on what constitutes 'identity', under which matches are deleted.

(2.41)



As in the HRA, the RC-internal head noun raises (together with OP, or the relative pronoun) to [Spec, CP]. When the derivation reaches PF, the lower, RC-internal head noun will be deleted under identity, leaving only the higher, RC-external head noun to be pronounced, but stranding the relative pronoun. Essentially then, as noted by Douglas (2016: 15), '[t]he fundamental difference between the HRA and MA lies in how the top-most copy of the RC head gets to its position. In the HRA, we have movement, whilst in the MA, we have base-generation.'

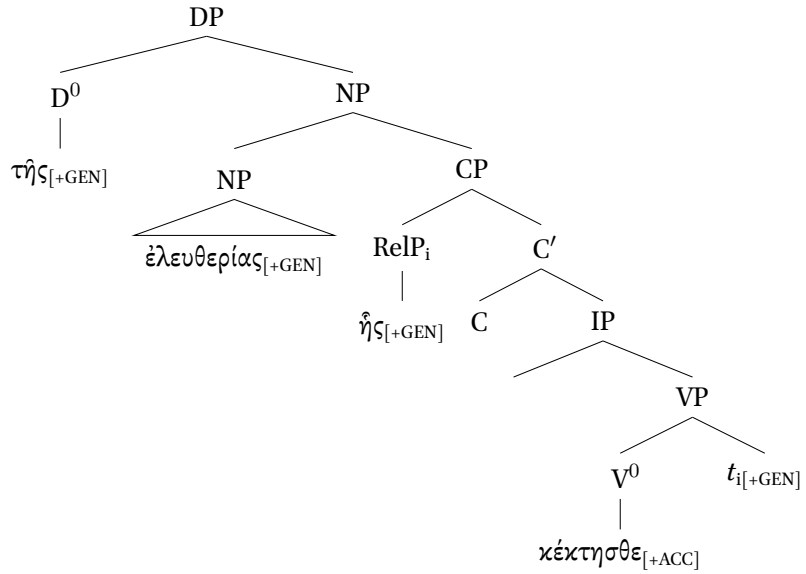
Advocates for the MA argue that it solves the issues such as those enumerated by Salzmann (2017) for the HRA. Nevertheless, the MA brings with it its own issues. Indeed, the criticisms of the sort made by Brame (1968), Schachter (1973) and Vergnaud (1974) of the HEA are arguably applicable to the MA: as long as we understand the head noun to be base-generated outside the RC—whether it is 'matched' with the internal head or not—we cannot account for reconstruction effects such as those observed in ex. (2.32).[32] Nevertheless, some authors (e.g., Douglas, 2016: 26–9) have argued that the MA can account for these reconstruction effects. Once again, we find ourselves caught in an ongoing dispute that is somewhat theory-internal, with diagnostics that are difficult to apply to the languages under investigation in this thesis. Once again, we may shed some light on the issue by looking at case effects.

Under the MA, we get the first of the three case patterns (the 'regular' one) for free. Since the top-most instance of the head noun is generated externally to the RC, we predict that its case will be assigned within the matrix clause (GEN). On the other hand, since the relative pronoun is base-generated within the RC, together with the lower instance of the head noun, we predict that it will reflect the case assigned according to its grammatical function within the clause. On the face of it, this is a simpler account than one in which the head-noun is somehow 're-assigned' its case according to the external D°. Yet to account for the 'attraction' of the relative pronoun to the case of the head noun, another mechanism is required. A detailed account of case attraction under the MA is given in Salzmann (2017: 411–7).

[32]This claim is maintained in Bhatt (2002, 2015), Sauerland (1998, 2003), Hulsey & Sauerland (2006), Aoun & Li (2003), Cinque (2013, 2020).

He notes that case attraction is most likely when the head noun bears a more oblique case than the relative pronoun, following the hierarchy GEN > DAT > ACC > NOM. In other words, when a relative pronoun undergoes case attraction, it is attracted from a less oblique case to a more oblique case. Citing the work of McFadden (2004) and Assmann (2013), Salzmann (2017: 410) argues that we should understand cases as feature-bundles: the more oblique a case, the more features it bears. Thus NOM can be schematised as [α], ACC as [α β], etc. If the probe (i.e., the case-assigner) has a subset of the case features of the goal (i.e., the case-assignee), the latter can **check** the case requirements of former without entering into an agreement relation. A second requirement for his model is that the relative pronoun can be merged into the derivation with **either** the case features assigned by the RC verb **or** the matrix verb. This allows Salzmann to posit structures such as the following to account for instances of case attraction (adapted for ex. 2.36):

(2.42)

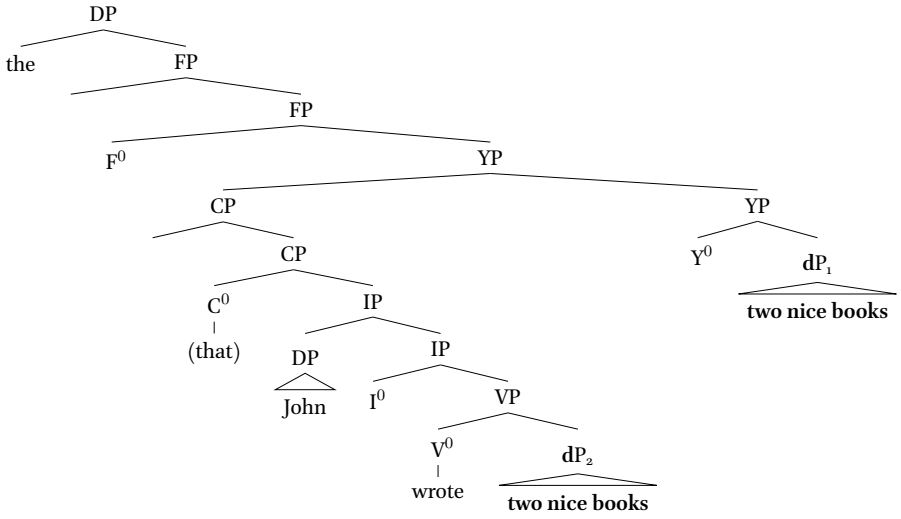


The derivation proceeds as follows. The relative pronoun, ἧς is merged into the derivation bearing the (unchecked) case features assigned to the head noun in the MC (GEN). The relative pronoun (RelP) then enters into a checking relation with κέκτησθε, allowing the verb to 'discharge' its accusative case-probe; yet because RelP bears the case feature [*i*GEN], which is more oblique than the one on the verb, [*u*ACC], according to Salzmann (2017: 412), it 'remains active for case checking.' It can then move to [Spec, CP] where it enters another checking relation. This time N° (ἐλευθερίας) acts as a probe. And since ἐλευθερίας bears the same case features as its goal (ἧς), the former checks the remaining features on the latter, at which point it ceases to be active for any further case checking.

So, the MA may be able to handle conventional attraction phenomena too. Yet Salzmann (2017: 417, n.56) admits that it cannot account for "inverse attraction": it falls into exactly the same trouble as the HEA. Because the instance of the head NP that is spelled out is base-generated external to the RC, there is no mechanism for it to receive its case from the relative pronoun. The field thus remains quite divided as to which of the HRA or the MA provides a more cogent account for the syntactic facts. Depending on the precise details of a given RC both intra- and cross-linguistically, each appears to have its own strengths. As

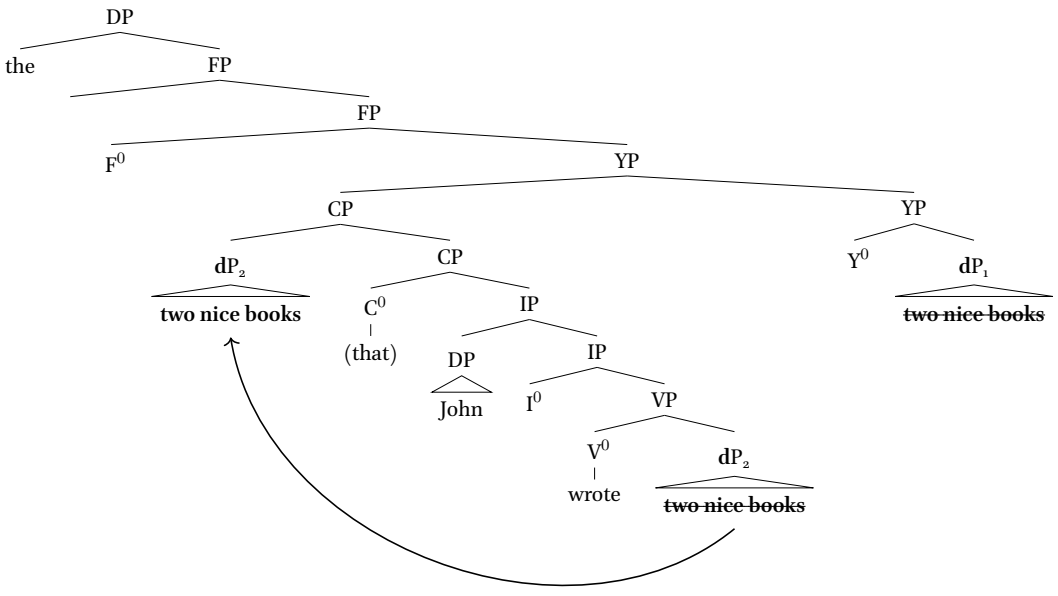
such, in a bid to square the circle, certain authors have permitted a 'mixed account', of the sort developed by Sauerland (1998, 2003). In short, we may hold that within the same grammar, RCs can instantiate either the HRA or the MA construction, with different grammatical effects accordingly. A recent account of this sort is Cinque (2013, 2020). He holds that all RCs begin with the following structure:[33]

(2.43)



In cases where the HRA makes the correct predictions, Cinque postulates that dP_2 raises to [Spec, CP]. At PF, both dP_1 and the lower copy of dP_2 are deleted under identity, yielding a structure that could have been derived by the HRA.

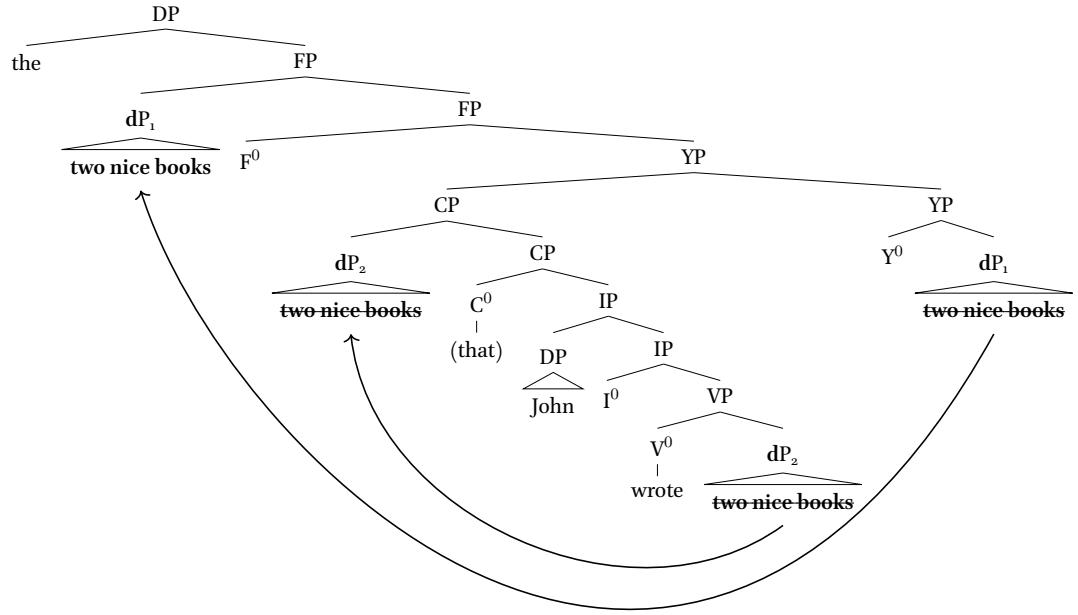
(2.44)



[33]In this notation, FP and YP are both generic functional projections within the NP whose specifiers act as 'landing' sites for NP-internal movement, on which see Cinque (1994) and Cinque (2010: 24–8). dP is used to mean 'the chunk of the extended projection of the NP modified, c-commanded, by the RC [which] is 'indefinite' (only contains weak determiners in Milsark's (1974) terms)' (Cinque, 2020: 14).

On the other hand, where the MA makes the correct predictions, it is dP_1, base-generated externally, that raises, to [Spec, FP]. In this scenario, dP_2 is still raised to [Spec, CP] within the RC, but is deleted at PF under identity.[34]

(2.45)



Thus, Cinque ingeniously collapses both matching and raising constructions into one underlying structure, the difference being encoded not in fundamentally different derivations, but rather as variation between which of the two instances of the head noun, dP_1 or dP_2, raises. To bring this back to the touchstone of case attraction—on which Cinque (2020) actually has fairly little to say—this model could allow for a HRA-style derivation to account for "inverse attraction", but an MA-style derivation to account for other case patterns accordingly. More generally, Cinque (2020: 31) posits the following features as characteristic of dP_1 raising (generating the HRA structure, as in ex. 2.44):

a. (obligatory) reconstruction (the Head has to be interpreted inside the RC)
b. sensitivity to strong islands
c. refractoriness to resumptive pronouns
d. possible occurrence in anti-pronominal contexts
e. incompatibility with *wh*-pronouns (unless these are assimilated to an invariant relativizer)
f. no extraposition
g. no stacking
h. sensitivity to weak islands

In all other instances, according to Cinque (2020), it is dP_2 that raises, generating the MA structure, as in ex. (2.45). It is worth noting that these distinctions in some way replicate the earlier debates

[34]Although this is not demonstrated with Cinque's examples, deletion under identity will only delete the phonetically identical components of dP_2 and not, therefore, the relative pronoun (OP), which surfaces in its position within [Spec, CP].

on HRA vs. MA, and as such are subject to the same difficulties when applied to ancient IE languages. Consider, for instance, the claim that HRA-style RCs show 'incompatibility with *wh*-pronouns'. This restriction is demonstrated by considering (maximalising) amount RCs, a type of RC in which Cinque (2020: 22) suggests a 'raising'/HRA-style derivation is 'forced'. He notes (2020: 28) that these are allegedly incompatible with *wh*-pronouns:

(2.46) Marv put {everything, those things, the things} ({that, *which}) he could in his pocket.
(Carlson, 1977: 526)

He further argues that speakers who disagree with this grammaticality judgement[35] 'reanalyse such pronouns as stylistic variants of *that*/∅'.[36] In English at least, any supposed counter-evidence to the claim that *wh*-pronouns are incompatible with amount RCs can thus be explained away. Yet Cinque (2020) does not appear to consider languages in which inflected relative (i.e., *wh*-) pronouns are the **only** way of introducing finite RCs, to the exclusion of an uninflected complementiser. Clearly such languages instantiate amount RCs, as in the following example from Ancient Greek (Probert, 2015: 140):

(2.47) εἰ ἐτεὸν δὴ πάντα τελευτήσεις ὅσ' ὑπέστης /
if truly PTC all.ACC.PL.N finish.FUT.2SG which-amount.ACC.PL.N promise.AOR.2SG
Δαρανίδῃ Πριάμῳ
Dardanian.DAT.SG.M Priam.DAT.SG.M
'If you will really accomplish everything (**which**) you promised to Dardanian Priam'
(Hom. *Il.* 13.375–6)

These data do not fit neatly into Cinque's characterisation of amount relatives. I have further concerns with Cinque's (2020) account for correlative clauses, which I discuss in §2.2.2.2 below. Nevertheless, it seems clear that whether or not the two approaches can be unified à la Cinque, both the HRA and the MA are required to account for 'plain' relatives. For this reason, I do not adopt a strong stance as to which is preferable here; both will feature in my analysis of PIE RCs in §5.

2.2.2.2 Correlative relatives

So far, I have focused mainly on the structure of the RC, and not the MC. This is because in 'plain' relatives of the type discussed in §2.2.2.1, the RC occupies a DP position within the MC, often as an argument of the main verb. In other words, if the relativised head noun is the subject of the sentence, the RC will occupy the subject position in the MC:

(2.48) [$_{\text{MC}}$ [$_{\text{Simple DP}}$ The girl] is tall.]

(2.49) [$_{\text{MC}}$ [$_{\text{Relativised DP}}$ The girl [$_{\text{RC}}$ who is standing over there]] is tall].

In correlative clauses, however, the RC does <u>not</u> occupy a DP position within the MC; that position is instead occupied by a correlative demonstrative pronoun, as in the following form:

[35] I include myself in this number.

[36] He attributes this idea to Andrew Simpson, reported in Aoun & Li (2003: 244, n.15).

(2.50) [$_{RC}$ Which girl is standing], [$_{MC}$ **she** is tall]

What arises from this distinction is the fact that while 'plain' RCs can be 'embedded' within the MC, correlatives cannot; the 'embedded' position is reserved for the correlative demonstrative:

(2.51) [$_{MC}$ I told the [$_{RC}$ student who was writing a thesis] to take a break]

(2.52) [$_{RC}$ Which student was writing a thesis], [$_{MC}$ I told **them** to take a break]

In English, while correlative clauses are (just about) grammatically possible, they are a highly marked relativisation strategy, and are usually avoided in favour of 'plain' relatives. However, there are some languages for which correlatives are quite unmarked, including certain ancient IE languages. Much of the generativist literature on the topic has been based on Hindi and other closely related Indo-Aryan languages, since they are some of the better studied modern languages that employ correlatives as an unmarked relativisation strategy; as such, many of the examples in the following section come from Hindi, such as the following (Dayal, 1996: 160):[37]

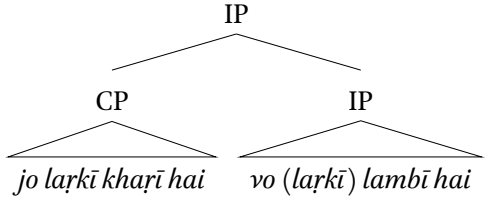
(2.53) *jo* *laṛkī* *khaṛī* *hai* *vo* (*laṛkī*)
REL.NOM.SG girl.NOM.SG.F standing.SG.F be.PRES.3SG DEM.NOM.SG girl.NOM.SG.F
lambī *hai*
tall.NOM.SG.F be.PRES.3SG
'Which girl is standing, she is tall' = 'The girl who is standing is tall'

Note also that in correlative clauses, the head noun can be spelled out twice: once in the RC, and once in the MC. This is another feature that is not possible in 'plain' RCs, where even if there are two instances of the head noun underlyingly (as in MA), only one is spelled out.

In §2.2.2.1 above, I concluded that 'plain' RCs could be accounted for with either the HRA or the MA, both of which ultimately lead to complex DPs that can occupy argument positions within the MC. Despite some obvious similarities, correlatives do not immediately follow from the same grammatical structures, and for this reason they require their own theory. Historically, there have been two main schools of thought surrounding the structure of correlative RCs. These can be broadly defined as **Uniform** versus **Non-Uniform** accounts of correlatives. The former treats the underlying structure of correlative RCs to be identical to 'plain' RCs, while that latter treats correlatives as structurally distinct.

I start with the Non-Uniform approach to correlative clauses. For Non-Uniformists, the correlative RC is not a DP, like a 'plain' RC, but rather a CP, which is adjoined to the IP of the MC (Srivastav, 1991; Dayal, 1996; de Vries, 2002; Davison, 2009). This is represented below for ex. (2.53):

(2.54)



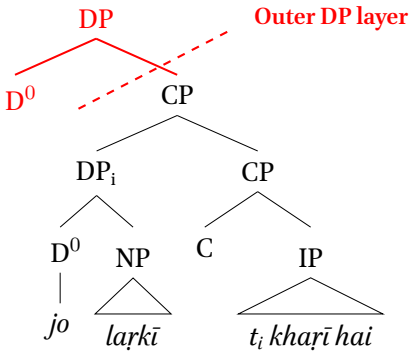
[37] When transliterating Hindi, I use *ṛ* to represent the retroflex tap /ɽ/, in the absence of a syllabic /ɾ̩/. In general, however, I use IAST to transliterate from Devanāgarī in exactly the same way for both Sanskrit and Hindi.

De Vries (2002: 40) lists four features of correlative RCs that suggest they are CPs rather than DPs:

(a) Correlative RCs do not occur in DP positions
(b) Correlative RCs never have an external determiner
(c) Correlative RCs never have an external Case ending or another nominal marking
(d) Correlative RCs never have an external (affixed) adposition

There is a conceptual simplicity to seeing the relative part of a correlative clause as a CP, especially when combined with the HRA adopted for 'plain' relatives, repeated below:

(2.55)



The difference between a correlative clause and a 'plain' relative boils down to one core structural difference: the **outer DP layer**, which is marked above the dotted line in **red** ex. (2.55). In a 'plain' relative, the external D° is responsible for nominalising the relative CP, allowing the relativised DP to occupy an argument position within the clause. In correlative clauses, this outer D° is not merged into the derivation, and the RC remains a CP. This means it cannot occupy an argument position within the matrix clause, and so the position must be occupied by a co-indexed correlative DP.

One of the major criticisms of the Non-Uniform approach to correlatives is that it requires a different structure to a 'plain' relatives. According to Cinque (2013, 2020), however, there are no languages that use correlatives **exclusively** for relativisation. In Hindi, for example, ex. (2.53) can be paraphrased using a centre-embedded, postnominal 'plain' RC:

(2.56) *vo* *laṛkī* [$_{RC}$ *jo* *khaṛī* *hai*] *lambī* *hai*
DEM girl.NOM.SG.F REL standing.SG.F be.PRES.3SG tall.NOM.SG.F be.PRES.3SG
'The girl [$_{RC}$ who is standing] is tall'

To Cinque, this suggests that correlatives are, by and large, simply a specialised form of 'plain' relative (to be discussed below).[38]

Another set of arguments against the correlatives-as-independent-CPs approach comes from Bhatt (2003), who argues that a Uniform hypothesis is preferable.[39] Building on the observations of Dayal

[38]It is worth noting, however, that 'embedded' RCs are quite marked in modern Hindi; they are also poorly attested up until the proliferation of texts translated from English to Hindi in the late 19th century, which may suggest their existence in contemporary varieties arose as a result of language contact with English (Puri, 2011).

[39]Previous authors arguing in favour of a Uniform approach to correlatives include Verma (1966), Junghare (1973), Kachru (1973), Wali (1982), Subbarao (1984), Bains (1989) and Mahajan (2000).

(1996), Bhatt (2003) demonstrates that correlatives in Hindi are bound by various locality constraints: they are sensitive to island effects; the RC and the correlative pronoun appear to form a constituent; and they are subject to reconstruction effects, as exemplified with Binding condition C[40] below (Bhatt, 2003: 513):[41]

(2.57) *jo* *laṛkī$_i$* *sītā-ko$_j$* *pyār* *kartī* *hai,* *us-ne$_{i/*j/k}$*
REL.NOM.SG girl.NOM.SG.F Sita.ACC.SG.F love doing.SG.F be.PRES.3SG, DEM.ERG.SG
us-ko$_{*i/j}$ *ṭhukrā* *diyā*
DEM.ACC.SG rejection give.PERF
'Which girl$_i$ loves Sita$_j$, she$_{i/*j/k}$ rejected her $_{*i/j}$'

Bhatt argues that a structure in which the RC constitutes a CP base-generated outside the matrix clause cannot account for such locality effects. He therefore argues that the relative CP is adjoined to the correlative pronoun (DemP, within the matrix clause), and subsequently fronted:

(2.58) [$_{RC}$ *jo* *laṛkī$_i$* *sītā-ko$_j$* *pyār* *kartī* *hai*]$_m$, *us-ne$_{i/*j/k}$*
[$_{RC}$ REL.NOM.SG girl.NOM.SG.F Sita.ACC.SG.F love doing.SG.F be.PRES.3SG]$_k$, DEM.ERG.SG
[$_{Dem\text{-}XP}$ t_m *us-ko$_{*i/j}$*] *ṭhukrā* *diyā*
[$_{Dem\text{-}XP}$ t_m DEM.ACC.SG] rejection give.PERF
'Which girl$_i$ loves Sita$_j$, she$_{i/*j/k}$ rejected her $_{*i/j}$'

Bhatt argues that this account for all 'simple' (i.e., single-headed) correlatives in Hindi. However, Hindi also exhibits **multiple-headed correlatives**, such as the following:

(2.59) *jis* *laṛkīne* *jis* *laṛke* *ke-sāth* *khelā* *usne* *us-ko*
REL.ERG.SG girl.ERG.SG.F REL.ERG.SG boy.ERG.SG.M with play.PERF she.ERG he.ACC
harāyā
defeat.PERF
'Which girl played with which boy, she defeated him'='Whichever girl defeated whichever boy she played with'

Multiple-headed correlatives do not exhibit the same locality effects observed in single-headed correlatives (Bhatt, 2003: 516).

(2.60) *jis-ne$_i$* *rām-ko$_j$* *jise$_k$* *diyā,* *us-ne$_j$* *us-se$_k$* *us$_i$* *kī* *tarīf*
REL.ERG.SG Ram.ACC REL.DAT.SG give.PERF DEM.ERG.SG DEM.INS.SG DEM.OBL of praise
kī
do.PERF
'Who$_i$ gave Ram$_j$ to whom$_k$, he$_k$ praised him$_i$ to him$_j$' = 'Person A$_i$ gave Ram$_j$ to Person B$_k$, and he$_j$ (= Ram) praised him$_i$ (= Person A) to him$_k$ (= Person B)'

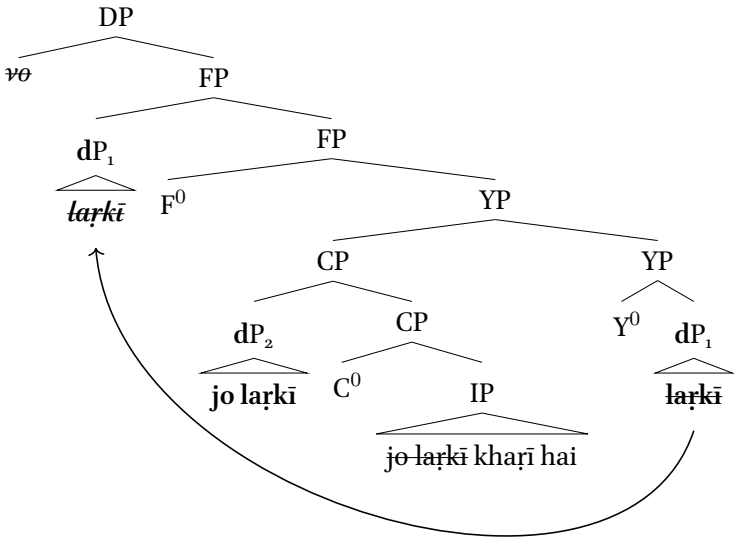
[40] That an R-expression (here, *Sītā-ko* 'Sita') cannot have an antecedent (here, *us-ne* 'she') that proceeds and c-commands it. To account for this effect in ex. (2.47), we must understand *Sītā-ko* as base-generated below *us-ne*.

[41] Bhatt (2003: 512, n.16) notes that there is 'considerable speaker variation here'.

Note that in ex. (2.60), for the R-expression *rām-ko* ('Ram') to be felicitous under Binding Condition C, it cannot originate below the co-referenced demonstrative *us-ne* ('he'). As such, Bhatt argues that for multiple-headed relatives only, the relative CP must be base-generated at the left edge of the clause, adjoined to IP, as in the non-Uniform analysis of correlatives. For the majority of correlatives, being only single-headed, the Uniform account of correlatives is preferred because it can account for the various locality constraints put forward. Put another way, Bhatt's criticism of traditional non-Uniform approaches to correlatives can be characterised as 'generaliz[ing] the worst case' (Cinque, 2013: 210), namely that of multiple-headed correlatives.

Cinque's (2013, 2020) approach is quite different from Bhatt's, but is still largely Uniform in nature, in the sense that he derives correlatives from the same starting point as 'plain' relatives. Essentially, he holds that the relative part of a correlative sentences is straightforwardly a left-dislocated relative **DP** (not CP), which is resumed by the correlative pronoun in the matrix clause. To exemplify this, let us take ex. (2.53) again. The starting point would be the 'plain' RC below (with the MA-style derivation as the default):

(2.61)



To derive the attested form, Cinque posits that the external D° *vo* and head noun (dP$_1$) undergo a form of 'recoverable deletion' (Cinque, 2013: 209). This leaves us with the RC *jo laṛkī khaṛī hai*, which Cinque understands to be a DP with the underlying structure posited in ex. (2.61). This DP is base-generated in the MC and then raised to somewhere at the left edge of the clause. The syntactic detail concerning the nature of the fronting operation seem to vary cross-linguistically: Cinque (2020: 129) notes that the resulting construction 'may apparently be either an English-type Left dislocation/Hanging Topic (Kashmiri), or a German-type Contrastive Left Dislocation (German, Bulgarian), or a Romance-type Clitic Left Dislocation (as in the 'correlatives' of Italian).' In any event, the fronted relative DP is resumed in the MC, variably with repetition of the head noun or with a demonstrative alone, yielding the attested form:

(2.62) [$_{\text{Topic}}$ [$_{\text{DP}_i}$ *jo* *laṛkī* *khaṛī* *hai*]], *vo*$_i$
[$_{\text{Topic}}$ [$_{\text{DP}_i}$ REL.NOM.SG girl.NOM.SG.F standing.NOM.SG.F be.PRES.3SG]] DEM.NOM.SG
lambī *hai*
tall.NOM.SG.F be.PRES.3SG
'[$_{\text{Topic}}$ The girl who is standing], she is tall'

Cinque (2020) justifies this approach with reference both to the evidence from locality discussed by Bhatt (2003), and by the logic that since correlatives are a 'non-exclusive relativization strategy', they do not merit *a priori* a distinct structure form 'plain' relatives. Cinque (2020: 139–41) also criticises de Vries' (2002: 40) observations, listed above, that suggest correlative RCs are CPs rather than DPs. He offers a selection of phenomena from various languages in response to each of de Vries' points, arguing that a DP-analysis to be preferable. These include:

a. Left dislocation accounts for the fact correlative RCs do not occur in DP positions
b. Correlatives in Itzaj Maya (Hofling, 2000) can be governed by an external D° *á* > Correlatives **can** have an external determiner
c. Correlatives in Austronesian (Formosan) language Isbukun Bunun can be followed by a topic maker *hai* > Correlatives **can** have a 'nominal marking'
d. Correlative RCs can be preceded by a preposition in Italian > Correlatives **can** have an external adposition

On this basis, Cinque concludes that 'the DP status of correlative relatives is confirmed' (2020: 141). Nevertheless, he must still concede that the only viable treatment of multiple-headed correlatives is one in which they are 'free adjunct free RCs', i.e., CPs (Cinque, 2020: 127). He notes that 'a DP analysis for such cases is out of the question since the correlative CP cannot have two external heads' (2020: 127). Conceptually, then, it must be conceded that bare correlative relative CPs exist; they are simply a different construction from single-headed relatives. In sum, arguments from locality in Hindi and sundry grammatical phenomena noted by Cinque speak against a bare-CP analysis of some correlative RCs; in the absence of certain locality constraints, as in multiple-headed correlatives, a bare-CP approach is warranted. I believe this presents a satisfactory account of the facts in (at least) Hindi and other closely related Indo-Aryan languages.

As with the distinction between MA vs. HRA for 'plain' relatives, however, these arguments present some difficulty when applied to the ancient IE languages. We cannot diagnose island constraints, nor reconstruction effects of the sort observed in Hindi. Though they are rare, we do see multiple-headed correlatives in, e.g., Sanskrit:[42]

(2.63) *yasya-yasya*$_i$ *matam* *yad-yad*$_j$ *śrotum* *icchāmi* *tasya*$_i$
REL.GEN.SG.M opinion.NOM.SG.N REL.NOM.SG.N hear.INF want.PRES.1SG DEM.GEN.SG.M
tat$_j$
DEM.ACC.SG.N
'Whatever the opinion$_i$ is of whoever$_j$, I desire to hear that$_i$ from him$_j$' (MBh 1.188.6c)

[42] See further Hock (1989: 96).

Thus, even under Cinque's theory, we must concede that bare-CP correlative RCs are possible. The question is whether we have reason to believe that other, single-headed correlatives, are more convincingly derived under a Uniform approach than under a non-Uniform approach. My belief is that they are not. For example, the ancient IE languages present no serious challenge to de Vries' (2002: 40) criteria (see §5.2); while I am happy to accept Cinque's exceptions as genuine, I do not believe they 'confirm' the status of **all** correlative RCs as DPs. I believe instead that this is a feature that can vary parametrically; I will argue in §5 that for PIE, it is more likely that correlative RCs were CPs than DPs.

The question remains, therefore: at what level does the relative CP adjoin to the MC? In ex. (2.54) above, I have represented it as adjoining to IP. In Hindi, there is good evidence for this from minimality violations and other locality constraints (Dayal, 1996; Bhatt, 2003; Davison, 2009). Yet again, such diagnostics cannot be applied in the same way to ancient IE languages. In Sanskrit, for example, there is no clear evidence for such minimality constraints of the sort observed in Hindi (Davison, 2009); moreover, as established by Hock (1989), there is solid evidence that Sanskrit correlative structures are formed of two adjoined CPs; similarly strong evidence can be posited for Hittite, and extended to Latin and Ancient Greek. The question of adjunction position is thus another feature of RCs that may vary parametrically: I address the question of the adjunction site in PIE in §5.2.

I thus conclude this section with a summary of the generative account for RCs I adopt in this thesis.

- 'Plain' relatives are DPs, which take the RC as either a complement or an adjunct. The head NP raises from within the RC to [Spec, CP]: it may be matched by an external instance of the head noun, and subsequently deleted (MA), or there may be no external head noun, in which case the head noun that has raised from within the RC is spelled out (HRA)
- Correlatives have a similar structure to 'plain' relatives, but **can** lack the **outer** DP layer (i.e., they can be DPs or CPs)

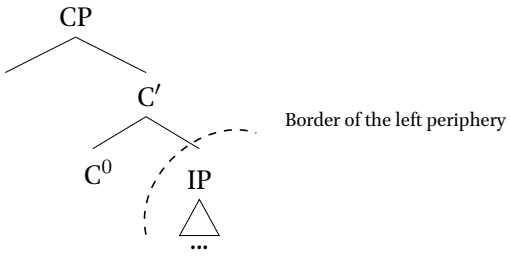
Much is left unsaid in this summary: I aim to balance the elements of the theory I take to be 'given' for the purposes of this thesis (e.g., the theoretical validity of the HRA and the MA), and those which require investigation precisely as part of my reconstruction for PIE (e.g., the categorical status of correlative RCs), which I will return to in §5.

The last set of theoretical preliminaries that must be addressed pertain to the specifics of the [Spec, CP] position.

2.3 The left periphery

The **left periphery** constitutes what I have so far labelled as C° and anything that dominates it.

(2.64)



While this remains a useful shorthand, it has been established since Rizzi (1997) that this part of the clause is more complex than a single projection. In the first place, the CP domain is responsible for two core grammatical functions: Force and Finiteness. Force gives the sentence its 'type': declarative, interrogative, exclamative etc. In this sense Rizzi (1997) describes it as "looking outwards"; it contextualises the utterance within the discourse. Finiteness, on the other hand, "looks inwards"; it dictates whether the verb will be finite (i.e. bear features of mood and personal agreement, e.g. main verbs) or non-finite (i.e., not bearing the features above, e.g. participles, gerunds). For instance, the English complementiser *that* marks the sentence as both [+declarative] (Force) and [+finite] (Finiteness). The distinct functions of Force and Finiteness are therefore attributed to two syntactic heads: Force° and Fin(iteness)°. As noted by Roberts (2004: 300), that a complementiser is associated with the properties of two functional heads—Force° and Fin°—is comparable to the fact 'a finite verb is associated with properties of V° (thematic structure) and T° (tense).'

Rizzi (1997) established that we can diagnose the different positions of Force° and Fin° by observing the distribution of complementisers with respected to fronted topics. Consider the following (Rizzi, 1997: 288):

(2.65) *Credo* ***che*** *il tuo libro, loro lo apprezzerebbero molto*
believe.PRES.1SG that the your book they it appreciate.CND.3PL much
'I believe that your book, they would appreciate it a lot'

(2.66) **Credo, il tuo libro,* ***che*** *loro lo apprezzerebbero molto*
believe.PRES.1SG the your book that they it appreciate.CND.3PL much
*'I believe, your book, that they would appreciate it a lot'

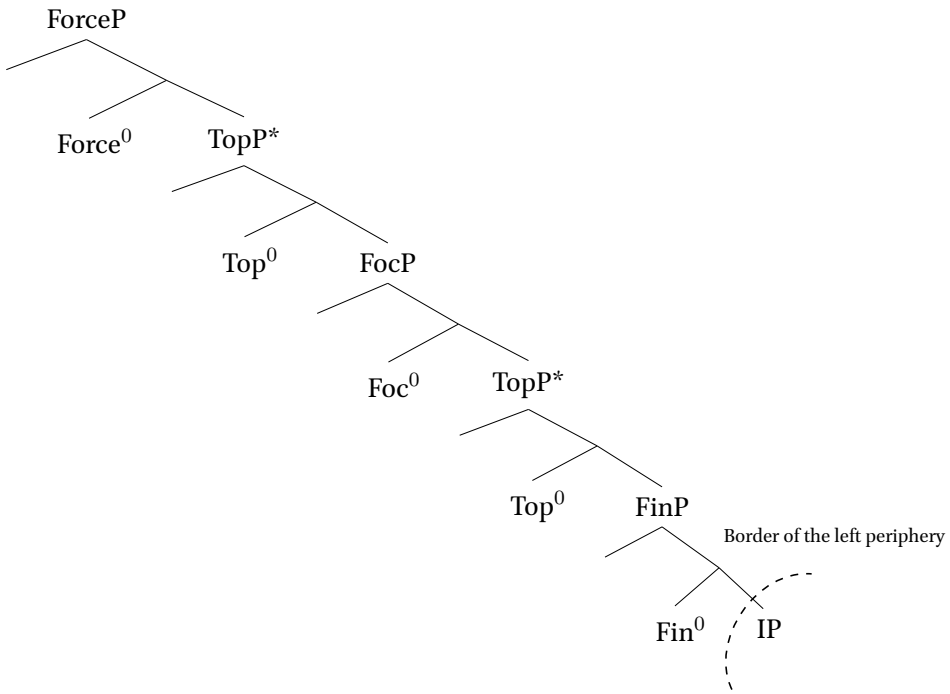
(2.67) **Credo* ***di*** *il tuo libro, apprezzar-lo molto*
believe.PRES.1SG of the your book appreciate.INF-it much
*'I believe I, your book, appreciate it a lot'

(2.68) *Credo, il tuo libro,* ***di*** *apprezzar-lo molto*
believe.PRES.1SG the your book of appreciate.INF-it much
'I believe, your book, I appreciate it a lot'

In these examples, we can compare the relative ordering of the topicalised DP *il tuo libro* with two different complementisers: *che* and *di*. Note that the topicalised DP must follow *che* but must precede *di*. On this basis Rizzi argues that *che* manifests in Force°, the "top" of the left periphery, while *di* manifests in Fin° (the "bottom"). Both would occupy C° in pre-Rizzian analyses: here we can see that this is not a unitary position.

Central to Rizzi's analysis of the left periphery is the Topic-Focus complex. This is an "accessory" component of the left periphery, visible only when a sentence contains a topicalised or focalised element, or both. It is located between Force° and Fin°, and constitutes maximally a single Focus head, Foc°, nested between an unlimited number of Topic heads (Top°). The full structure is given below (adapted from Rizzi, 1997: 297):

(2.69)



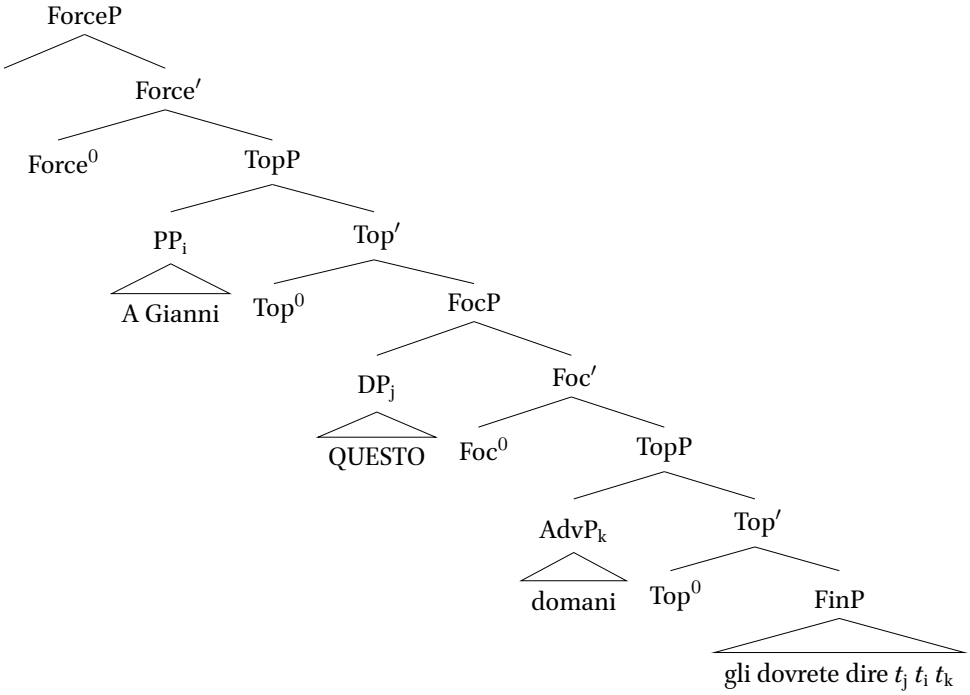
Top° and Foc° may bear strong features (see §1.3) and so attract topicalised/focalised elements to their specifier position. Thus Topics move to [Spec, TopP] and Foci move to [Spec, FocP], as in the following (Rizzi, 1997: 291):[43]

(2.70)	*A*	*Gianni,*	*QUESTO,*	*domani,*	*gli*	*dovrete*	*dire*
	to	Johnny	this	tomorrow	to-him	must.FUT.2PL	say.INF

'To Johnny, THIS, tomorrow, you should say'

[43] Focalised elements are written in CAPS.

(2.71)



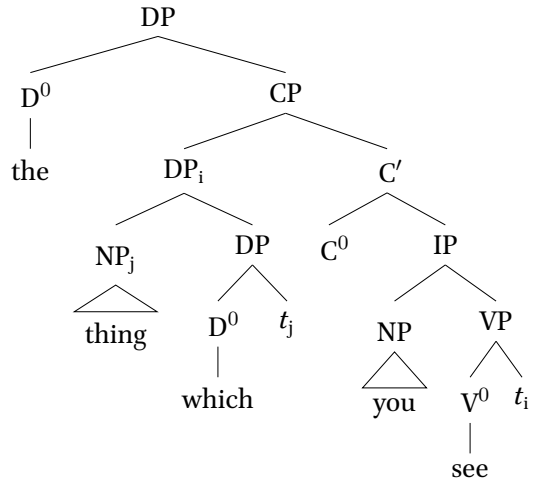
This theory relies on a clear understanding of the interpretive difference between **Topic** and **Focus**. In simple terms, a Topic is what the sentence is 'about', and 'is generally associated with the aspect of 'given' information' (Frascarelli, 2000: 2). On the other hand, Focus is often conceived of as information that is 'new'. Both Topics and Foci can be 'contrastive', yielding the following possibilities (Neeleman & Vermeulen, 2012: 5):

	Topic	Focus
No contrast	aboutness topic	new information focus
Contrast	contrastive topic	contrastive focus

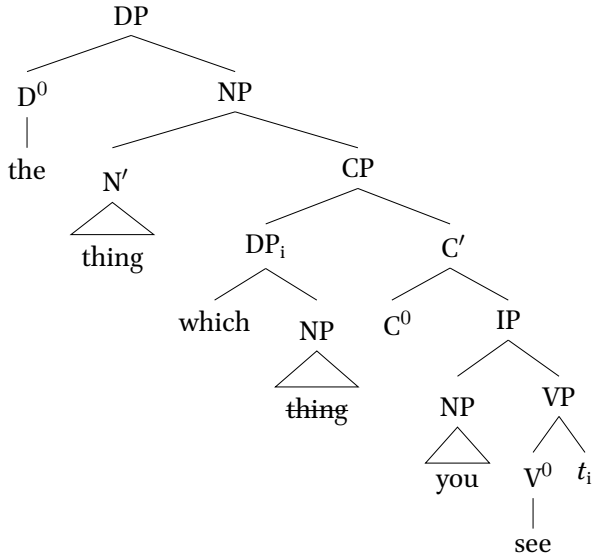
Rizzi (1997) notes that Focus is also **quantificational**, in a way that Topic is not: while quantified expressions (*everything, nothing* etc.) can be focalised, they cannot be topicalised. Additionally, Focus, unlike Topic, is **unique**: there can only be one Focus per sentence, while Topics can proliferate. This is encoded in Rizzi's model by TopP projections being recursive (marked with an asterisk* in ex. 2.69 above) while there is only one FocP.

With this structure established, we must revisit our analysis of RCs. Whether we adopt the HRA or the MA, the RC contains a CP complete with its own left periphery.

(2.72) **Head-Raising Analysis:**

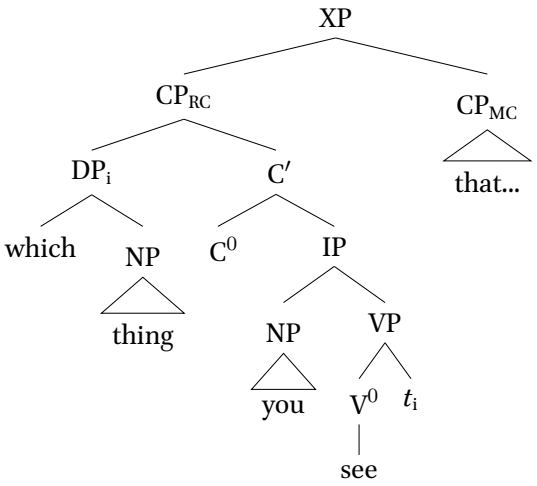


(2.73) **Matching Analysis:**



The same is also true, of course, of correlative clauses:

(2.74) **Correlative clause:**



In every case, the relative pronoun is fronted to left periphery of the RC, represented here by the generic [Spec, CP] position. This movement is traditionally referred to as "*wh*-movement", and treated

analogously to the fronting of interrogative pronouns.[44] However, with a finer understanding of the left-periphery, we observe the fact that interrogative and relative pronouns do not target the same position in the left periphery. According to Rizzi (1997), interrogatives target [Spec, FocP]:[45] this is justified on both interpretive grounds (interrogatives are quantificational, and often focalised) and distributional grounds, since interrogatives are not compatible with other Foci. Relative pronouns, on the other hand, can co-occur with Foci, as noted by Bianchi (1999: 191, n. 63):

(2.75) John is the kind of person **who** [$_{\text{Focus}}$ under no circumstances] would I be willing to talk to.

Since FocP is not recursive, [Spec, FocP] cannot be the landing site for fronted relative pronouns. We can also eliminate [Spec, FinP], since relative pronouns can be followed by Topics:

(2.76) This is the man who, [$_{\text{Topic}}$ his car], he gave away.

We are thus left with two viable positions for the relative pronoun: [Spec, ForceP] and [Spec, TopP]. Rizzi (1997) argues that relative pronouns occupy [Spec, ForceP] since they cannot be preceded by Topics in Italian. This does not seem to hold cross-linguistically, however: in ancient IE languages, topicalisation around the relative pronoun is quite common. I discuss this at length in §4, but for now note that this was observed for Latin (and Hungarian) by Bianchi (1999: 192), who cites examples such as the following:

(2.77) *spectatores* *optumos,* / [$_{\text{Topic}}$ *Fidem*] ***qui*** *facitis*
spectator.ACC.PL.M best.ACC.PL.M Faith.ACC.SG.F REL.NOM.PL.M do.PRES.2PL
maxumi
greatest.NOM.PL.M
'Most excellent audience, **who** esteem Faith most highly' (Plaut. *Cas.* 1–2)

On the basis of this evidence, Bianchi hypothesises that the relative pronoun occupies its own [Spec, TopP]. This may be justified on interpretive grounds: the relative phrase is, in a sense, what the RC is 'about'. Yet there seems to be some cross-linguistic variation here, as this does not account for the impossibility of topicalisation around the relative pronoun in languages such as Italian and English. There also remains the question of the ordering of the relative pronoun and the FocP projection. As such, the landing site for the relative pronoun is something that requires language-specific investigation: I address this question for PIE in §4 and §5.

Finally, there is question of the position of the head noun. It is first worth noting that in correlative clauses in ancient IE languages, unlike in English and Italian, the head noun does not itself need to raise to the left periphery, but can stay in its base-generated position, as in the following:

[44] Languages may lack *wh*-movement, in which case they are referred to as "*wh-in-situ*". As we shall see in §3 and §5, there is some evidence for *wh-in-situ* strategies for RCs in PIE.

[45] I do not address the position [Spec, IntP] here which Rizzi (2001) invokes to explain the position of indirect interrogative words (e.g. English 'if') or 'higher *wh*-elements' (e.g. Italian *perché* 'why?'), which goes beyond the scope of the current discussion.

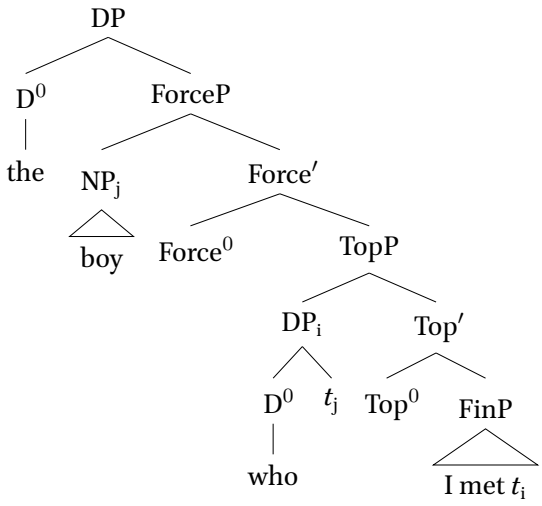
(2.78) ***yáḥ*** *tvā́m* *agne* ***havíṣpatiḥ*** *dūtám*
REL.NOM.SG you.ACC.SG agni.VOC.SG offering-master.NOM.SG messenger.ACC.SG
deva *saparyáti*
god.VOC.SG worship.PRES.3SG
'The **master of the offerings who** worships you the messenger, O divine Agni...' (RV 1.12.8ab)

When the entire relativised DP is raised, the head noun does not usually raise above the relative pronoun, but stays as its complement.[46]

(2.79) ὅν τινα μὲν βασιλῆα καὶ ἔξοχον ἄνδρα
REL.ACC.SG.M INDEF.ACC.SG.M PTC king.ACC.SG.M and eminent.ACC.SG.M man.ACC.SG.M
κιχείη
encounter.AOR.OPT.3SG
'**Whichever king** or eminent man he encountered...' (Hom. *Il.* 2.188)

As for 'plain' relatives, if we follow the MA, the version of the head noun that is spelled out is base-generated externally to the RC, and so is not implicated in the left-periphery of the RC at all. On the other hand, under the HRA, the head noun is strictly base-generated RC-internally and fronted. For this reason, Bianchi (1999) dedicates significant discussion to the position of the head noun, ultimately deciding that it raises to [Spec, ForceP], rather than [Spec, DP] as suggested in ex. (2.72) above. The revised structure is therefore the following (Bianchi, 1999: 191):

(2.80)



Since this is only really an issue for the HRA, and since raising of the head noun is not a necessary feature of ancient IE correlatives, I will take no strong view on the position of the head noun in the left periphery. Instead, I will focus on establishing the position of the relative pronoun in the left periphery (i.e., the nature of relative *wh*-movement), which affects all types of RCs under all possible analyses, representing a central question in the reconstruction of PIE RCs.

[46] It is possible for the head noun to raise above the relative pronoun, but this is likely a separate movement to a different [Spec, TopP] and not a feature of its being relativised *per se*.

Chapter 3

The Relative Pronoun(s) of Proto-Indo-European

Relative clauses are generally introduced in the archaic Indo-European languages by a relative pronoun. In some languages, this pronoun is descended from a form *k^{w}í-/*k^{w}ó-, while in others it is descended from a form *yó-. In this chapter, I survey the syntactic and semantic behaviour of the descendants of these pronouns in the attested languages. This includes a discussion of both their relative and non-relative uses. I conclude that neither *k^{w}í-/*k^{w}ó- nor *yó- can be excluded as a relative pronoun in Proto-Indo-European, and that together they reflect what was a unitary syntactic category in the proto-language: *REL.

3.1 A brief survey

The distribution of the relative pronouns across the IE language families is as follows:

- Descendant of ***k^{w}í-/*k^{w}ó-** as a relative pronoun:
  - Anatolian
  - Italic
  - Tocharian
- Descendant of ***yó-*** as a relative pronoun:
  - Armenian
  - Balto-Slavic
  - Celtic
  - Greek
  - Indo-Iranian
  - Phrygian
- Other
  - Albanian: uncertain. Relative *që* probably borrowed from Latin *qui* (Orel, 1998: 360)
  - Germanic: uses *þa- < PIE *só-/*tó-, variously as an inflected pronoun (cf. Got. relative pronouns in *-ei*) or an uninflected subordinator (cf. OE *þe*) optionally combined with an inflected demonstrative

As Clackson (2007: 173) notes, '[t]his distribution cuts across other isoglosses separating the IE languages and does not seem to reflect a 'dialectal' difference of the parent language.' Moreover, there is the added complication that the **yó-* languages consistently inherit some form of **kʷí-/*kʷó-*, but in one or more of its non-relative uses (see §3.3.3 below).[47] On the other hand, those languages which use **kʷí-/*kʷó-* as a relative pronoun often lack forms that are clearly traceable back to a PIE pronoun **yó-*. As such, neither of these pronouns can be excluded *prima facie* as a relative pronoun in PIE. I therefore address the behaviours of each of them in turn.

3.2 **yó-*

Formally, **yó-* can be considered to fit into a broader paradigm of anaphoric pronouns in **i-/*ey-*[48] (e.g. Lat. *is, ea, id* ('he', 'she', 'it'), Skt. *ayám, iyám, idám* ('id.') etc.). Such an idea was posited in the first place by Windisch (1869), and has since gained currency. Essentially, we can understand **yó-* as the zero-grade of a stem **(e)y-* + thematic vowel **o*. If we also allow for an *e*-grade of the athematic stem, we can generate such forms as the following:

- Zero-grade, athematic: **i-* > Lat. *is*, Lit. *jis* (anaphoric)
- *e*-grade, athematic: **ey-* > Lat. *eum*, Skt. *ayám* (anaphoric)
- Zero-grade, thematic: **yó-* > AGk. ὅς, Skt. *yáḥ*, Cb. *ios* (relative)

Given the formal plausibility and the semantic closeness of relative and anaphoric pronouns, I am willing to accept this hypothesis. However, as I will argue below, I do not think we are bound to reconstruct the thematised form **yó-* itself as originally anaphoric. I therefore do not accept *a priori* that the relative use of the form **yó-* is a post-PIE innovation. I make this case in §3.2.1, and consider alternative functions of **yó-* in §3.2.2.

3.2.1 Relative uses of **yó-*

Descendants of **yó-* are attested directly as relative pronouns in six branches of IE: Indo-Iranian, Greek, Armenian, Phrygian, Celtic and Balto-Slavic. These are exemplified below with Sanskrit, Ancient Greek, Classical Armenian,[49] Phrygian,[50] Celtiberian[51] and OCS[52] respectively:

(3.1) ***yáḥ*** *árvantam* *prathamáḥ* *adhi* *átiṣṭhat*
REL.NOM.SG.M racehorse.ACC.SG.M first.NOM.SG.M upon stand.IMPF.3SG
'**He who** was the first to mount the racehorse' (RV 1.163.9d)

[47] There is some doubt as to whether the Armenian interrogative/indefinite pronoun *i-* is from PIE **kʷi-*, since the total disappearance of **kʷ-* here is unexpected (Martirosyan, 2009: 299).

[48] This pronoun is sometimes reconstructed with an initial laryngeal, as **h₁(e)y-* (cf. Beekes, 2011: 226). Similarly, **yó-* is often reconstructed as **Hyó-*. I take no strong stance on the phonology here, and omit the laryngeals for simplicity.

[49] Example from Meyer (2013: 37).

[50] Example from Lubotsky (1988: 8–11).

[51] Example from Ziegler (1993: 252). I will discuss in detail the evidence for relative **yó-* in the other Celtic languages in §4.3; for now we can take Celtiberian as representative of Proto-Celtic.

[52] Example from the *Codex Zographensis* (Jagić, 1879: 3).

(3.2) ὃς μέγα πάντων / Ἀργείων κρατέει καί οἱ
REL.NOM.SG.M mightily all.GEN.PL.M Argive.GEN.PL.M rule.PRES.3SG and DEM.DAT.SG.M
πείθονται Ἀχαιοί
obey.PRES.3PL Achaean.NOM.PL.M
'**Who** rules mightily over all the Argives and whom the Achaians obey' (Hom. *Il.* 1.78–9)

(3.3) *yaynžam gay Yisows ənd nosa i geł mi*
then go.PRES.3SG Jesus with DEM.ACC.PL into village.ACC.SG INDEF.ACC.SG
***orowm** anown ēr get'samani*
REL.DAT.SG name IMPF.3SG Gethsemane
'Then Jesus went with them to a village **whose** name was Gethsemane' (Matthew 26:36)

(3.4) ***ios** esait : materey : eveteksetęy : ovevin : onoman : daψet*
REL.NOM.SG DEM.DAT.SG : mother.DAT.SG : ?.DAT.SG : ?.ACC.SG : name.ACC.SG : put.SUBJ.3SG
'Who(ever) puts [?] name on this [?] mother...' (W-01b)

(3.5) ***iomui** liśtaś TiTaś sisonti ...*
REL.DAT.SG ? ? SOW.PRES.3PL ...
'To whom(ever) they *liśtaś TiTaś* sow...' (Z.09.01)

(3.6) *tako bo izgŭnašę proroky **iže** běšę pręžde*
so for persecute.AOR.3PL prophet.ACC.PL.M REL.NOM.PL.M be.IMPF.3PL before
vas
you.GEN.PL
'For so they persecuted the prophets **who** were before you' (Matthew 5:12)

In each of these languages, **yó-* functions in a virtually identical way. It agrees in number and gender with its antecedent, introduces restrictive, maximalising and appositive RCs, and can occur either in correlative or 'plain' RCs (see §2.2 for terminology). Moreover, whether in a 'plain' relative or a correlative, descendants of **yó-* have a strong tendency to be placed clause-initially. Following the analysis of RCs in §2.2, we understand this fronting to be movement to the left-periphery (i.e., *wh*-movement). There are, however, some instances in which a descendant of **yó-* occurs in non-initial position. As noted by Hale (1987, 1996), Hock (1982, 1989), Kiparsky (1995), Lowe (2014) and Probert (2015) *inter alios*, descendants of **yó-* can be preceded by another topicalised or focalised constituent. These constructions are very well attested in Vedic, and appear also in Ancient Greek:

(3.7) *prá **yáḥ** vām mitrā varuṇā jīráḥ*
forth.PW REL.NOM.SG.M you.ACC.DU Mitra.VOC Varuna.VOC quick.NOM.SG.M
dūtáḥ ádravat
messenger.NOM.SG.M run.IMPF.3SG
'The quick messenger **who** ran to you, O Mitra and Varuna' (RV 8.101.3)

(3.8) κακῶς δ' οἵ πέρ μιν ἔρεζον / μνηστῆρες,
badly PTC REL.NOM.PL.M PTC him.ACC.SG.M do.IMPF.3PL suitor.NOM.PL.M
τοὺς πάντας ἐτίσατο ᾧ ἐνὶ οἴκῳ
DEM.ACC.PL.M all.ACC.PL.M avenge.AOR.3SG his.DAT.SG.M in houseDAT.SG.M
'The suitors who did him evil, he took vengeance upon them all in his house'
(Hom. *Od.* 23.56–7)

Additionally, in the Rig Veda, Skt. *yá-* < **yó-* can exceptionally be found with up to **two** fronted constituents preceding it:

(3.9) *hávīmabhiḥ* *hávate* ***yáḥ*** *havírbhiḥ*
invocation.INS.PL call.PRES.3SG REL.NOM.SG offering.INS.PL
'He **who** calls with invocations and offerings' (RV 2.33.5a)

The structures underlying these word orders will be treated in detail in §4.

All the examples so far have included a finite verb within the RC. There are a set of RCs, however, in which there is no overt form of the verb: these are traditionally referred to as 'nominal' relatives. Examples are attested directly in Indo-Iranian (here exemplified with Vedic) and Ancient Greek.

(3.10) *yuktágrāvnaḥ* ***yáḥ*** *avitā́* *suśipráḥ*
stone-mover.GEN.SG.M REL.NOM.SG.M helper.NOM.SG.M beautiful-lipped.NOM.SG.M
sutásomasya
Soma-presser.GEN.SG.M
'**Who** [is] the beautiful-lipped helper of the one who set the stones in motion, (and) of the one who pressed the Soma' (RV 2.12.6cd)

(3.11) Αἴαντές τε δύω Τεῦκρός θ', ὃς ἄριστος Ἀχαιῶν
Ajax.NOM.PL.M and two Teucer and REL.NOM.SG.M best.NOM.SG.M Achaean.GEN.PL.M
/ τοξοσύνῃ
archery.DAT.SG.F
'The two Ajaxes and Teucer, **who** [is] the best of the Achaeans in archery' (Hom. *Il.* 13.313–4)

'Nominal' relatives are usually quite short, often consisting of only one constituent other than the relative pronoun itself. Those constructed with **yó-* are syntactically parallel to those with **k^{w}í-*/**k^{w}ó-* as attested in Latin and Hittite (see §3.3.2). An apparent peculiarity of 'nominal' relatives in **yó-* is that they can occur with a constituent fronted around the relative pronoun, as in exx. (3.7–9) above:

(3.12) *ná* *āsīt* *rájas* *ná* *u* *vyomā́* *parás* ***yád***
NEG be.IMPF.3SG air.NOM.SG.N NEG and sky.NOM.SG.M beyond REL.NOM.SG.N
'There was no atmosphere nor a heaven **which** [was] beyond' (RV 10.129.1b)

The fronting of constituents above the relative pronoun in 'nominal' relatives is likely to be the genesis of the Balto-Slavic definite adjectival endings (on which see §3.2.2 below).

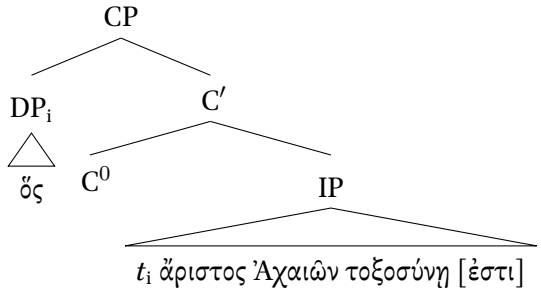
The traditional account of 'nominal' relatives is essentially that these clauses are part of the broader set of 'nominal' sentences, which involve Omission of the Copula (OC).

(3.13) *sá* *janāsa* *índraḥ*
DEM.NOM.SG.M people.VOC.PL.M Indra.NOM.SG.M
'He, oh peoples, [is] Indra' (RV 2.12 *passim*)

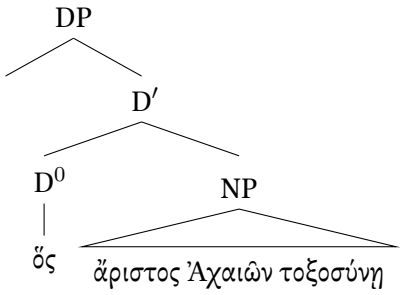
'Nominal' relatives can therefore be understood to instantiate the same structure as other RCs by supplying an elided form of the copula. Some authors, however, following Benveniste (1957/1958), have argued

that **yó-* in this context does not function as a relative pronoun, but as something 'article-like'. The argument follows that we should interpret a sentence such as ex (3.11) not as 'Teucer, who is the best...' but simply as 'Teucer, **the** best..'. The two alternatives are presented in structural terms below:

(3.14) Τεῦκρός θ'...



(3.15) Τεῦκρός θ'...



Ex. (3.14) shows the 'traditional' hypothesis, where ὅς introduces an RC, while in (3.15) ὅς acts as a simple D^0 akin to the definite article. At first glance, either of these analyses is possible. However, if we take into account the case expressed on both the relative/'article-like' pronoun and the nominal it agrees with, we can diagnose which structure is more plausible. In finite OC constructions, the subject obligatorily stands in the **nominative**. Clearly this is the case for exx. (3.10–11); this is the regular pattern for 'nominal' relatives in Ancient Greek and Vedic, even when the constituent the 'nominal' relative describes occurs in a different case:

(3.16) *agním* *tám* *manye* ***yáḥ*** *vásuḥ*
Agni.ACC.SG.M DEM.ACC.SG.M think.PRES.1SG REL.NOM.SG.M good.NOM.SG.M
'I ponder that Agni, **who** [is] good' (RV 5.6.1a)

If we posit a structure such as ex. (3.15) for the relative above, it is a mystery as to how *yaḥ* is assigned nominative case, rather than accusative as the object of *manye* ('I ponder'). Under the analysis in (3.14), the case would be assigned by a phonologically null form of *ásti* ('is'). In these languages, therefore, we should prefer the analysis in (3.14) across the board, unless we wish to posit an entirely arbitrary distinction between 'article-like' **yó-* when the antecedent occurs in the nominative, but a relative pronoun elsewhere.[53]

In Avestan, however, we find examples of the relative pronoun in a 'nominal' relative surfacing in an oblique case, usually the accusative, in agreement with the head noun:[54]

[53] See also Probert (2015: 409–10) for arguments to this effect; I share her conclusions regarding the innovative status of the 'article-like' **yó*.

[54] Examples from West (2011: 82).

(3.17) *tām* ... *vī.adarəsəm* ... ***yəm*** *Mazdąm* *Ahurəm*
DEM.ACC.SG ... see.IMPF.3SG ... REL.ACC.SG Mazda.ACC.SG lord.ACC.SG
'Him I have discerned, Mazda **the** Lord' (Y 45.8)

(3.18) *tāiš* *šiiaoϑənāiš* ***yāiš*** *vahištāiš* *fraēšiiāmahī*
DEM.INS.PL action.INS.PL REL.INS.PL best.INS.PL urge.PRES.1PL
'With these actions, **the** best ones, we urge...' (Y 35.4)

Clearly these are not examples of OC sentences: the case of the antecedent, the relative pronoun and the following epithets are all assigned in the MC. In contrast to the Vedic examples, these case forms are more of a mystery if we posit this as an OC structure. Rather, it seems clear that the forms of *ya-* here introduce simple DPs that are straightforwardly in apposition to the antecedent. In this context, the 'article-like' structure posited in ex. (3.15) is the better explanation.

The situation is complicated, however, by the fact that Avestan also exhibits 'nominal' relatives with regular OC case behaviour:

(3.19) *vaēdōdūm* ... *ahūm* ***yə̄*** *vaŋhāuš* *manaŋhō*
know.IMPV.2PL ... life.ACC.SG REL.NOM.SG good.NOM.SG mind.GEN.SG
'Find for yourselves...a life **which** [is] of good thought' (Y 53.5)

The Avestan data therefore require us to admit both ex. (3.14) and (3.15) as possible structures. The former is unproblematic, since Avestan uses *ya-* as a relative pronoun with regular RCs, and makes extensive use of OC constructions elsewhere. I would argue that what we see in exx. (3.17–18) is the actualisation of an ongoing reanalysis. We have seen already that in clauses where all nominals occur in the nominative, there is a structural ambiguity between a relative OC clause and a simple DP introduce by 'article-like' *ya-*. If speakers abduce the 'article-like' structure of (3.15) from these all-nominative occurrences, and extend it to other case forms, the outcome is tokens such as exx. (3.17–18). I would argue, therefore, that the 'article-like' use of *ya-* in Avestan is most plausibly understood as an innovation: the 'nominal' relative, as attested in other branches, is reanalysed as a simple DP in apposition to the antecedent. Since there is no evidence for this reanalysis in either Vedic or Ancient Greek, I believe the 'article-like' behaviour of Av. *ya-* is not inherited from PIE.

A final feature of relative **yó-* is its use as a **quasi-complementiser**. In short, an invariant form, usually the neuter nom./acc. sg., can be used to introduce various types of subordinate clause that are not, strictly speaking, relatives. In Vedic, *yád* can be used to mean 'when', or 'as', as well as introducing conditional and purpose clauses:[55]

(3.20) ***yád*** *ha* *yánti* *márutaḥ* *sám* *ha* *bruvate*
REL EMPH go.PRES.3PL Maruts.NOM.PL together EMPH speak.PRES.3PL
'**As** the Maruts go, they speak together' (RV 1.37.13)

(3.21) ***yád*** *agne* *syā́m* *ahám* *tvám*, *tvám* *vā* *ghā* *syā́ḥ* *ahám*,
REL Agni.VOC.SG be.OPT.1SG I you you or EMPH be.OPT.2SG I
syúr *te* *satyā́ḥ* *ihá* *āśíṣaḥ*
be.OPT.3PL your true.NOM.PL here wish.NOM.PL

[55] Examples from Macdonell (1916: 242–3).

'**If** I were you, Agni, or you were me, your wishes would come true' (RV 8.44.23)

(3.22)

yád	*nūnám*	*aśyā́m*	*gátim*	*mitrásya*	*yāyām*	*pathā́*
REL	now	reach.OPT.1SG	way.ACC.SG	Mitra.GEN.SG	go.OPT.1SG	path.INS.SG

'**In order to** find a way now, I would go along the path of Mitra' (RV 5.64.3ab)

Elsewhere, this use of a form of **yó-* appears to be mirrored in the first element of AGk. complementiser ὅτι ('that') and conjunction ὅτε ('when'). Additionally, alongside the neuter nom./acc sg. forms of **yó-*, we see forms that are apparently built on oblique cases used as complementisers or conjunctions: these included Skt. *yā́d* 'as long as', *yadā́* 'when', Av. *yā* ('id.') and AGk. ὡς 'that', 'as'.

It is possible that the use of various forms of **yó-* in this way is inherited. Nevertheless, I do not see in this behaviour an 'earlier' use of **yó-* than its function as a relative pronoun. In the first place, although the neut. nom./acc. singular is the basis of several conjunctions and complementisers, the fact that we also see them built on various oblique forms means that we cannot reconstruct a singular form of the proto-complementiser with confidence. However, what these forms appear to share is that their morphology is largely the same as we find synchronically in **adverbs**: the neut. nom./acc. is productive in forming adverbs in both Greek and Sanskrit, while we also find fossilised instrumentals (e.g., ἄν-ω 'upward', κάτ-ω 'downward') used adverbially. Taken as adverbs built on a relative stem, these forms could be interpreted in the first place as something like 'in which way', 'how', 'as' etc., the use which is retained in Vedic *yathā́* and AGk. ὡς. Moreover, from a syntactic point of view, subordinate clauses formed with quasi-complementiser **yó-* are similar to those in which **yó-* has a strictly pronominal use. The relevant form most commonly occurs clause-initially; in Vedic, at least, it may optionally be preceded by one or two fronted constituents, just like the 'true' relative pronoun. Quasi-complementiser **yó-* also regularly forms correlative clauses with a corresponding adverbial built on the demonstrative stem **só-/tó-*.

To me, all this implies that in PIE, various adverbial forms of **yó-* were used to introduce a variety of subordinate clauses. This does not, however, mean that these forms were, strictly speaking, complementisers base-generated in C° in PIE. Admittedly, it is quite challenging to diagnose the difference between a subordinating adverb fronted to (somewhere in) [Spec, CP] and something in C°. The obvious test would be to look for sentences with an overt C° co-occurring with adverbial **yó-*. Unfortunately, there are no overt complementisers that can be securely reconstructed back to PIE (Windhearn, 2021: 7).[56] Each language seems to build complementisers from different lexical material, and often during the course of its attested history (Hackstein, 2013). In Ancient Greek, almost all the elements that appear as complementisers synchronically can be traced back to forms of **yó-*. The same is largely true of Sanskrit, with the exception of quotative *íti*, which has a distribution that is considerably different to that of *yád*: namely, it occurs generally clause-finally rather than clause-initially. Other languages build their complementisers on other nominal stems, such as **só-/*tó-* or **kʷó-*, sentence connectives such as **dé*, or

[56]The possible exception to this is the enclitic connective **kʷe*, which appears to have some reconstructable complementiser properties, and is clearly not adverbial (Keydana, 2018: 2215–6). However, these data are obscured by (a) interpretational challenges concerning subordination, and (b) the status of **kʷe* as a clitic, whose surface position involves the interaction of syntax and prosody (see §3.1). As such, I am not sure that it can help us here, though further studies may seek to shed some light on this possibility.

material that is not inherited at all. This does not necessarily mean that PIE lacked overt complementisers, simply that if they did exist they were comprehensively supplanted in the attested languages.

Even if we allow for the quasi-complementiser behaviour of **yó-* in PIE itself, I am inclined to believe that this behaviour is an innovation following from the use of **yó-* as a relative pronoun. As far as grammaticalisation is concerned, it is quite common for an inflected relative pronoun to develop into an uninflected complementiser; yet according to Hendery (2012: 153–4), there is only one supposed example of an uninflected complementiser becoming an inflected pronoun.[57] Furthermore, if we were to assume that the 'original' form was an uninflected complementiser, **yó(d)*, we would have to discount the shared etymology with the anaphoric pronoun **e(y)-*, a decision which lacks motivation elsewhere. For these reasons, I will continue to use the term 'quasi-complementiser' to describe such uses of **yó-*. Although across all these languages the meaning has gone beyond 'in which way' *vel sim.*, to something we would certainly translate with a complementiser in English, from a syntactic perspective I do not believe that we are reconstructing something originally base-generated in C°.

In sum, where a descendant of **yó-* is directly attested as an inflecting relative pronoun, it is not possible to attribute this use to an earlier, non-relative function. The data from these languages therefore offer strong support in favour of the hypothesis that **yó-* was used a relative pronoun in PIE.

3.2.2 Non-relative uses of **yó-*

3.2.2.1 Demonstrative

As mentioned above, we may accept that **yó-* is ultimately derived from the same root as the demonstratives in **(e)y-*. However, there is only one branch of IE in which something conceivably reconstructable as **yó-* is attested with a purely demonstrative use, and that is Balto-Slavic. Both OCS and Lithuanian inherit a demonstrative/personal pronoun that appears to be at least partially derived from **yó-*; these are shown in Figures 3.1 and 3.2 below.

The evidence from these languages may lead us to hypothesise that **yó-* was a demonstrative or anaphoric pronoun in PIE; that is, it had the same function that survives in the descendants of **(e)y-* and did not function as a relative pronoun. There are two pieces of evidence that may support this hypothesis: (1) the shared etymology of **yó-* and **(e)y-*, and (2) the fact that anaphoric pronouns very commonly become relative pronouns, but not *vice versa*. However, there are several factors that should make us question this hypothesis.

The first is that from a phonological point of view, while parts of these paradigms could be derived from **yó-*, there are others that could be derived from **(e)y-*. Indeed, Pokorny (1959) and Fraenkel (1960) attribute Lit. *jis* entirely to **(e)y-*. Fraenkel notes only that Lit. conjunctions *jéi* ('if'), *jeĩb* ('in order to') and *jóg* ('that') belong to the *'Demonstrativstamm' *yo-, *ye-* (1960: 194), while Pokorny adds that the OCS *-jǫ, -ję, -jь* specifically are better taken from **yó-*. More recently, Derksen seems to suggest that the

[57] This is the nonstandard PDE use of 'that's' as a possessive relative in sentences such as 'The language that's complementiser became a pronoun' (Seppänen, 1997). As Hendery (2012: 153) notes, this example is slightly dubious because 'that's' may have developed by analogy with 'whose' (who + 's). Equally, it's possible that English 'that' is underlyingly a relative pronoun (Kayne, 2014), in which case the point is moot.

Sg.	Masc.	Fem.	Neut.
Nom.	**jь*	**ja*	**je*
Acc.	*jь*	*jǫ*	*je*
Gen.	*jego*	*jejȩ*	*jego*
Dat.	*jemu*	*jeji*	*jemu*
Ins.	*jimь*	*jejǫ*	*jimь*
Loc.	*jemь*	*jeji*	*jemь*
Pl.	**Masc.**	**Fem.**	**Neut.**
Nom.	**ji*	*jȩ*	*ja*
Acc.	*jȩ*	*jȩ*	*ja*
Gen.	*jichъ*	*jichъ*	*jichъ*
Dat.	*jimъ*	*jimъ*	*jimъ*
Ins.	*jimi*	*jimi*	*jimi*
Loc.	*jichъ*	*jichъ*	*jichъ*

Table 3.1: Old Church Slavonic 3sg personal pronoun (Lunt, 2001: 62).

Sg.	Masc.	Fem.
Nom.	*jìs*	*jì*
Acc.	*jį̃*	*ją̃*
Gen.	*jõ*	*jõs*
Dat.	*jám*	*jái* (*jaĩ*)
Ins.	*juõ* (*júo*)	*jà* (*ją́*)
Loc.	*jamè* (*jam̃*)	*jojè* (*jõj*)
Ill.	*ján*	*jõn*
Pl.	**Masc.**	**Fem.**
Nom.	*jiẽ* (*jíe*)	*jì*
Acc.	*juõs* (*júos*)	*jàs* (*jás, jáis, jós*)
Gen.	*j̃u*	*j̃u*
Dat.	*jíems*	*jóms*
Ins.	*jaĩs*	*jomìs* (*jõms*)
Loc.	*juosè*	*josè*
Ill.	*júosna*	*jósna*

Table 3.2: Lithuanian 3sg personal pronoun (Senn, 1966: 191–2).

paradigm of OCS **jь* can be derived wholly from **yó-* (2007: 208), while the paradigm for Lit. *jis* exhibits a mixture of forms derived from **yó-* and **e(y)-* (2014: 212).

Indeed, in his later comment on the matter, Derksen (2014: 212) 'prefer[s] to assume that both in Baltic and Slavic the demonstrative pronoun **h₁e* [i.e. **(e)y-*] and the relative pronoun [i.e. **yó-*] were conflated'. This was the view of Vailliant (1950–1977: II/2, 423), held also by Flier (1974: 67); the latter explicitly comments on the fact that OCS relative pronoun is formally identical to the demonstrative in **jь*, only with the compulsory addition of the 'intensifying suffix' *-že* (cf. ex. 3.6 above).[58] This points towards a functional as well as a formal conflation between **yó-* and **(e)y-*.

The Balto-Slavic data are therefore somewhat obscure. Nevertheless, we can see quite clearly that at least some forms derivable from **yó-* are attested with an anaphoric value, and this remains to be accounted for. Some authors argue that Proto-Balto-Slavic inherited **yó-* with its relative function, which subsequently developed into the demonstrative.[59] Others argue that **yó-* was inherited with a

[58] Little can be concluded from the fact that Lithuanian has a relative pronoun in *k-* < **kʷó-*, as given its very late attestation (16th century CE), this could well have been an internal development akin to that of English *who, which*. This hypothesis is further strengthened by the appearance of the aforementioned Lit. conjunctions that appear to show forms derived from **yó-*.

[59] This view is perhaps more popular among Indo-Europeanists, starting with Brugmann (1886–1893: II/2, 331). It is also the view held by Hermann (1912), Fraenkel (1947), Zinkečius (1957) , C. Koch (1992) and Petit (2009).

demonstrative function, and subsequently became a relative pronoun in OCS, via the same grammaticalisation pathway as, e.g., PIE **só/tó-* in some dialects of Ancient Greek.[60] Of these two, the latter may seem more likely typologically, but one crucial piece of evidence suggests that **yó-* was inherited with, at the very least, some relative function: this is the Balto-Slavic **definite adjective declension**.

This declension, preserved most clearly in Lithuanian, is formed by the combination of an adjective + *jis*: the very same morpheme as the personal pronoun in table 3.2 above.

	Basic		Definite	
	Masc.	Fem.	Masc.	Fem.
Nom.	*gẽras*	*gerà*	*geràsis*	*gerój̨i*
Acc.	*gẽrą*	*gẽrą*	*gẽrąjį*	*gẽrąją*
Gen.	*gẽro*	*gerõs*	*gẽrojo*	*gerõsios*
Dat.	*gerám*	*gẽrai*	*gerájam*	*gẽrajai*
Ins.	*gerù*	*gerà*	*gerùoju*	*gerą́ja*
Loc.	*geramè*	*gerojè*	*gerãjame*	*gerõjoje*
Ill.	*gerañ*	*gerõn*	*geraĩan*	*gerõjon*

Table 3.3: Basic (indefinite) and Definite forms of Lit. *gẽr-* 'good' (Senn, 1966: 143, 163).

In the general absence of a definite article in this language family, the definite form of the adjective, as the name suggests, marks the NP it describes as definite, often in an emphatic way.

(3.23) *gẽras* *vaĩkas*
good.NOM.SG.INDEF child.NOM.SG
'A good child'

(3.24) *geràsis* *vaĩkas*
good.NOM.SG.DEF child.NOM.SG
'**The** good child' (Senn, 1966: 360)

The definite adjective declension therefore provides another piece of the puzzle concerning the original function of Proto-Balto-Slavic **yó-*. Those who argue that **yó-* was in the first place an anaphoric pronoun hold that definite NPs such as:

(3.25) *geràsis* *žmogùs*
good.NOM.SG.DEF man.NOM.SG
'The good man'

should historically be analysed as:

[60] This view is argued in the first place by Miklosich (1868–1879), and finds subsequent support in Otrębski (1956–1965), Rosinas (1975), Ballester (2001).

(3.26) **geras=jis* *žmogùs*
good.NOM.SG=PRO.NOM.SG man.NOM.SG
'Man=he good' > 'The good man' (Petit, 2009: 327)

On the other hand, those who hold that **yó-* was a relative pronoun hypothesise that the historical analysis should be:

(3.27) **geras=jis* *žmogùs*
good.NOM.SG=REL.NOM.SG man.NOM.SG
'Man=who [is] good' > 'The good man' (Petit, 2009: 340)

Although there has been much back and forth on this question, the arguments in favour of an original relative function of **yó-* are quite compelling. In the interest of space I do not enumerate these in full, but summarise from Petit's (2009) thorough overview of the arguments put forward by various authors:

1. Since anaphoric *jis* is strictly used as a personal pronoun that **cannot** occur as a determiner governing a noun, its apparent use as a determiner in the structure posited in ex. (3.26) lacks plausibility (Petit, 2009: 337–8)
2. Whereas the structure hypothesised in ex. (3.26) is unattested elsewhere in IE, there are clear parallels for (3.27) in the 'nominal' relatives discussed in §3.2.1 above, cf. especially the data from Avestan in exx. (3.17–8) (Petit, 2009: 340–1, 345–8)
3. Further to the Avestan data, Iranian later shows a development of a relative pronoun into a determiner (the '*ezāfe*' construction) which provides a close typological parallel from within the IE family (Petit, 2009: 341–5)

In summary, alongside the OCS relative pronoun *i* + *že*, the definite adjectival declension clearly points to a descendant of **yó-* acting as a relative pronoun in Proto-Balto-Slavic. While we also have evidence for an anaphoric function in the personal pronouns in **jь*, the fact that its paradigm is a conflation of **(e)y-* and **yó-* weakens the argument that the anaphoric meaning is somehow the 'original'.

We should also remember that, despite their various archaic tendencies, the Balto-Slavic languages are attested significantly later than almost all the other languages that clearly exhibit a relative pronoun descended from **yó-*. While we must not ignore the Balto-Slavic data, the same is true of, e.g., Celtic, Greek and Indo-Iranian. On balance, therefore, I do not believe that a contestable hypothesis about the inherited value of **yó-* in Balto-Slavic is enough to supersede the clear-cut evidence from the other branches that **yó-* was used as a relative pronoun in PIE.

3.2.2.2 Sentential connective

Another purported use of **yó-* in PIE is that of a **sentential connective.** This connective is supposedly attested in both an orthotonic form **yó*, and an enclitic form **yo*. In this section, I discuss the evidence for such a particle and how it could be related to the relative pronoun.

The strongest evidence for such a particle comes from Hittite, where it is attested as *-ya* when preceded by a vowel, and *-a* with gemination of the preceding consonant elsewhere. It has the function of a conjunction usually translatable as 'and'.[61]

(3.28) *nu=tta* DINGIR.MEŠ ^{d}É.A-*aš=**ša*** *ḫattannaš* LUGAL-*uš*
CONN=you.ACC.SG god.NOM.PL Ea.NOM.SG=and wisdom.GEN.SG king.NOM.SG
aššuli *paḫšantaru*
health.DAT.-LOC.SG protect.IMPV.3PL
'May the gods **and** Ea, the king of wisdom, keep you in good health'
(HKM 3:18–20)

(3.29) ^{LÚ}A.ZU=***ya*** *kuššan* *apāš=pat* *pāi*
doctor.GEN.SG=and fee.ACC.SG DEM.NOM.SG.MASC=EMPH give.PRES.3SG
'...**and** he will (also) pay the doctor's fee himself.' (KBo 6.2 i 19)

Hittite -(*y*)*a* is straightforwardly enclitic. Its syntax is more or less equivalent to that of Latin *-que*, AGk. τε or Skt. *ca* ('and'): namely, it follows the second of two coordinated elements (X Y -(*y*)*a*). As Hoffner & Melchert note, -(*y*)*a* is 'the only connective which joins *individual* words' (2008: 400, emphasis theirs), as in ex. (3.28). It is also used to conjoin clauses, in which case it regularly appears after the first accented word in a clause, as in ex. (3.29).[62]

Elsewhere, we supposedly have the orthotonic form of the connective **yó* in Germanic, e.g., Got. *jah* ('and'), and Toch. A *yo* ('id.') (Dunkel, 1982: 180–1).

(3.30) ***jah*** *jabai* *taihswo* *þeina* *handus* *marzjai* *þuk*,
and if right.NOM.SG your hand.NOM.SG offend.OPT.3SG you.ACC.SG,
afmait *þo* ***jah*** *wairp* *af* *þus*
cut-off.IMPV.2SG it and throw.IMPV.2SG from you.DAT.SG
'**And** if your right hand offends you, cut it off **and** cast it away from you' (Matthew 5:30)

(3.31) *oṅk* ***yo*** *k$_u$li*
man and woman
'A man **and** a woman' (TS 340 b2)

Syntactically, these forms stand quite apart from Hittite **-(*y*)*a*. They do not appear to be enclitic, and tend to be placed after the first of two coordinated items (X *yo*/*jah* Y); in this way their behaviour is more like Lat. *et* or AGk. καί ('and'). This is a strict rule for Got. *jah*; Toch. A *yo*, however, seems able to appear after the second of two coordinated elements, but only in verse (Thomas, 1966: 269):

(3.32) (*:kā*)*ruṇikāṃ* *pältsäkyo* *knānmuneyo* *āṣträṃn* ***yo*** *r*(*i*)*ṣakk* *oki:a*
compassionate.OBL.SG spirit.INS.SG knowledge.INS.SG pure.OBL.SG and *rishi* like
'With a compassionate mind **and** with a pure knowledge, like a sage...' (TS 77 b5)

[61] Examples from Hoffner & Melchert (2008: 400–1).

[62] Not counting the conjunctions *takku*, *mān*, on which see Hoffner & Melchert (2008: 395,400). I discuss the complexities of the syntax of Hit. -(*y*)*a*, and its position in the left periphery, in §4.4.

Some authors, including Thomas (1966), have argued this syntagm to be a conservatism, though it is also possible that it is a a feature of poetic genre. The picture is further complicated by the existence of the instrumental suffix *-yo*, whose etymology also remains obscure.

Some (e.g., Dunkel, 1982) add to this set of *comparanda* the Mycenaean particle that is written variously as *o* or *jo*, apparently in free variation,[63] and appears to fill the pragmatic function of 'introducing' the discourse.[64]

(3.33)	***o*-*wi*-*de***	*pu$_2$*-*ke*-*qi*-*ri*	*o*-*te*	*wa*-*na*-*ka*	*te*-*ke*	*au*-*ke*-*wa*	*da*-*mo*-*ko*-*ro*
	and?-see.AOR.3SG	Phuge^{w}rins	when	king.NOM.SG.M	set.AOR.3SG	Augewas	*damokoron*

'Phuge^{w}rins saw when the king appointed Augewas as *damokoros*...' (PY Ta 711.1)

Yet again, we have in Myc. (*j*)*o* something that is clearly not enclitic,[65] standing clause-initially. And unlike the *comparanda* from Tocharian A and Gothic, Myc. (*j*)*o* is never used to co-ordinate anything, whether phrasal or clausal.

If we do accept the cognacy of such connectives, the question arises: do they share an origin with the relative pronouns descended from **yó*-? And if so, how are the two uses—the connective and the relative—related? Several authors, including Watkins (1963) and Dunkel (1982) reconstruct a sentential connective **yó*/**yo* for PIE. Indeed, Watkins (1963) goes so far as to use the putative enclitic connective **yo* to account for the forms of the relative verb in Old Irish (a topic I return to in §5.1), ultimately suggesting that its relative subordinating value arose secondarily out of its original meaning as a connective. While I see the appeal of such hypotheses, I remain unconvinced for several reasons.

First, there are some issues of etymology. It is not absolutely certain that Hittite -(*y*)*a* is derived from a putative PIE **yo*. While this was the favoured etymology for Puhvel (1984: 7–8), more recently Kloekhorst (2007: 378–9) rejects it altogether, hypothesising instead a Proto-Anatolian form **Ho*, most likely from PIE **h$_3$e*. Moreover, with the possible exceptions of Myc. *to-so-jo* and Toch. A *yo* in poetry, outside Anatolian there are no attested cognates of -(*y*)*a* that behave in the same way syntactically. We have orthotonic connectives from **yó*- (as in Gothic and Tocharian A), and enclitic relative pronouns from **yo*- (as in (Pre-)Old Irish and Gaulish), but no other enclitic connectives from **yo*. Taken together with Kloekhorst's phonological misgivings, I am inclined to believe that the etymology of Hit. **-(*y*)*a* cannot be secured outside the Anatolian branch.

The evidence from Mycenaean is likewise unconvincing. As mentioned above, orthotonic (*j*)*o* behaves not at all like the other *comparanda* mentioned in this section, in that it does not coordinate anything. For further discussion, see Thompson (2002), who suggests instead that (*j*)*o* is—as traditionally conceived—descended from the relative pronoun **yó*-. He argues that we should understand it as a relative adverb /ho:/, 'in which way', 'how', which moves to [Spec, CP] and triggers raising of the verb to C°. This brings it in line with the other adverbial forms of **yó*- mentioned in §3.2.1 above. More recently,

[63] Though see Probert (2008) for a potential syntactic explanation.

[64] Example from Thompson (2002: 317).

[65] There is one, disputable use of Myc. *jo* where it appears to be enclitic, substituting for *-de* ('and') in the phrase *to-so-jo* (Bader, 1975). The alternative hypothesis is that *to-so-jo* is simply the genitive of *tos-* 'this'. The latter is, admittedly, syntactically inexplicable; yet the hypothesis that this an enclitic from of **yo* unparalleled elsewhere in the Greek corpus, seems unlikely.

Probert (2008) argues that Myc. (*j*)*o* is straightforwardly a relative pronoun which inflects according to its case function. This leaves only the very questionable instance of enclitic *-jo* in *to-so-jo*, which does not constitute anything reliable enough to reconstruct from.

We are left then with Tocharian A and Germanic. Considering the phylogenetic distance between these languages, and their late dates of attestation, it seems quite unlikely that these two branches alone inherited and maintained the 'original' function of the connective **yó*, to the exclusion of all other branches. Such an inheritance would be all the more remarkable considering that even Tocharian B shows no evidence for such a form. Further studies may seek to account for the Germanic and Tocharian forms, but I do not see reconstruction of connective **yó*, superseding relative **yó-*, as the most prudent response to such isolated data.

To conclude this section then, I will proceed with the understanding that the only securely reconstructable function of PIE **yó-* is that of a relative pronoun. I will return to the implication of this finding in §3.4 below.

3.3 **kʷí-/*kʷó-*

As I mentioned above, **kʷí-/*kʷó-* is attested in every branch of IE as an interrogative and/or indefinite pronoun, but in only three of these as a relative pronoun from their earliest stages (Anatolian, Italic, Tocharian). A straightforward application of the comparative method will therefore lead to our reconstructing **kʷí-/*kʷó-* as an interrogative/indefinite; its use as a relative pronoun is somewhat more perplexing. Theories from the late 19th and early 20th centuries, before the decipherment of Hittite and Tocharian, tended to suppose that **kʷí-/*kʷó-* was not a relative pronoun in PIE; instead, focusing on Latin, they aimed to explain how the relative use may have developed from the interrogative (e.g., Wegener, 1874) or indefinite (e.g., Kroll, 1910) functions. Since the Hittite evidence in particular supported the theory that the use of **kʷí-/*kʷó-* as a relative pronoun was not a post-PIE innovation, scholars have remained somewhat conflicted on the matter. I return to the stances of various authors on this question in §3.4 below. For now I will work from the position that its relativising function is plausibly inherited. More precisely, I believe that the descendants of **kʷí-/*kʷó-* can be used as *comparanda* when reconstructing the behaviour of relative pronouns in PIE. For this reason, it is prudent to summarise the grammatical features of **kʷí-/*kʷó-* and its descendants; as with **yó-* above, I will discuss both its relative and non-relative uses.

3.3.1 Vocalism

There is evidence for both an *o*-stem, **kʷó-*, and an *i*-stem **kʷí-/*kʷé-*,[66] variant of this pronoun. Both options are attested, for example, in the Latin paradigm:

- i-stem forms: *quis, quid, quem* (< **kʷím*), *quibus*

[66] **kʷí-* is found only in the nominative and accusative forms of the *i*-stem variant, while **kʷé-* is the basis of the rest of the paradigm.

- o-stem forms: *quī* (< *k^wói*), *quod, cum* (< *k^wóm*), *quō, quorum*

In languages that attest both vocalisms, there is rarely a discernible grammatical pattern to the distribution of *i*-stem and *o*-stem forms. In some languages one vocalism is generalised throughout the paradigm. This is the case for Hittite, which generalises the *i*-stem forms:

	Singular	Plural
nom. com.	***kuiš***	***kuiēš***, *kuēš*, (*kueuš*)
acc. com.	***kuin***	***kuiuš***, *kuiēš*, (*kueuš*, *kuiš*, *kue*)
n.-a. neut	***kuit***	***kue*** (*kuie*)
gen.	***kuēl***	—
d.-l.	***kuedani***	***kuedaš***
abl.	*kuēz*(*za*)	

Table 3.4: Attested forms of the Hittite interrogative/relative Pronoun (Hoffner & Melchert, 2008: 149).

In Gothic, on the other hand, the interrogative pronoun is primarily based on the *o*-stem vocalism, with the *i*-stem preserved only in masc./neut. gen. *hvis* and fem. dat. *hvizai*.

Singular	Masculine	Feminine	Neuter
Nom.	*hvas*	*hvō*	*hva*
Acc.	*hvana*	*hvō*	*hva*
Gen.	*hvis*	—	*hvis*
Dat.	*hvamma*	*hvizái*	*hvamma*

Table 3.5: Attested forms of the Gothic interrogative pronoun.

It has been suggested, foremost by Sihler (1995: 395–400) and Ringe (2017: 69), that PIE had two distinct paradigms: an *i*-stem substantive **pronoun** (a full NP, Eng. 'who?') and an *o*-stem pronominal **adjective** (Eng. 'which?'): I will refer to this as the **Sihler-Ringe Hypothesis**. A key part of their claim rests on the forms of the Latin interrogative *quis* < *k^wí-s* vs. *quī* < *k^wó-i*. Ringe argues that *quis* is properly an interrogative pronoun, while *quī* is an interrogative adjective. This distinction is taught in Latin reference materials such as Kennedy (1962: §99), who gives the following examples:

(3.34) Quis venit? *Who comes?* (**quis**, interrogative)

(3.35) Quī homō venit? *Which man comes?* (**quī**, interrogative)

However, as noted by Adams (2016: 48), '[t]he school rule, that the distinction between *quis* and *qui* in interrogative clauses is that the former is substantival and the latter adjectival, has little foundation...except in the neuter singular'. This is particularly evident in early Latin poetry: according to Lodge

(1962: 718), the string *quis homo* occurs in Plautus no fewer than 17 times. On the other hand, interrogative *qui* is preferred under either of two conditions, regardless of its pronominal vs. adjectival use: (i) in indirect questions (ii) when the following word begins with /s/ (Adams, 2016: 48–52). As such, the evidence from archaic Latin offers little support to the Sihler-Ringe hypothesis.

Sihler (1995: 397) also brings purported evidence from Avestan to bear, claiming that:

> In Avestan, such forms as *čahmāi* dat., and *čahyā* gen., from **k^{w}e-*, are pronouns, interrog. and indef.; *kahmāi*, *kahyā* by contrast, from **k^{w}o-*, are adjectival.

Yet this assertion is also rather questionable. The following examples show at the very least that there are exceptions to Sihler's rule in Avestan:[67]

(3.36) *ciš* *ahī* ***kahiiā*** *ahī*
INT.NOM.SG be.PRES.2SG INT.GEN.SG be.PRES.2SG
"Who are you and **whose** are you?" (Y. 43.7)

(3.37) *kat* *mōi* *urvā* *isē* ***cahiiā*** *avaŋhō*
Q PARTICLE PRO.DAT.1SG soul.NOM.SG ?seek.PRES.3SG INDEF.N.GEN.SG support.N.GEN.SG
"Does my soul seek **any** help?" (Y. 50.1)

In ex. (3.36), we see the use of *kahiiā* < **k^{w}ó-syo* as a substantive, while (3.37) shows *cahiiā* < **k^{w}é-syo* used as an adjective, in direct contradiction to Sihler.

Thus, I would argue that there is insufficient evidence from either Latin or Avestan for us to reconstruct the original grammatical function of *i*-stem vs. *o*-stem forms of **k^{w}-*. Though it is tempting make this ablaut pattern make grammatical sense, I do not accept the Sihler-Ringe hypothesis, and leave the investigation of this alternation to further study. I shall make no further comment on the matter, continuing to use **k^{w}ó-* as a stand-in for all vocalisms, since we cannot posit with confidence any difference in grammatical function between the *o*-stem and the *i*-stem form.

3.3.2 Relative uses of **k^{w}ó-*

Descendants of **k^{w}ó-* are used as relative pronouns in many IE branches, though these uses often developed within from their interrogative/indefinite use within the attested history of the languages. However, those branches which use **k^{w}ó-* from their earliest attestations are Hittite, Italic and Tocharian.

In these languages, relative **k^{w}ó-* behaves in a way that is syntactically very similar to relative **yó-* discussed in§3.2.1 above. It agrees with its antecedent in number and gender; it introduces both 'plain' RCs and correlative RCs. It generally occurs clause-initially, but can optionally be preceded by another topicalised element:[68]

[67] I thank Timothy Barnes for pointing me to these examples.
[68] Hittite example from Huggard (2011: 96), Tocharian from Adams (2015: 30).

(3.38) *iouesat* *deiuos* ***qoi*** *med* *mitat* *nei*
call-to-witness.PRES.3SG god.ACC.PL.M REL.NOM.SG.M me.ACC.SG send.PRES.3SG if-not
ted *endo* *cosmis* *uirco* *sied*
you.ACC.SG in friendly.NOM.SG.F maiden.NOM.SG.F be.PRES.SUBJ.3SG
'**The one who** sends me calls the gods to witness: if the young girl is not kind towards you...'
(CIL I² 4)

(3.39) *nu=šši=šš̌an* ***kuit*** *šaḫḫan* LUGAL-*uš* *dāi* *nu*
CONN=him.DAT=PTC REL.ACC.SG.N service.ACC.SG.N king.NOM.SG.M put.PRES.3SG CONN
apāt *ēšš̌ai*
DEM.ACC.SG.N do.PRES.3SG
'Any service **which** the King imposes on him, he shall do that.' (KBo 6.4 iv 15-16)

(3.40) ***k$_u$ce*** *te* [*ma*]*nt* *wñāwa,* *tu* *ṅke* *weñau* *anaiśai*
REL.ACC.SG.N thus EMPH speak.PRET.1SG, DEM.ACC.SG now speak.SUBJ.1SG clearly
'**What** I have so said, that will I now speak clearly' (PK-AS 7B-a5/6^c)

From these patterns, it would seem that we can treat relative **k^{w}ó-*, like **yó-*, as generally moving to the left-periphery,[69] with the possibility of other constituents being topicalised/focalised around it.

Finally, like **yó-*, **k^{w}ó-* can also be used to introduce 'nominal' relatives:[70]

(3.41) *salvete,* *Athenae,* ***quae*** *nutrices* *Graeciae*
be-well.IMPV.2PL Athens.VOC.PL.F REL.NOM.PL.F nurse.NOM.PL.F Greece.GEN.SG.F
'Greetings, Athens, (you) who (are) the nurse of Greece' (Plaut. *Stich.* 649)

(3.42) DINGIR.LÚ.MEŠ DINGIR.MUNUS.MEŠ *ŠA* LUGAL *U* MUNUS.LUGAL-*UTTI*
gods goddesses of king and queen
kuiēš *daranteš* ***kuiēš*** *ŪL* *daranteš*
REL.NOM.PL invoked.PPL.NOM.PL REL.NOM.PL NEG invoked.PPL.NOM.PL
'Gods and Goddesses of the King and Queen, (those) **who** (are) invoked and (those) **who** (are) not invoked' (KUB 6.46 iii 42–3)

All the languages that have 'nominal' relatives in **k^{w}ó-* strictly obey the case rule of OC clauses: in finite 'nominal' relatives, **k^{w}ó-* and the relevant nominals within the RC occur in the nominative case regardless of the case of the antecedent, as in the examples above. This is to say there are no 'article-like' uses of descendants of **k^{w}ó-* akin to that of Avestan *ya-*.

Additionally, in a fashion similar to that of **yó-*, those languages that use **k^{w}ó-* as a relative pronoun also form (quasi-)complementisers from a petrified form of the nominal stem. This is seen most clearly with Latin *quod* 'that', 'because' and *cum* < *quom* 'since', 'when', but also occurs with Hittite *kuit* 'that' (Hoffner & Melchert, 2008: 426):

[69] Although there is some question on this matter in Hittite, discussed in §3.3.2.1 below.

[70] Latin example from Pompei (2014: 441), Hittite from Hoffner & Melchert (2008: 426). Although 'nominal' relatives are rarer in Classical Latin, they are reasonably well attested in the earlier authors (Benveniste, 1966: 220).

(3.43) *maḫḫan=ma* LÚ.MEŠ ^{URU}*Aššur* *auēr* URU.DIDLI.ḪI.A BÀD=*kan* ***kuit***
when=but people-of-Aššur see.PRET.3PL cities fortified **that**
zaḫḫiyaz *katta* *daškeuwan* *teḫḫun*
battle.ABL.SG under take.SUP put(=begin).PRET.1SG
'But when the Assyrians saw **that** I had begun to capture fortified cities in battle'

(KBo 4.4 iv 28–9)

Moreover, as with descendants of **yó*- that are used as (quasi-)complementisers, those built on the **k*^{w}*ó*- stem expand their use as a complementiser within the attested history of the language. For the same reasons as with **yó*-, I do not see it as at all plausible that **k*^{w}*ó*- originated as a complementiser. Indeed, the claim is even less plausible for **k*^{w}*ó*- since this use of *kuit* is limited to the later Hittite texts. Yet again, it is the neuter nom./acc. sg. that we find here,[71] suggesting an original adverbial function.

In terms of its behaviour as a relative pronoun, then, **k*^{w}*ó*- is for the most part identical to **yó*-, at least as far as our reconstructions are concerned. There are, however, three features of **k*^{w}*ó*-clauses that appear to separate them from **yó*-. These are as follows:

1. Languages that use **k*^{w}*ó*- as a relative pronoun exhibit a higher proportion of correlative clauses in their earliest attestations, and a lower proportion of 'plain' relatives, than languages which have relative **yó*-
2. Correlative clauses formed with **k*^{w}*ó*- are more likely to occur in the order RC-MC, while correlatives in **yó*- are more likely to occur in the order MC-RC
3. RCs introduced by **k*^{w}*ó*- are more frequently restrictive (or inherently maximalising) in the earliest stages of their respective languages than those introduced by *yó*-, which are more likely to be appositive

The earliest discussion in this vein was Sturtevant (1930), who argued that taken together, these phenomena should lead us to reconstruct two different 'types' of RC for PIE: a restrictive type introduced by **k*^{w}*ó*- and an appositive type introduce by **yó*-. Sturtevant reached these conclusions by comparing Hittite, Old Latin and Sabellic texts with Greek evidence from the Homeric poems and Sanskrit evidence from the Rig Veda. The difficulty in building Sturtevant's observations into our reconstructions, however, is that these texts fall into two quite distinct categories. The oldest Latin and Hittite texts are prosaic and often legal, while the Ancient Greek and Sanskrit texts are poetic. As such, it may well be the case that what we are observing here is not a difference in the features of **yó*- vs. **k*^{w}*ó*-, but rather a difference in genre between texts (Hock, 1993, 1997, 2000; Clackson, 2007; Probert, 2015). A further investigation of this possibility for each of the three points of difference is therefore warranted.

Perhaps the easiest of these patterns to explain by genre is the variation between restrictive and appositive RCs. In legal texts, the function of the RC is generally to specify the objects or individuals relevant to a particular case, as in the following:

[71] There is evidence for a possible neuter singular form **k*^{w}*óm*/**k*^{w}*ím* alternating with **k*^{w}*ód*/**k*^{w}*íd*, as attested in Vedic *kím* vs. *kád* (Sihler, 1995: 398).

(3.44) *cui* *testimonium* *defuerit,* ***is*** *tertiis*
REL.DAT.SG.M testimony.NOM.SG.N be-lacking.FUT.PERF.3SG DEM.NOM.SG.M third.ABL.PL.M
diebus *ob* *portum* *obvagulatum* *ito*
day.ABL.PL.M before door.ACC.SG.M demand.SUP go.IMPV.3SG
'Who(ever) lacks testimony, let **him** go to the door (of the witness) every third day to demand it' (Twelve Tables II.3)

Strictly speaking this is an example of an inherently maximalising RC rather rather than a restrictive, but both of these can be set in opposition to appositive RCs, which are very rare in this genre. In poetry, on the other hand, the RC can serve to present elaborate descriptions of characters already known to the audience. To quote Hock (2000: 177, emphasis mine):

> Appositive RCs are felicitous under two conditions: (i) they provide something like an embellishment, and (ii) they invite the 'inference' (à la Grice 1975) that there is a special reason for providing information about a **well-defined entity that ordinarily does not require further characterization.**

Clearly these conditions vary with genre. In fact, even within Sanskrit, in the earliest prose works, characterised by Hock as 'turgid' and 'didactic' (2000: 178), we see relatively fewer appositive RCs than in the Vedic hymns. Elsewhere in IE, Probert (2015: 440) notes that 'in Greek and in Latin, inherently maximalizing constructions are prominent in prescriptive texts because these constructions are good for defining categories of people or things to whom, or to which, prescriptions are to apply.' I would argue from these facts that at the very least, the relative frequencies of restrictive (+maximalising) vs. appositive RCs vary with genre. Given this is the case, it does not seem that we have sufficient evidence to posit a historic connection between the possible forms of the relative pronoun and the restrictivity of their clauses in the proto-language.

Once we ascribe variation in the number of restrictive vs. appositive RCs to genre, there is good reason to believe that the variation between pre- and postposed RCs is likewise, a reconstructional phantom. There is a cross-linguistic tendency for appositive RCs to follow their antecedent, which is generally introduced in the MC (Lehmann, 1984: 278). This tendency is so overwhelming that it has been argued to be a universal, and there is a substantial theoretical literature accounting for this in Minimalist terms (de Vries, 2002, 2006; del Gobbo, 2007, 2017). I discuss the position of appositives RCs in PIE in §5.6. For now it is sufficient to note that the prevalence of appositive RCs would imply a prevalence in post-posed RCs: both patterns are in turn explained purely by genre.

This leaves only one supposed syntactic difference between **yó-* and **kʷó-* RCs: 'plain' relatives vs. correlatives. The problem with this observation is that while 'plain' relatives are more common in the earliest Greek and Sanskrit texts than they are in Old Latin or Hittite, correlatives still abound, syntactically identical to the so-called "Hittite/Latin style" correlatives. Moreover, while 'plain' relatives are less common, they are still attested in archaic Latin and Hittite. In fact, as Cinque (2013: 217) notes, we do not know of any languages for which correlatives are the only type of RC. What, then, are we to make of these patterns? Again, this could be linked to the difference between restrictive and appositive structures, the former corresponding to correlatives and the latter 'plain' relatives. Cross-linguistically, this pattern is dominant enough that it led Grosu & Landman (1998) to suggest that correlative clauses could never

be appositive (see §2.2.2.1). This claim is contradicted by the evidence from Sanskrit (Hettrich, 1988); additionally, even in the Hittite corpus, where appositive RCs are scant, we find examples of appositive correlatives:[72]

(3.45) *nu* *Ḫūtupiyanzaš* ***kuiš*** DUMU LUGAL KUR ^{URU}*Palā* *maniyaḫḫieskit*
CONN Hutupiyanza.NOM.SG REL.NOM.SG son king land Pala.ACC govern.PRET.3SG
nu=šši ^{URU}*Wāsumana* *kūruriyaḫta*
CONN=himDAT.-LOC.SG Wasumana.NOM wage-war.PRET.3SG
'Hutupiyanza, **who**, (being) the king's son, was governing Pala, upon **him** Wasumana waged war' (KUB 14.29.1.17–8)

On the other hand, among the few examples of 'plain' relatives in Hittite, we find at least some that are restrictive:[73]

(3.46) *nu* 8 DUMU.MEŠ-*uš* *uwadanzi* MUNUS-*ni=ššan* ***kuiēš*** *nāwi*
CONN 8 boy.ACC.PL bring.PRES.3PL woman-towards REL.NOM.PL not-yet
pānzi
go.PRES.3PL
'They bring eight boys **who** have not yet gone to a woman' (KUB 9.31 ii 9–10)

Thus the hypothesis that correlative = restrictive and 'plain' relative = appositive does not hold in the attested languages, whether their relative pronoun is descended from **k^{w}ó-* or **yó-*. For this reason, I find it unconvincing. Again, it could well be that the use of correlatives structures is yet another feature of the prosaic genre of the earliest Hittite and Latin texts; this is a variable we simply cannot control for. And finally, even if we were to accept the correlative-as-restrictive hypothesis, this would tell us nothing of the uses of **k^{w}ó-* vs. *yó-*, since descendants of each are used to form both correlatives and 'plain' relatives in their respective languages.

To summarise: I do not believe the evidence is strong enough for us to reconstruct a "**yó*-type" RC vs. a "**k^{w}ó*-type". The propensity of descendants of **k^{w}ó-* in Old Latin and Hittite to occur in restrictive and maximalising correlatives is not, I think, a pattern inherited from PIE. For this reason, I will not ascribe these features to **k^{w}ó-*, but allows them to be a possibility for **both** **yó-* and **k^{w}ó-*. I discuss the implications for this decision in §3.4 below.

3.3.2.1 'Determinate' and 'indeterminate' relatives in Hittite

There is one branch of IE where the use of **k^{w}ó-* as a relative pronoun appears to show a syntactic pattern that is not comparable to anything we find with **yó-*. This is the Hittite (and probably Proto-Anatolian)[74] distinction between so-called **determinate** and **indeterminate** RCs. These were first discussed in Held (1957) and Garrett (1994), with renewed investigations in Samuels (2005), Huggard (2011, 2015) and Yates (2014) *inter alia*. In what we may refer to as the 'traditional' account of Hittite RCs (see, e.g., Hoffner &

[72] Example from Hahn (1949: 367).
[73] Example from Hoffner & Melchert (2008: 426), who refer to such examples as 'indefinite relatives'.
[74] See Garrett (1994: 49) and Yates (2014).

Melchert (2008: 424–5)), the position of the relative pronoun affects the semantic interpretation of the RC, under a generalisation referred to as the 'Held-Garrett Rule' (HGR, Yates, 2014). The HGR consists of two parts: (i) when the relative pronoun occurs in absolute clause-initial position, or immediately after the clitic-chain, the reading is of an 'indeterminate' RC, in which the antecedent has not been established in the discourse yet:[75]

(3.47) ***kuiš*** *ammel* *āppan* LUGAL-*uš* *kīšari* *nu* ^{URU}*Ḫattušan*
REL.NOM.SG me.GEN.SG after king.NOM.SG become.PRES.3SG CONN Ḫattuša.ACC.SG
āppa *ašāši* *n=an* *nepišaš* ^{d}IŠKUR-*aš* *ḫazziet*[(*tu*)]
again settle.PRES.3SG CONN=he.ACC.SG sky.GEN.SG storm-god.NOM.SG strike.IMPV.3SG
'**Who(ever)** becomes king after me and resettles Ḫattuša, let the Stormgod of heaven strike him down' (KBo 3.22: 49–51)

(ii) where the pronoun is preceded by one or more accented words (i.e., not including the clitic chain), the semantics are determinate, and refer to an antecedent whose existence has already been established:[76]

(3.48) KUR.KUR.ḪI.A ***kue*** *dannatta* *ammuk* EGIR-*pa* *ašešanun*
land.ACC.PL REL.ACC.PL empty.ACC.PL I.NOM.SG again settle.PRET.1SG
nu=mu=kan *apē=ya* *ḫūmanda* *arḫa* *dāš*
CONN=me.DAT.SG=PTC this.ACC.PL=and all.ACC.PL away take.PRET.3SG
'The empty lands **which** I resettled, all those he took away from me' (Ḫatt. iii 57–8)

'Indeterminate' relatives are therefore semantically equivalent to conditional clauses, as noted by Garrett (1994: 44), such that ex. (3.47) could be rephrased as '**If** anyone becomes king...'. It is worth noting that elsewhere in the archaic IE languages, indeterminacy is often expressed by the use of irrealis moods (subjunctive or optative). This strategy is not available in Hittite, given its limited verbal morphology in comparison to the 'nuclear' IE languages.

The distinction between 'determinate' and 'indeterminate' relatives, which has been the subject of much discussion among Anatolianists, has various implications for our reconstruction of PIE relativisation strategies. Let us start with an account for 'determinate' RCs, since they are in many ways more syntactically unexpected than their 'indeterminate' counterparts. Recall that in 'determinate' relatives, *kuiš* **must** be preceded by another lexically accented word; i.e., 'determinate' *kuiš* cannot occur clause-initially. The language seems to be fairly ambivalent as to what the lexically accented word preceding *kuiš* is. It could be the relativised NP, as in ex. (3.48) above; but it could also be a completely unrelated XP from within the RC. These options are schematised as follows (Huggard, 2011: 95):[77]

(3.49) (a) [[N + REL] ...
(b) [XP [REL + N]...
(c) [XP [N + REL] ...

[75] Example from Hoffner & Melchert (2008: 424).
[76] Example from Hoffner & Melchert (2008: 424).
[77] Where N = head noun.

(d) [N XP [REL] ...

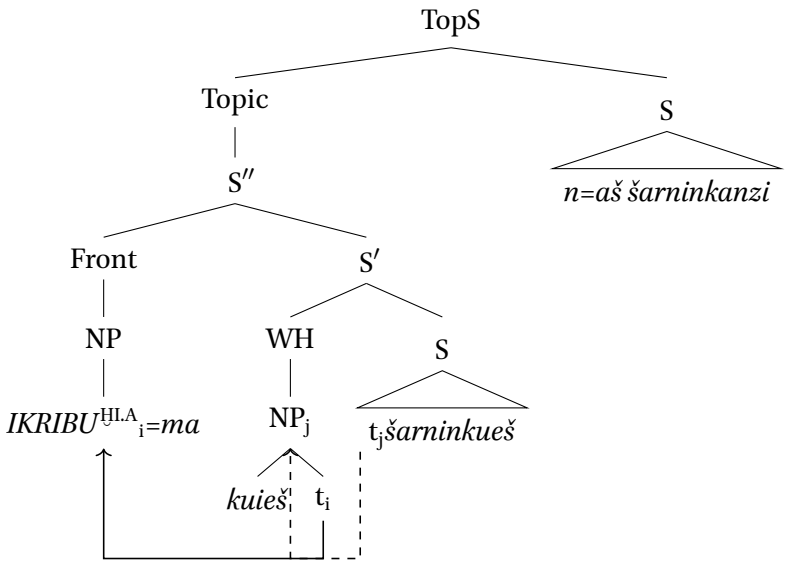
In fact, in the absence of any other material in the clause other than the relative pronoun and the verb itself, the relative pronoun follows the verb:

(3.50) *paprizzi* ***kuiš*** 3 GÌN KUBBABAR *pāi*
sully.PRES.3SG REL.NOM.SG 3 half-shekels-of-silver give.PRES.3SG
'The one **who** sullies gives three half-shekels of silver' (KBo 6.2 I 57)

Garrett (1994) accounts for this word order as follows. There are two relevant left-peripheral slots: the higher of these is 'Front' and the lower is 'WH'. First, the entire relativised DP (i.e. rel. pro. + head noun) is raised to 'WH' (i.e., *wh*-movement). Then, in the case of 'determinate' relatives, another constituent must **obligatorily** move to 'Front'. This is illustrated with the following example and tree, reproduced from Huggard (2011: 94):

(3.51) *nu* *IKRIBU*$^{\text{ḪI.A}}$*=ma* ***kuieš*** *šarninkueš* *n=aš*
CONN votive-offering.PL=but REL.NOM.PL compensatory.NOM.PL CONN=them.ACC.PL
šarninkanzi
pay-in-compensation.PRES.3PL
'But the votive offerings **which** are to be made in restitution, those they will make in restitution'
(KBo 2.2 iv 7f)

(3.52)



This 'fronting' is supposedly motivated by discourse factors, i.e, Topic and Focus. The argument is that in 'determinate' relatives, some constituent other than the relative pronoun is obligatorily focused or topicalised, raising it above the relative pronoun, yielding the attested word order. Garrett's idea stems from Hale (1987), who argues that we could account for non-initial occurrences of relative *yá-* in Vedic in this way (I return to this in §4). However, there is a crucial difference between Vedic and Hittite. In Vedic, while it is certainly possibly for a topicalised or focalised element to precede the relative pronoun, in the vast majority of tokens the relative pronoun is clause-initial. On the other hand, in Hittite 'determinate' relatives, the relative pronoun is **never** clause-initial. Given the rigidity of this rule in Hittite, we are

forced to assume that every single instance of a 'determinate' RC exhibits another constituent that has been topicalised or focalised above the relative pronoun, which is itself (allegedly) fronted. While this hypothesis can generate the attested word orders, it is poorly motivated semantically: why should it be the case that, unless something (anything!) *other than the relative pronoun* is topicalised/focalised, the antecedent is understood as 'indeterminate'? This question remains open under the double-fronting hypothesis.

An appealing alternative is put forward by Huggard (2011, 2015). Huggard's position is that in the first place, the relative pronoun in 'determinate' relatives is not fronted at all, and instead remains in its base-generated position within *v*P. It is still possible for other constituents to be fronted according to discourse factors, but these movements are optional, just as they are for simple declaratives. In this regard, Huggard builds upon the model of Garrett (1994) where topicalisation/focalisation is possible in RCs, but differs in that the relative pronoun itself is not fronted. This accounts for examples such as the following, where, rather than positing that everything preceding relative *kuin* is somehow 'fronted', only the topic ('3000 deportees of Dukkama') is fronted, while the PP ('to my house') remains adjoined to *v*P, and the relative pronoun itself remains in the base-generated object position directly preceding the verb.

(3.53) NAM.RA ^{URU}*Duggama* 3 *LIM* NAM.RA *INA*
deportees(collective).ACC.SG Dukkama.GEN.SG 3 thousand deportees.ACC.SG into
É=*YA* ***kuin*** *uwatenun* *n=an=za=an* ÉRIN.MEŠ
house=my REL.ACC.SG bring.PRET.1SG CONN=them.ACC.SG=PTC=PTC infantry
ANŠE.KUR.RA.ḪI.A *iyanun*
cavalry make.PRET.1SG
'The deportees of Dukkama, 3000 deportees, **whom** I brought into my house, I made them infantry and cavalry' (KBo 4.4 iv 22–7)

Yet this still cannot account for the fact the 'determinate' *kuiš* **never** occurs clause-initially: even if it remains in its base generated position, it should be able to precede the verb in a sentence such as ex. (3.50) above. To account for the attested patterns, Huggard (2015) argues that *kuiš*, which remains syntactically in its base-generated position, is also **enclitic**, per the phonological evidence established in Kloekhorst (2014). I will return to the implications of this clitic behaviour in §5.1.1.3.

As an interim summary, looking only at 'determinate' RCs in Hittite, we may draw the following conclusions about relative *kuiš*:

- It does not move to the left periphery, but stays in its base-generated position
- It is enclitic, and must have an appropriate prosodic host to its left

However, neither of these rules appears to be true for 'indeterminate' relatives, where *kuiš* occupies clause-initial position, implying both that it has been fronted and that it is not enclitic. For Garrett (1994), the explanation was that *kuiš* in 'indeterminate' relatives was itself fronted out of 'WH' and into the position that was reserved in 'determinate' relatives for *some other constituent* (the position he calls 'Front'). This hypothesis is more convincing than his account of 'determinate' relatives. Clearly *kuiš* is

fronted in these structures; it makes sense that this fronting is somehow implicated in the 'indeterminate' semantics. Accordingly, Huggard's (2015) account of 'indeterminate' relatives is more similar to Garrett's than his account for 'determinate' relatives. Indeed, his starting point is the observation, made by Garrett (1994: 44), that 'indeterminate' relatives are semantically equivalent to **conditional clauses**. Huggard (2015) takes this observation one step further by arguing that 'indeterminate' relatives are better termed ***wh*-conditional correlatives**. Huggard (2015: 129) provides as parallels the following from Mandarin and Serbo-Croatian respectively:

(3.54) ***shei*** *xian jinlai, wo xian da shei*
who first enters, I first hit who
'If X enters first, I hit X first'

(3.55) ***ko*** *ima magarca, taj ga i bije*
who has donkey that.one it and beats
'If anyone owns a donkey, he beats it'

At this point, it is worth noting that the grand claim of Huggard (2015) is that all instances of Hittite *kuiš*—interrogative, relative, indefinite—instantiate the same lexical item: an **indefinite polarity item**. A full appraisal of his analysis is beyond the scope of this thesis, but a crucial part of his claim is that Hittite *kuiš*, like the PIE indefinite *k^wi- from which it is supposedly derived, is prosodically deficient (i.e., enclitic). This prosodic deficiency is evidenced directly in 'determinate' relatives, as noted above. The question is: how do we account for the clause-initial position of prosodically deficient *kuiš* in 'indeterminate' relatives/*wh*-conditional correlatives? Huggard's (2015: 120) solution is to posit that in these constructions, indefinite *kuiš* is fronted to [Spec, ContrFocP][78] and so receives a pitch accent, allowing it to occur clause-initially. This theory quite similar to Roussou's (1998) account of AGk. indefinite τις vs. interrogative τίς. Roussou argues that indefinite τις receives its interrogative semantics **and** pitch accent by moving to [Spec, FocP]. In comparison, Huggard is simply arguing for a different left-peripheral position, with correspondingly different semantics, to account for indefinite *kuiš* receiving its accent and contrastive (and so conditional) semantics in Hittite.

On the whole, I find Huggard's account of Hittite 'indeterminate' relatives (*qua wh*-conditional correlatives) rather convincing. Given that the only structural difference between 'indeterminate' and 'determinate' relatives is the fronting of *kuiš*, it is parsimonious to assume that this movement is what accounts for the 'indeterminate'/conditional semantics of such clauses. There are also cogent distributional arguments for the existence of a distinct position [Spec, ContrFocP], which I return to in §4.4.3. However, Huggard's analysis of all instances (interrogative, indefinite, relative) of Hittite *kuiš* as 'indefinites whose core is a polarity item' (2015: 120) does not necessarily imply a diachronic grammatical pathway from PIE indefinite *k^wi- > Hit. "*wh*-word" *kuiš*. Rather, it is a way of conceptualising "*wh*-words"—that is, indefinite, interrogative and relative pronouns—**synchronically**. To illustrate this, let us consider Arsenijević (2009), who is cited several times in Huggard (2015), and provides some of the theoretical context for the analysis of *wh*-conditional correlatives in Hittite. Arsenijević (2009) proposes an account

[78] [Spec, CntrFocP] in Huggard's notation.

of *wh*-conditionals in which he argues that 'the minimal semantic content shared by all *wh*-expressions is the weak_∃ quantification, i.e. the non-specific interpretation with weak existential commitment' (Arsenijević, 2009: 137).[79] In other words, "*wh*-words" by virtue of their grammatical function, are argued to share an underlying **semantic** property (weak existential quantification). Now, in a language where there is some degree of morphological identity between all three types of *wh*-word (e.g., Hittite or Latin) positing shared features of this sort may be more intuitive than in one where one of the three forms is distinct (e.g., Sanskrit or Ancient Greek). Yet I do not see any compelling reason to interpret this shared semantic content as evidence for a shared **etymological** origin. From a synchronic perspective too, I believe the possibility for some shared semantic content between indefinite and relative pronouns does not necessarily imply they are exactly the same lexical item; rather, the possibility of a shared feature is compatible with the null hypothesis that *kuiš* is a relative pronoun, and it is precisely its function as such that allows it to form *wh*-conditionals. I am therefore content to accept a definition of the relative pronoun which includes 'weak existential quantification' (per Arsenijević, 2009), an attribute shared by indefinite pronouns—but this does not imply that one use developed from the other, nor that the two are underlyingly the same lexical item.

Huggard (2015: 145) also hints at a diachronic analysis in claiming that '[w]ithin the archaic IE languages, early Latin and Hittite [i.e., those with a relative pronoun from **k^{w}ó-*], correlatives do show a conditional sense, but the correlatives in ancient Greek and the Indo-Iranian languages [i.e., those with a relative pronoun from **yó-*] do not'. The argument would follow that the conditional function of Hittite and Latin correlatives is inherited from the status of PIE **k^{w}ó-* as an indefinite polarity item, while its absence in languages with relatives derived from **yó-* is due to the fact **yó-* was originally a demonstrative,[80] and not a polarity item. However, Huggard's claim is false: Hock (1989: 95–6) lists several examples of Sanskrit RCs that are better translated as conditionals, noting in particular that 'Sanskrit generalizing relatives frequently mark the verb of the RC in the optative' [i.e., the mood used for the protasis of general conditions] (1989: 101, n.12). The same argument can be made for Ancient Greek: Probert (2015: 95) argues that '[t]he indefinite construction [viz. relative ὅς (± modal particle ἄν) + irrealis mood] is used in the equivalent of a conditional protasis when the generalization applies across occasions of a relevant kind.' To me, this suggests that the underlying grammatical feature that contributes a conditional reading in relative pronouns, however we choose to theorise it, has no implications for the morphological history of the relative pronoun. Rather, 'weak existential quantification' is simply a feature shared by indefinite and relative pronouns. As such, while I accept Huggard's (2015) analysis of Hittite 'indeterminate' relatives as *wh*-conditionals, I would hold that *kuiš* is still meaningfully a relative pronoun in these cases, and as such *wh*-conditionals, as much as they are a form of conditional clause, are also a form of relative clause.

Setting aside 'indeterminate' relatives, Huggard (2015) goes one step further, extending his hypothesis that *kuiš* is not a relative pronoun but an indefinite to 'determinate' relatives as well, which he

[79] 'Weak' in this context means simply 'non-specific', such that weak_∃x = "there exists some x" (as opposed to a specific x to be predicated over). See further Ladusaw (1994).

[80] Cf. the claims discussed in §3.2.2.1. Huggard (2015: 5) claims that 'the oldest reconstructible function of [**yó-*] was not as a relative pronoun', but the author remains agnostic as to what this older function was.

speculatively rechristens 'existential correlatives', as a counterpart to *wh*-conditional correlatives (2015: 152–69). Huggard (2015: 154) expresses concern that Hittite 'determinate' relatives do not show all the characterisations of correlative RCs established by de Vries (2002): namely (i) they do not exhibit *wh*-movement; (ii) they adjoin to CP, not IP; (iii) they are not always semantically maximalising and so can "stack". On this basis, he sees it prudent to consider a hypothesis in which these are not RCs at all, hence 'existential correlatives'. Yet when assessing de Vries' characteristics of correlative relatives, he rightly suggests that these generalisations may hold 'only for correlatives in certain languages such as Hindi', on which much of the generative literature on correlatives has focused (see §2.2.2.2). He even cites Davison (2009) who argues that Sanskrit correlatives also adjoin to CP (like Hittite), not IP (like Hindi); a point that was already argued in pre-Minimalist terms by Hock (1989) and Kiparsky (1995). Sanskrit correlatives therefore share property (ii) with Hittite; in fact, Sanskrit also shares property (iii), insofar as appositive correlatives are possible and even common (Hettrich, 1988).[81] Since Huggard explicitly limits his scope to Hittite, he does not offer further comment on this matter. Yet looking at the comparative situation, the only criterion for correlative-hood that Hittite 'determinate' correlatives fail to meet, which other ancient IE correlatives satisfy, is that of *wh*-movement. Aside from the fact that *wh*-movement appears to be optional elsewhere in ancient IE languages (see §5.1.2), on its own the lack of *wh*-movement is not sufficient to force a re-interpretation of *kuiš* as anything other than a relative pronoun: *wh*-movement varies parametrically. In fact, de Vries (2002: 173) mentions four languages (albeit closely related to each other), that seem to parallel the situation in Hittite:

> There are relative pronouns in interrogative format, e.g. *mìn* in Bambara. The predominant relative strategy is correlative...the relative pronoun and head noun are *in situ*.

Admittedly, the Hittite head noun need not be *in situ*, and can be topicalised/focalised according to the same rules as in declaratives. Nevertheless, I maintain that the only thing that is perhaps unexpected about Hittite 'determinate' relatives is the lack of *wh*-movement. Yet even this is not particularly troubling since, as Huggard (2011, 2015) demonstrates, Hittite does not seem to have *wh*-movement at all (including for interrogatives, on which see §3.3.3.1 below). While in synchronic terms I see the elegance of reducing every instance to *kuiš* to a single lexical item, I do not think there are sufficient grounds to abandon the null hypothesis that *kuiš* is meaningfully a relative pronoun in Hittite.

As such, I interpret the Hittite data concerning *kuiš*, taken as broadly representative of the archaic Anatolian situation, as follows:

- Hittite 'indeterminate' relatives are best understood as *wh*-conditional correlatives, which also occur elsewhere in IE
  - 'indeterminate' *kuiš* fronted to [Spec, ContrFocP], yielding its conditional semantics
  - 'indeterminate' *kuiš* functions as a relative pronoun
- Hittite 'determinate' relatives are functionally equivalent to conventional RCs elsewhere in IE
  - 'determinate' *kuiš* does not undergo *wh*-movement

[81] These are issues I return to for PIE in §5.

– 'determinate' *kuiš* is enclitic
– 'determinate' *kuiš* functions as a relative pronoun

This situation differs in several important ways from the other patterns of relative **k^{w}ó-* discussed so far. I will account for the various behaviours of *kuiš* in §5.1.1.3, including its apparently enclitic behaviour in 'determinate' relatives and its unique left peripheral position in [Spec, ContrFocP].

3.3.3 Non-relative uses of **k^{w}ó-*

As I mentioned above, **k^{w}ó-* is preserved as an interrogative and indefinite pronoun in every branch of IE.

3.3.3.1 Interrogative

**k^{w}ó-* as an interrogative tends to appear initially or near-initially across the IE language family. This is the regular position in almost all branches of IE, including Indo-Iranian, Italic, Greek, Celtic and Germanic. These languages are generally held to have *wh*-movement in interrogative clauses, that is obligatory movement of the interrogative pronoun (regardless of case) to the left periphery:

(3.56) ***káḥ*** *vā yajñáiḥ* *pári* *dákṣam* *te*
INT.NOM.SG.M or sacrifice.INS.SG.M around strength.ACC.SG.M you.GEN.SG
āpa, ***kéna*** *vā te* *mánasā* *dāśema*
obtain.PERF.3SG INT.INS.SG.N or you.ACC.SG mind.INS.SG.N strengthen.OPT.1PL
'**Who** has wholly obtained your strength by sacrifice? Or by **what** thought should we strengthen you?' (RV 1.76.1cd)

(3.57) ***quid*** *ais?* ***quid*** *nomen* *tibi* *est?*
INT.ACC.SG.N say.PRES.2SG INT.NOM.SG.N name.NOM.SG.N you.DAT.SG be.PRES.3SG
'**What** do you say? **What** is your name?' (Plaut. *Amph.* 364)

(3.58) ἔνθα τίνα πρῶτον, τίνα δ' ὕστατον ἐξενάριξαν /
then INT.ACC.SG.M first.ACC.SG.M INT.ACC.SG.M PTC last.ACC.SG.M slay.AOR.3PL
Ἕκτωρ...;
Hector.NOM.SG.M
'Then **whom** first, **whom** last, did Hector [and Ares] slay?' (Hom. *Il.* 5.703–4)

(3.59) ***cid*** *asmaith* *disunt* *tra*
INT.ACC.SG.N be.PRES.3SG.REL/good.NOM.SG.N from-here then
'**What** then is good therefrom?' (Wb 12d12)[82]

(3.60) ***ƕana*** *wileiþ* *ei* *fraletau* *izwis*?
INT.ACC.SG want.PRES.2PL that release.OPT.1SG you.ACC.SG
'**Whom** do you want me to release to you?' (Matthew 27:17)

[82] Example from García-Castillero (2020: 192).

Given our model of the left periphery (see §2.3), we assume these move to [Spec, FocP]. Theoretically, therefore, they can be preceded by a fronted topic in the higher [Spec, TopP]. And indeed, this word order is also attested:

(3.61) *índraḥ* ***kím*** *asya* *sakhyé* *cakāra*
indra.NOM.SG INT.NOM.SG.N it.GEN.SG friendship.LOC.SG do.PERF.3SG
'**What** did Indra do in friendship of it?' (RV 6.27.1b)

Such examples are discussed in detail in §4. Finally, it is also possible in some of these languages—at least Latin and Ancient Greek—to have multiple fronted interrogatives in the same clause:

(3.62) **τίς** **τίνος** αἴτιός ἐστι γενήσεται φανερόν
INT.NOM.SG.M INT.GEN.SG.N guilty.NOM.SG.M be.PRES.3SG become.FUT.3SG clear.NOM.SG.N
'It will become clear **who** is responsible for **what**' (Dem. 18.73)

(3.63) *considera,* *C. Piso,* ***quis*** ***quem*** *fraudasse*
consider.IMPV.2SG Gaius-Piso.VOC.SG.M INT.NOM.SG.M INT.ACC.SG.M cheat.PERF.INF
dicatur
be-said.PRES.SUBJ.3SG
'Consider, Piso, **who** is said to have cheated **whom**? (Cic. *QRosc.* 21)

The case is slightly more complicated for Hittite. Although several previous authors (Hale, 1987; Garrett, 1994; Hoffner & Melchert, 2008) have argued that Hittite also has *wh*-movement, it has been clear since at least Hoffner (1995) that the situation is not so straightforward. More recently, Goedegebuure (2009) established that almost all Hittite sentences with an initial interrogative pronoun fall into one (or both) of two categories:

1. The interrogative pronoun is the subject of the verb, as in ex. (3.64) below
2. The only constituent other than the interrogative pronoun is the verb, as in ex. (3.65) below

(3.64) ***kuiš****=war=an* *ḫaran* ^{d}*Pirwar*[*i*] ^{URU}*Ḫaššuwaza* *uwatez*[*zi*]
INT.NOM.SG=QUOT=him.ACC.SG eagle.ACC.SG Pirwa.DAT.SG Hassu.ABL.SG bring.PRES.3SG
'**Who** will bring the eagle from the city of Hassu to Pirwa?'
(KUB 48.99 obv 6–7)

(3.65) ***kuit****=ta* *memaḫḫi*
INT.ACC.SG=you.DAT.SG say.PRES.1SG
'**What** do I say to you?' (KBo 26.65 iv 23, 25)

Under such circumstances, we cannot say for certain whether the pronoun has been fronted to the left periphery or not. Where these conditions do not obtain, interrogative *kuiš* is not found initially. Most commonly, if the relevant form is the object of the clause it is found pre-verbally (i.e., in its base-generated position), while oblique forms appear 'distributed freely across the clause', as is the case in declaratives too (Goedegebuure, 2009: 947–8). On this basis, Goedegebuure (2009) argues that the position of interrogative pronouns matches that of DPs in declarative sentences. These patterns, when

taken together, suggest that Hittite is in fact a *wh-in-situ* language. The point is further elaborated upon by Huggard (2011, 2015), who draws a link between the lack of fronting in interrogatives and determinate relatives discussed in §3.3.2.1 above.

Since it is neither the focus nor the topic of this thesis, I will not offer any conclusions as to whether PIE interrogatives had *wh*-movement or not. Typologically speaking, *wh*-movement can be either gained or lost: additionally, it is possible for languages to lose *wh*-movement (including multiple fronting) in interrogatives but keep it in RCs, or for it to be optional under certain pragmatic conditions, as is the case in, e.g., Hindi. For this reason, although further studies should seek to address the question of interrogative placement, it does not bear directly on the reconstruction of RCs.

3.3.3.2 Indefinite

The third use of **k^{w}o-* that can be reconstructed is as an indefinite pronoun. When a form of **k^{w}o-* is used alone as an indefinite pronoun it is consistently **enclitic**, meaning that it is underlying unaccented, and cannot appear clause-initially. Like interrogative **k^{w}ó-*, indefinite **k^{w}o-* is found in every branch of IE, and is thus very likely an inherited use. It is attested in its bare form most clearly in Ancient Greek (τις) and Sabellic (e.g., Oscan **pís**), as well as some grammatical environments in OCS (Willis, 2013: 376); relics survive elsewhere, usually following a conjunction (exemplified below with Latin and Gothic) or negation (exemplified with Sanskrit).

(3.66) ὣδε δέ **τις** εἴπεσκεν ἰδὼν ἐς πλησίον ἄλλον
thus PTC INDEF.NOM.SG.M say.IMPF.3SG see.AOR.PPL.NOM.SG.M to near.ACC.SG.M other.ACC.SG.M
'Thus **someone** would say, turning to his neighbour' (Hom. *Il.* 2.271)

(3.67) *si-**quis** me quaeret, inde vocatote aliqui*
if-INDEF.NOM.SG.M/F me.ACC.SG seek.FUT.3SG from-there call.FUT.IMPV.2PL INDEF.NOM.PL.M/F
'If **anyone** asks for me, some of you must call me from there' (Plaut. *Stich.* 67)

(3.68) *jabai **ƕas** ni frijoþ fraujan Iesu Xristu, sijai anaþaima*
if INDEF.NOM.SG NEG believe.PRES.3SG lord.ACC.SG Jesus Christ, be.OPT.3SG anathema
'If **anyone** does not believe in the Lord Jesus Christ, let them be cursed' (1 Corinthians 16:22)

(3.69) *mā́=**kis** tokásya naḥ riṣat*
NEG=INDEF.NOM.SG offspring.GEN.SG our be-injured.INJ.3SG
'Let not **one** of our offspring be injured' (RV 8.67.11c)

Across IE, the regular paradigms of the indefinite pronoun are usually built on **k^{w}o-* but with diverse morphological extensions (Lat. *ali-*/*-que*/*-piam* etc., Hit. *-kki*/*-kka*, Skt. *cid*/*ca*/*caná* etc., Got. **-uh*). Only in these extended forms are indefinite pronouns attested as non-clitic; it is thus widely held that

PIE indefinite *k^wo- was enclitic. Indeed, even when there is morphological extension, the indefinite pronoun may still be enclitic, as is the case for Hittite *kuišk*i (Hahn, 1946; Watkins, 2010; Kloekhorst, 2014; Huggard, 2015; Sideltsev & Molina, 2015).[83]

There is some debate, however, as to whether the indefinite pronoun followed Wackernagel's Law (i.e., whether it was routinely moved to "second position", broadly conceived), or whether it remained within its base-generated phrase—I address this question in §4.4. For now, it is sufficient to note that by virtue of their being enclitic, bare forms of indefinite *k^wo- never occur clause-initially; this is in sharp contrast to the use of *$k^wó$- as a relative or interrogative, where this position is very common.

A pertinent issue concerning indefinite *k^wo- is its implication in the 'reduplicated' relativisers found in Lat. *quisquis* ('whoever') and Hit. *kuiš kuiš* ('id.'). Several different accounts have been posited for the origins of these forms, including various combinations of *$k^wó$-/*k^wo- in its interrogative, relative and indefinite meanings (Hahn, 1937: 397–9). The meaning that both Hit. and Lat. forms share is a generalising capacity, somewhat equivalent to the use of English *-ever*:[84]

(3.70) ***kuit*** ***kuit*** *ḫarakzi* *t=at* *šarnikzi*
REL.ACC.SG.N "-ever" destroy.PRES.MID.3SG CONN=it.ACC.SG.N pay-for.PRES.3SG
'**Whatever** is destroyed, this he pays for' (Laws §127)

(3.71) *video* *necesse* *esse* *eloqui* ***quidquid*** *roges*
see.PRES.1SG necessary be.INF say.INF REL.ACC.SG.N ask.PRES.SUBJ.2SG
'I see that I must tell you **whatever** you ask' (Plaut. *Asin*. 24)

Both parts of Hit. *kuiš kuiš* and Lat. *quisquis* inflect; the main difference between the two is that Lat. *quisquis* appears to be entirely univerbated, whereas Hit. *kuiš kuiš* can be interrupted by other constituents (especially the emphatic particle *imma*) and the two parts are 'felt to be two separate words' (Hahn, 1937: 388).

In both Hittite and Latin, *kuiš kuiš/quisquis* are often taken to be strictly relative; in both languages, however, there appear to be some instances whether the meaning is not relative, but are used as **free-choice indefinites**, generally translatable as 'any'. In the case of Hittite, the use of *kuiš kuiš* as a free-choice indefinite appears to be an innovation that took place during the language's attested history. As noted by Melchert (1985) and reported in Sideltsev (2018), there are no uses of *kuiš kuiš* as a free-choice indefinite in the Old Hittite corpus. Morever, when *kuiš kuiš* is attested as a free-choice indefinite, it is only found with the inclusion of the emphatic particle *imma*, an addition which is optional with generalising relatives:[85]

(3.72) ***kuit*** ***imma*** ***kuit*** *kukupalatar* GAM *NIŠ* DINGIR–*LÌ*
?INDEF.NOM.SG.N EMPH ?INDEF.NOM.SG.N kukupalatar down oath god
GAR-*ru*
lie.IMPV.3SG
'Let **any and every** *kukupalatar* be placed under oath' (KUB 26.1+ rev. iv 52–53)

[83]Not all of these authors use the specific term 'enclitic' to describe the status of Hit. *kuišk*i, but discuss phenomena that point towards its being characterised as such.

[84]Hittite example from Hahn (1937: 393).

[85]Example from Sideltsev (2018: 286).

It seems likely, therefore, that Sideltsev (2018) is right in arguing that indefinite use was a development of the relative use, a change for which there are cogent typological parallels (discussed below). The case for Latin is not as clear, since *quisquis* is also used as a free-choice indefinite quite frequently in the early Latin authors (Hahn, 1933: 32; Lodge, 1962: 519):

(3.73) *ego* *istos* *novi* *polypos* *qui* *ubi*
I DEM.ACC.PL.M know.PERF.3SG octopus.ACC.PL.M REL.NOM.SG.M when
quicquid *tetigerunt* *tenent*
?INDEF.ACC.SG.N touch.PERF.3PL hold.PRES.3PL
'I know those tentacled types who, when they touch **anything**, they keep (it)' (Plaut. *Aul.* 198)

In this way, its use overlaps almost entirely with *quisque*—the standard free-choice indefinite in Classical Latin—which is itself attested as a generalising relative in early authors (Hahn, 1933: 32; Lodge, 1962: 517):

(3.74) ***quisque*** *obviam* *huic* *occesserit* *irato,* *vapulabit*
?REL.NOM.SG.M in-the-way DEM.DAT.SG.M go-to.FUT.PERF.3SG angry.DAT.SG.M flog.FUT.3SG
'**Whoever** gets in his way when he's angry, he'll flog (them)' (Plaut. *Asin.* 404)

There is no direct evidence for a relative pronoun of the form *quis* in Latin. However, we see in the attested history of the language a development of innovative free-choice indefinites built on pronouns that are unambiguously relative in origin. These include forms such as *quivis* and *quilibet* ('anyone'). These are transparently derived from the relative *qui* + *vis* ('you want') and *libet* ('it pleases') respectively; as such, the ancestor of a construction such as *quilibet homo* ('any man') would have been something like 'which(ever) man **pleases** (you)'. In its genesis, then, *qui libet* is not a free-choice indefinite but a generalising RC. This explains why the univerbated form *quilibet* cannot itself introduce an RC: the relatival force of *qui* has been satisfied by *libet*. The grammaticalisation of such an RC, most often with verbs of wanting, is cross-linguistically common and fairly uncontroversial (Haspelmath, 1997: 133–4).

On the other hand, the forms *quisquis* and *quisque* do not contain any verbal material, and so can (and do) introduce RCs. Yet it is possible that their use as free-choice indefinites develops in the same way as *quilibet* and *quivis*: that is, in both cases, we see the development from generalising relative > free choice indefinite. The difference is that in *quilibet* and *quivis*, the generalising force was contributed by the verbs of 'wanting'; in *quisquis* and *quisque*, the generalising force was inherent to the pronominal forms; as such, there was no need for any particular lexical verb to be part of the construction. And since no verb was grammaticalised as part of the construction, *quisquis* and *quisque* retained their ability to introduce RCs, at least until the pre-Classical period in the case of *quisque*. If this hypothesis holds, *quisque* and *quisquis* would instantiate a fossilised form of a relative **quis* (<**k*^{w}*í-s*), present at an earlier, unattested stage of Latin but comprehensively supplanted by *qui* (<**k*^{w}*ó-i*) before our earliest texts.

In sum, we have reason to believe that both Hittite *kuiš kuiš* and Latin *quisquis* had the original function of a generalising relative pronoun. As for how these forms came into being, I believe there are two plausible hypotheses. The first is that they are straightforwardly a case of iterative reduplication,

probably of the relative pronoun. There is a structural parallel here with Sanskrit *yá- ya-*, which has a generalising force:

(3.75) ***yád=yad*** *áichat* *prajā́patau* *tád*
REL.ACC.SG.N=REL.ACC.SG.N want.IMPF.3SG Prajapati.LOC.SG DEM.ACC.SG.N
brahmacārí́ *prā́yachat*
brahmacarin.NOM.SG. deliver.IMPF.3SG
'**Whatever** he desired from Prajapati, the Brahmacharin delivered it' (AV 11.5.15)

However, despite its preponderance in later texts, there are only four occurrences of reduplicated *yá-* in the Rig Veda. Two of these are clearly adverbial, involving the forms *yátra-yatra* ('whithersoever', RV 6.75.6) and *yadā́-yadā* ('howsoever', 4.54.5); this leaves only two instances of a reduplicated relative pronoun.[86] By way of comparison, Grassmann (1873: 1066–7) lists seven tokens of generalising relatives ('whoever') in *yá- cit* and 17 tokens of *yá- ka- ca*, none of which is adverbial. Perhaps unexpectedly, the Rig Veda even contains more instances of a reduplicated **demonstrative** in correlative clauses than of the reduplicated relative.

(3.76) ***yáḥ*** *naḥ* *pṛtanyā́d* *ápa* ***tám=tam*** *ít*
REL.NOM.SG.M us.ACC.PL fight.OPT.3SG away DEM.ACC.SG.M=DEM.ACC.SG.M EMPH
dhatam
put.IMPV.2DU
'**Whoever** would give us battle, strike **him** down' (RV 1.132.6b)

De Vaan (2015: 44) cites six occurrences of this construction, only one of which is adverbial. There is some debate as to whether the reduplicated demonstrative is specifying ('exactly that one'), or generalising, as I have translated it in the example above. Hettrich (1988: 557–8) prefers the generalising value, while Klein (2003: 784–6) prefers the specifying value, ascribing it to *ít* ('only', which I have glossed as EMPH). De Vaan (2015) takes the middle ground, arguing that without *ít*, reduplicated *sá-/tá-* is generalising, while the combination with *ít* makes it specifying. I do not propose to resolve the issue here, but draw attention to de Vaan's conclusion that this use of reduplicated *sá-/tá-* could plausibly be understood as a Sanskrit-internal innovation; it is certainly possible that the same is true for reduplicated *yá-*.

Setting aside reduplication, then, we may alternatively understand these forms as a combination of relative **k^wó-* + indefinite *k^wo-*. This would also account readily for their relative force, but has more promising structural parallels such as Skt. *yá- cit*, *yá- ká- ca* 'who(ever)', and AGk. ὅστις 'id.'. In such a construction, indefinite **k^wo-* does not have the meaning 'someone', 'something', but acts in a more general 'domain-widening' capacity.[87] This concept is applied to AGk. ὅστις by Probert (2015), but is also a fair characterisation of the 'reduplicated' forms of Latin and Hittite. The central idea is that the function of the 'domain-widener' (e.g. Eng. *-ever*) is to extend the set evoked by the RC to include a number of entities 'larger than one might expect', or containing 'rather unexpected members' (Probert, 2015: 99). Probert illustrates this with the following Homeric example (2015: 100):

[86] RV 8.61.6; 2.25.1.

[87] See Jacobson (1995: 479–81), Grosu (2002: 148), Rawlins (2008: 211–16), as cited in Probert (2015: 99).

(3.77) καὶ τὸν Ἀχιλλεὺς θῆκεν ἀέθλιον οὗ
also DEM.ACC.SG.M Achilles.NOM.SG.M place.AOR.3SG prize.ACC.SG.N his.GEN.SG.M
ἑτάροιο / ὅστις ἐλαφρότατος ποσσὶ κραιπνοῖσι
companion.GEN.SG.M REL.NOM.SG.M swiftest.NOM.SG.M foot.DAT.PL.M fast.DAT.PL.M
πέλοιτο
move.PRES.OPT.3SG
'It too Achilles set down as a prize in honour of his companion, (for) **whoever** should be swiftest with his nimble feet' (Hom. *Il.* 23.748–9)

Probert remarks that 'ὅς would have picked out the same individual, but ὅστις...suggests that there is a rather large number of theoretically possible winners.' The same is true of Lat. *quisquis* and Hit. *kuiš kuiš*,[88] as well as Skt. *yá- ká- ca*:

(3.78) *numquam edepol me inultus istic ludificabit,*
never by-Pollux me.ACC.SG unpunished.NOM.SG.M DEM.NOM.SG.M tease.FUT.3SG
quisquis *est*
REL.NOM.SG.M be.PRES.3SG
'By Pollux, that bloke will never make a joke out of me unpunished, **whoever** he is'
(Plaut. *Amph.* 1041)

(3.79) *nu ḫantezziyaš* LÚ-*aš* ***kuit kuit*** *p*[*ešta*] *ta=šše*
CONN first.NOM.SG man.NOM.SG **whatever**.ACC.SG give.PRET.3SG CONN=him.DAT./LOC.SG
šarnikzi
make-restitution.PRES.3SG
'**Whatever** the first man paid, he shall give in restitution' (Laws §28a)

(3.80) *idám āpaḥ prá vahata* ***yát kím ca*** *duritám*
DEM.ACC.SG water.VOC.PL forth carry.IMPV.2PL **whatever**.NOM.SG difficulty.NOM.SG
máyi
me.LOC.SG
'Take this away, oh waters, **whatever** difficulty is within me' (RV 1.23.22ab)

From these data, it seems reasonable to hypothesise that **k*^{w}*o-*, when combined with a relative **k*^{w}*ó-* **or** **yó-*, had the abstract role of a 'domain-widener' in PIE rather than a purely indefinite sense.

To summarise: I will hold that PIE had an enclitic pronoun **k*^{w}*o-*. In its bare form, it functioned as an indefinite pronoun. In combination with a relative pronoun, it had the function of a 'domain-widener'. Its descendants in the attested languages were usually reinforced with morphological material from various sources; these expanded forms facilitated the development of non-clitic indefinite pronouns.

3.4 Which "which" is which?

In light of the above, I turn to the question that has long perplexed Indo-Europeanists: what did PIE use as a relative pronoun? The opinions of the authors I have mentioned so far fall into three broad categories, categorised by Luján (2009: 222):[89]

[88] Examples from Hoffner & Melchert (2008: 151).

[89] I exclude here authors such as W. P. Lehmann (1974) who argue that PIE had no relative pronouns at all.

(A) The oldest relative construction used **k*^{w}*ó-* as the relative pronoun (Sturtevant, 1930; Hahn, 1946, 1949, 1964; Szemerényi, 1980).

(B) The oldest relative construction used **yó-* as the relative pronoun (Delbrück, 1893–1900; Hirt, 1937; Gonda, 1954, 1955).

(C) Both possibilities were available in PIE, either for different semantic types (Haudry, 1973, 1979; Hettrich, 1988; Hajnal, 1997), or as dialectal variants (Schmitt-Brandt, 1973; Lehmann, 1980)

My hypotheses, summarised below, do not fall neatly into any of these three categories:

- On **yó-*:
  - **yó-* was **probably** used as a relative pronoun in PIE
  - Other pronominal uses of **yó-* (as a thematised form of **(e)y-*) cannot be reconstructed
  - Relative **yó-* cannot be securely connected with a putative sentential connective **yo*
- On **k*^{w}*ó-*:
  - **k*^{w}*ó-* was **possibly** used as a relative pronoun in PIE
  - **k*^{w}*ó-* was **certainly** an interrogative pronoun in PIE
  - **k*^{w}*o-* was **certainly** an indefinite pronoun in PIE, and **probably** acted as a 'domain-widener'
- On both:
  - Neither **yó-* or **k*^{w}*ó-* can be firmly associated with a particular semantic or syntactic type of RC (i.e. restrictive vs. appositive, 'plain' vs. correlative)
  - The use of **yó-* or **k*^{w}*ó-* as a complementiser is secondary to their use as relative pronouns

My judgement that **yó-* was probably a relative pronoun in PIE is based on the fact that we cannot securely reconstruct any use for it other than that of a relative pronoun. Any purported anaphoric used is based on conjecture beyond the attested languages. It is plausible that at an earlier stage—what we may call "pre-PIE"—**yó-* was an anaphoric pronoun, which was later recruited as a relative pronoun. However, there is simply no comparative evidence that speaks in favour of this reconstruction: it is, rather, internal reconstruction within PIE based on typological patterns. This is not the methodology of this thesis, and so I will not explore this possibility.

The case for **k*^{w}*ó-* is more complex. In the first place, we can be confident of at least two non-relative uses of **k*^{w}*ó-* in PIE: indefinite, and interrogative. It is therefore plausible that its use as a relative pronoun developed out of either of these uses. Nevertheless, we encounter the same difficulty as above: we are largely reduced to PIE-internal reconstructions based on typological patterns observed elsewhere. A possible exception to this is the evidence from Anatolian. The peculiarities of 'determinate' vs. 'indeterminate' relatives, together with the possible synchronic analysis of Hittite *wh*-words as a unitary lexical item, open up the possibility that **k*^{w}*ó-* was not a relative pronoun in PIE. This argument may find some support in the status of Hit. *kuiš* in 'determinate' relatives as an enclitic, much like PIE indefinite **k*^{w}*o-*. However, I have argued that the synchronic analysis of Hittite *wh*-words as indefinite polarity items does not imply a diachronic pathway from indefinite to relative. Moreover, the necessary semantic content of

'weak existential quantification' required to derive *wh*-conditionals appears to be shared by descendants of **k^{w}ó-* and **yó-* alike. As for the status of Hit. *kuiš* as enclitic, this is an issue I address in §5.1: the upshot is that since there are also enclitic reflexes of relative **yó-*, we need not hypothesise this behaviour is inherited from PIE indefinite **k^{w}o-*, and likely developed in tandem with other grammatical changes. Thus there is still no clear comparative evidence that allows us to exclude the possibility that **k^{w}ó-* was a relative pronoun in PIE.

So, what are the implications for a reconstruction of RC syntax in PIE? I have argued that neither **yó-* nor *k^{w}ó-* can be readily discounted as a relative pronoun, and, crucially, that we cannot reconstruct with confidence any difference in their syntactic behaviour when used as such. What this means is that—as far as the **syntax** is concerned—we can reconstruct a single functional category in Minimalist terms: *REL. The phonetic content of *REL, whether **yó-* or **k^{w}ó-*, simply cannot be ascertained. We may imagine various scenarios: perhaps one was the "original", the other supplanting it in the relevant daughter languages. Perhaps both co-existed within the same grammar, in free variation or tied to sociolinguistic factors. Perhaps the difference was one of register; perhaps the difference was one of dialects. The plausibility of these scenarios, however, is not something that can be measured by Minimalist reconstruction, since the grammars of the attested languages, whether **k^{w}ó-* or **yó-*based, point towards a relative pronoun with a fairly coherent set of syntactic features, which I reconstruct in §5.

In sum, if we are to pursue the aetiology of relative **yó-* vs. relative **k^{w}ó-*, we move beyond what can reasonably be called PIE—understood as the shared ancestor of the attested IE languages at the latest stage of genetic unity—towards something like "pre-PIE" (the ancestor of the ancestor of the attested IE languages). That is not to say such questions are altogether unanswerable, but they lie outside the scope of this thesis. As such, I will proceed with the reconstructed functional category *REL in PIE, whose syntax can be reconstructed by comparison of **both** **k^{w}ó-* and **yó-* languages.

Chapter 4

Clisis and the Left Periphery in Proto-Indo-European

In this chapter, I analyse the left periphery of PIE with specific reference to the interaction between pragmatic fronting (topicalisation and focalisation) and clitic placement ("Wackernagel's Law"). This constitutes a mapping out of the CP-layer in PIE, which forms a crucial part of relative clause structure, and lays the groundwork for analysing the precise syntactic behaviour of the relative pronoun, *REL.

4.1 Onwards and upwards: The "front" of the clause

In §3, I established that we should reconstruct a single lexical item (*REL) as the relative pronoun for PIE. A core feature of the descendants of *REL is that, with the probable exception of 'determinate' relatives in Anatolian, they are consistently found near the front of the sentence. Under the theoretical assumptions established in §2, we understand this 'fronting' as *REL moving to the left periphery: specifically, [Spec, TopP]. But we know that this can vary cross-linguistically; as such, a thorough analysis of the ancient IE data within this framework is necessary for the purposes of reconstruction.

It is fair to say that the left periphery of ancient IE languages is quite active.[90] Ancient IE languages seem to allow for a position[91] at the beginning of the clause[92] which can be filled by an element that is either **topicalised** (exx. 4.1–2) or **focalised** (exx. 4.3–4), according to the interpretive characteristics discussed in §2.3.

(4.1)

μητέρι$_{[Topic]}$	μοι	μνηστῆρες	ἐπέχραον	οὐκ	ἐθελούσῃ
mother.DAT.SG.F	me.DAT.SG	suitor.NOM.PL.M	assault.IMPF.3PL	NEG	willing.DAT.SG.F

'**As for my mother**, the suitors beset her against her will' (Hom. *Od.* 2.50)

(4.2)

nu	É	^{d}***IŠTAR***$_{[Topic]}$	^{m}*Dudḫaliyaš*	DUMU=*YA*	*tapardu*
CONN	house	Ištar	Tudḫaliya.NOM.SG	son=my	govern.IMPV.3SG

[90] Authors who have written on this topic from a comparative perspective include Hale (1987); Kiparsky (1995); Krisch (1990, 2017); Keydana (2011, 2018).

[91] While authors such as Keydana (2011) and Spevak (2010) have argued Topics and Foci occupy the same projection, according to the theoretical framework adopted in Rizzi (1997) I treat them as occupying distinct projections (TopP and FocP). This is further motivated when we understand interrogatives as occupying [Spec, FocP], since they co-occur with topics.

[92] In Hittite this position immediately follows the initial clitic chain introduced by a sentence connective such as *nu*.

'**As for the house of Ištar**, let Tudḫaliya, my son, govern it' (KUB 1.1 iv 77–8)

(4.3) ***nullam***$_{[\text{Focus}]}$ *ego me vidisse credo magis anum*
none.ACC.SG.F I me.ACC.SG see.PERF.INF believe.PRES.1SG more old-woman.ACC.SG.F
excruciabilem quam illaec est
tormentable.ACC.SG.F than DEM.NOM.SG.F be.PRES.3SG
'I believe I have seen **no** old woman more deserving of torment than she is' (Plaut. *Cist.* 653–4)

(4.4) ***sárvam***$_{[\text{Focus}]}$ *sá pūtám aśnāti svaditám*
all.ACC.SG.M DEM.NOM.SG.M pure.ACC.SG.M drink.PRES.3SG sweetened.ACC.SG.M
mātaríśvanā
Mātariśvan.INS.SG.F
'He drinks **all** the pure [essence], sweetened as it is by Mātariśvan' (RV 9.67.31cd)

As established in §3, both relative and interrogative pronouns (again, with probable exceptions in Anatolian), are also situated somewhere within the left periphery:

(4.5) ***quid*** *apud hasce aedis negoti est*
INT.NOM.SG.N among DEM.ACC.PL.F house.ACC.PL.F business.GEN.SG.N be.PRES.3PL
tibi?
you.DAT.SG
'**What** is your business among these here houses?' (Plaut. *Amph.* 350)

(4.6) ***qui*** *sibi mandasset delegati ut*
REL.NOM.SG.M himself.DAT.SG order.PLUPERF.SUBJ.3SG assign.PPP.NOM.PL.M that
plauderent
applaud.IMPF.SUBJ.3PL
'He **who** commissioned people assigned to applaud him...' (Plaut. *Amph.* 83)

(4.7) τίς τ' ἄρ σφωε θεῶν ἔριδι ξυνέηκε
INT.NOM.SG.M and PTC them.ACC.DU god.GEN.PL.M strife.DAT.SG.F set-together.AOR.3SG
μάχεσθαι;
fight.INF
'And **which** of the gods sent these two to fight in contention?' (Hom. *Il.* 1.8)

(4.8) ὅν τινα μὲν βασιλῆα καὶ ἔξοχον ἄνδρα
REL.ACC.SG.M INDEF.ACC.SG.M PTC king.ACC.SG.M and eminent.ACC.SG.M man.ACC.SG.M
κιχείη
encounter.AOR.OPT.3SG
'**Whichever king** or eminent man he encountered...' (Hom. *Il.* 2.188)

(4.9) ***káḥ*** *naḥ mahyái áditaye púnaḥ dāt*
INT.NOM.SG.M us.ACC great.DAT.SG.F aditi.DAT.SG.F again give.INJ.3SG
'**Who** would give us again to great Aditi?' (RV 1.24.1c)

(4.10) ***yáḥ*** *te agne námasā yajñám*
REL.NOM.SG.M you.ACC.SG Agni.VOC.SG.M obeisance.INS.SG.N sacrifice.ACC.SG.M
íṭṭe
implore.PRES.3SG
'**Who**(ever) calls you to the sacrifice with an obeisance, O Agni...' (RV 5.12.6a)

In some IE languages it also appears to be possible for interrogatives to be preceded by a topicalised constituent:

(4.11) *múkham* *kím* *asya*
mouth.NOM.SG.N INT.NOM.SG.N his
'As for his mouth, what was it?' (RV 10.90.11c)

In addition to the elements enumerated above, a category that often appears in the left periphery are the local particles (P-words), which are often fronted when they appear separately from their respective verb (traditionally referred to as *tmesis*). Left peripheral preverbs are most prevalent in Vedic and Ancient Greek:

(4.12) ***prá*** *rudréṇa* *yayínā* *yanti* *síndhavaḥ*
forth.PW rudra.INS.SG.M travelling.INS.SG.M go.PRES.3PL river.NOM.PL.M
'The rivers go **forth** with journeying Rudra' (RV 10.92.5a)

(4.13) ἐς δ' ἐρέτας ἐπιτηδὲς ἀγείρομεν, ἐς δ'
onto.PW PTC rower.ACC.PL.M sufficiently collect.PRES.1PL onto.PW PTC
ἑκατόμβην / θείομεν
hecatomb.ACC.SG.F place.AOR.OPT.1PL
'**Onto** [the ship] let us gather sufficient rowers, and **onto** it let us put a hecatomb'
(Hom. *Il.* 142–3)

These examples seem to suggest that a lot of material ends up in the left periphery, moved out of its base-generated position. However, establishing the 'landing site' of elements within the left periphery of ancient IE languages presents some methodological difficulties. Rizzi (1997) was able to establish the distributional facts for TopP and FocP by grammaticality judgements in his native language; clearly the same is not possible for the languages under examination in this thesis. First, we cannot generate novel sentences to test their grammaticality; we can only examine the corpus as it survives, and test our hypotheses against the observed patterns. Second: the Topic-Focus complex is manifest not only in patterns of word-order, but also **discourse function**; the fronting of constituents is a syntactic process that serves a pragmatic function. Yet the precise discourse function of a given left-peripheral element in an ancient text is rarely securely diagnosable. Although in context we can sometimes make a fairly cogent argument as to whether a given element is topicalised, focalised, or just happens to occur first in the sentence, we can rarely claim *certainty* from context alone, especially in multi-authored and non-narrative texts. This means that in some way, we are perversely more reliant on the distributional syntactic evidence to diagnose the discourse function than we might be if we had access to speaker judgements. Yet the left periphery of the ancient IE languages is remarkable insofar as it exhibits some clear and even inviolable word-order constraints, of the sort that can rarely be established for these languages. These constraints, as well as implicating clause-level information, affect the position of the fronted relative pronoun; this makes them of central importance when reconstructing RC syntax.

Aside from fronted pronouns, local particles and topicalised/focalised constituents, there is another class of elements that occupy the left periphery in ancient IE languages: **clitics**.[93] In some ways,

[93] For the remainder of this section, these will be marked with an equals sign < = > in quoted examples.

clitics can complicate the picture: they are yet another set of elements to account for when mapping the left periphery of PIE. Yet because their occurrence in the left periphery is not a matter of (optional) pragmatic fronting, their distribution is relatively predictable synchronically. This means they provide a valuable diagnostic for assessing the position of other elements in the left periphery, whose movement there could be due to various discourse-oriented features, and whose 'landing site' is therefore variable. On the other hand, since clitics routinely occur within the left periphery, a proper account for clisis necessarily requires an articulated account of the left periphery in its own right.

There is a final reason that a comprehensive treatment of clisis and the left periphery is central to understanding RCs, and this is that descendants of *REL in various languages have been argued to exhibit clitic behaviour. This is a question I address in detail in §5.1; to motivate my analysis of *REL, I must first establish a coherent account of both clitics and non-clitics in the left periphery, and how we can tell the difference.

This chapter is structured as follows: in §4.2, I establish some preliminaries for our understanding of clitics and their movement to the left periphery ("Wackernagel's Law"). In §4.3 I take the Vedic 'initial string'—a touchstone for this topic within IE—and treat it in detail as a case study in disentangling the left periphery. With this groundwork established, in §4.4 I offer a comparative analysis of the left peripheries from three other major branches of ancient IE: Latin, Ancient Greek and Hittite. These form the basis for my reconstruction of PIE left periphery, a crucial component of relative clause structure.

4.2 Clisis

4.2.1 Clitic characteristics

The term 'clitic' is a back-formation from 'enclitic' (< AGk. ἐγκλίνω 'lean on'), a phonological term employed by early Greek grammarians to refer to words that do not bear their own inherent accent, but affect the accentuation pattern of the previous word. For example:

(4.14) ὁ ἄνθρωπος ἔστηκεν
the.NOM.SG.M man.NOM.SG.M stand.PERF.3SG
'The man is standing'

(4.15) ἄνθρωπός =τις ἔστηκεν
man.NOM.SG.M =INDEF.NOM.SG.M stand.PERF.3SG
'A (certain) man is standing'

Note here that the final syllable of ἄνθρωπος ('man') is usually unaccented, but when followed by the enclitic indefinite pronoun τις it acquires an additional accent on its final syllable. The traditional explanation for this phenomenon is that enclitics form part of the phonological word immediately to their left, and the 'extra' accent is essentially a prosodic repair strategy:

> Since an enclitic, on losing its accent, forms a part of the preceding word, the writing ἄνθρωπος τις would violate the rule...that no word can be accented on a syllable before the antepenult.
>
> (Smyth, 1956: §184)

From these phonological descriptions, a syntactic rule may be gleaned: since enclitics must cause accentuation of a **preceding** syllable, they definitionally cannot occur clause-initially. However, this definition excludes orthotonic (i.e. accent-bearing) postpositives such as μέν, δέ, γάρ κτλ. which also do not occur clause-initially, and often follow the first phonological word of a clause.[94] Enclitics can be contrasted with **pro**clitics, which behave in a similar way but require an appropriate prosodic host to their right; Nida coined the term the term **clitic** as an umbrella term for both enclitics and proclitics, providing the following definition (1946: 76, n.17):

> A clitic is a form which phonologically combines with an element with which it does not form a morphological construction.

Clitics are later subdivided into three categories by Zwicky according to their syntactic distribution in his pivotal work *On Clitics*, summarised below (1977: 5–7).

1. Clitics: have an accented non-clitic counterpart.
   (a) "Special": These have a distinct syntactic distribution from their non-clitic counterparts, and often cannot be derived from their accented counterpart by regular synchronic phonological processes, e.g. AGk. 3sg pronoun μιν vs. αὐτόν, αὐτήν; Skt. 2sg pronoun *te* vs. *túbhyam, tavá*.
   (b) "Simple": These occupy the same syntactic position as their non-clitic counterparts, and often appear the outcome of regular synchronic phonological processes, e.g. Lat. *-st* ('is') vs. *est*; PDE *'ll* vs. *will*.
2. Bound words: do not have an accented non-clitic counterpart, e.g. Gk. τε; Lat. *-que*; Skt. *ca* etc. ('and')

In discussing the syntactic distribution of clitics, Zwicky's categories introduce a new complexity. Whereas the earlier phonological definitions of enclitics implicitly prohibit their appearance in first position, they do not *prescribe* their appearance in, say, "second position" (2P). Yet, as we will see below, this is where certain enclitics regularly appear, forming a distinct syntactic distribution from their non-clitic counterparts. For our purposes then, we may divide clitic characteristics into two categories:

1. Phonological characteristics: Prosodic deficiency, lack of accent
2. Syntactic characteristics: [for **en**clitics] Cannot appear clause-initially, usually appear in 2P

Part of the difficulty in defining clitics is that they instantiate these characteristics to varying degrees; for example, the aforementioned orthotonic postpositives of Ancient Greek, while not strictly exhibiting all the phonological characteristics of a clitic, clearly show the syntactic characteristics of an enclitic. Since I am primarily concerned with syntax, I will not dwell on the specific phonetic criteria a word must fulfil to be a clitic, working instead with the umbrella term of 'prosodic deficiency' where the syntax-phonology interface is concerned; for this reason, I will refer to orthotonic postpositives as clitics too.[95]

[94] See van Emde Boas *et al.* (2019: 704–6) for a recent, brief overview.

[95] This approach has been taken by previous authors including Goldstein (2016a,b) for Ancient Greek, Lowe (2014) for Sanskrit, and Sideltsev (2017) for Hittite.

4.2.2 "Second position": Wackernagel's legacy

Building on Delbrück's (1878) observations from Sanskrit, and adding to them *comparanda* from other IE languages, Wackernagel (1892) established what we have come to refer to as "Wackernagel's Law". Although the author uses the word *Gesetz* ("law") in the title, he never actually states his discovery in such terms. As such, the law is often quoted as simply "Enclitics move to second position". This does not adequately capture his thesis, which clearly excludes some enclitics, such as Av. *cīṭ* (1892: 403), and includes some accented (i.e. non-enclitic) words, e.g. Gk. μέν, δέ (1892: 377–8). A more precise formulation is therefore given below:

(4.16) **Wackernagel's Law**
The following words occupy 2P:

(a) Enclitic pronouns

(b) Enclitic conjunctions and grammatical particles

(c) Orthotonic postpositives

(d) Verbs (insofar as V2 is part of the same phenomenon)

Wackernagel's inclusion of V2 on this list has faced serious criticism (Hock, 1982; Kiparsky, 1995); I will not treat verb movement as part of the same phenomenon, and exclude it from the discussion. We may also collapse categories (b) & (c), since they are differentiated only by the presence or lack of an inherent pitch accent, which does not factor into my syntactic analysis—both are clitics for our purposes. We are thus left with two relevant categories of second-position, or "Wackernagel" clitics:[96]

WL1: clitic pronouns

WL2: clitic conjunctions and grammatical particles

Some clitics that belong to WL2 exhibit 2P effects variously at either the clause-level or the phrase-level according to scope. A typical example is AGk. τε 'and'. When coordinating clauses, τε occupies 2P within the **clause**:

(4.17) ὁ =δ' ἐγγύθεν ὀξὺ λεληκὼς / ταρφέ '
DEM.NOM.SG.M =PTC from-near sharply shriek.PERFP.NOM.SG.M relentlessly
ἐπαΐσσει, ἑλέειν =**τέ** =ἑ θυμὸς ἀνώγει
rush-at.PRES.3SG seize.AOR.INF =and =her.ACC.SG.F spirit.NOM.SG.M command.PRES.3SG
'He lets out a piercing cry from nearby and rushes at her relentlessly, **and** his heart commands him to seize her' (Hom. *Il.* 22.141–2)

When coordinating noun phrases, however, τε occupies 2P within the **phrase**:

[96] These labels are used by Keydana (2011), reflecting distinctions made earlier in Hale (1987) and Krisch (1990).

(4.18) ὄφρα σὺ χαίρῃς / τιμὴν ἀρνύμενοι
that you.NOM.SG rejoice.PRES.SUBJ.3SG honour.ACC.SG.F gain.PRESP.NOM.PL.M
Μενελάῳ σοί =τε, κυνῶπα / πρὸς Τρώων
Menelaus.DAT.SG.M you.DAT.SG =and dogface.VOC.SG.M from Trojan.GEN.PL.M
'...in order to make you happy, striving to win recompense for Menelaus **and** you, dogface, from the Trojans' (Hom. *Il.* 1.159–60)

I will call the latter use 'phrasal' clitics: these do not participate in "Wackernagel's law": rather, they stay syntactically in their base-generated position and do not raise to the left periphery. As such I will not offer any detailed discussion of phrasal clitics.

Wackernagel's Law dictates that the enclitics enumerated above occupy 2P. However, as research on 2P clitics *passim* has demonstrated, the definition of 2P evades precise description. Goldstein (2016b: 60, n.9, emphasis mine) remarks that:

> One comes away with the impression that [for Wackernagel], second position was defined graphically: that is, that it referred to the **second orthographic word after a major mark of punctuation**.

This is to say that "regular" instances of Wackernagel's law, traditionally conceived, define 2P as simply following the first orthographic word, e.g.:

(4.19) ὅ =σφιν εὖ φρονέων ἀγορήσατο καὶ
DEM.NOM.SG =them.DAT.PL well think.PRESP.NOM.SG.M address.AOR.3SG and
μετέειπεν
speak-with.AOR.3SG
'He addressed **them** in the assembly, well-disposed, and spoke among them' (Hom. *Il.* 1.73)

(4.20) *per* =***te*** *deos* *oro* *et* *nostram* *amicitiam,*
by =you.ACC.SG gods.ACC.PL.M beg.PRES.1SG and our.ACC.SG.F friendship.ACC.SG.F
Chremes
Chremes.VOC.SG
'I beg **you**, Chremes, by the gods and our friendship' (Ter. *An.* 538)

As Goldstein (2016b: 60, n. 9) aptly notes, this definition is not 'linguistically real'. Considering the clitic characteristics enumerated above, we may amend this definition to refer to the first **prosodic** word (ω). This would allow us to admit some elements such as short PPs which can sometimes constitute a single ω (Probert, 2003: 133–42). This may account for some apparent exceptions to Wackernagel's law, e.g.:

(4.21) [PP ἐν τῷ] =ῥά =σφι κύκησε γυνὴ
in DEM.DAT.SG.M =PTC =them.DAT.PL mix.AOR.3PL woman.NOM.SG.F
ἐϊκυῖα θεῇσιν / οἴνῳ Πραμνείῳ
resembling.NOM.SG.F goddess.DAT.PL.F wine.DAT.SG.M Pramnian.DAT.SG.M
'In this the woman, like the goddesses, mixed [a potion] for them with Pramnian wine'
(Hom. *Il.* 11.638–9)

Yet another set of "exceptions" to Wackernagel's law are those where phonology alone cannot solve the issue; 2P must be understood as the second **syntactic** constituent. This order is most prevalent in Sanskrit:

(4.22) ***amṛtatvám*** ***rákṣamāṇāsaḥ*** *=enam* *devā́ḥ* *agním*
immortality.ACC.SG protecting.PPL.NOM.PL =him.ACC god.NOM.PL agni.ACC.SG
dhārayan *draviṇodā́m*
preserve.IMPF.3PL giver-of-goods.ACC.SG
'**protecting (their) immortality**, the gods preserved him as Agni, giver of goods' (RV 1.96.6cd)

Its also possible, albeit rarer, for 2P clitics to appear as neither the second word nor the second constituent, in "third position" or later by any measure:

(4.23) *utá* *=vā* *yáḥ* ***=naḥ*** *marcáyāt* *ánāgasaḥ* *arātīvā́*
and =or REL.NOM.SG.M us.ACC harm.SUBJ.3SG innocent.ACC.PL evil.NOM.SG.M
mártaḥ *sānukáḥ* *vṛ́kaḥ*
mortal.NOM.SG.M eager.NOM.SG.M wolf.NOM.SG.M
'Or the evil mortal or eager wolf who would harm us, innocent as we are...' (RV 2.23.7ab)

(4.24) μήτηρ =δ' οὔ =με φίλη πρίν =γ' εἴα
mother.NOM.SG.F =PTC NEG =me.ACC.SG dear.NOM.SG.F before =PTC allow.IMPF.3SG
θωρήσσεσθαι
arm-oneself.INF
'And my dear mother did not allow **me** to array myself before...' (Hom. *Il.* 18.189)

Moreover, these examples demonstrate the fact, as noted by Hale (1987), Hock (1989) and subsequent work, that WL1 and WL2 have different distributions: the latter regularly precedes the former:[97]:

(4.25) *kéna* ***=vā***$_{\text{WL2}}$ ***=te***$_{\text{WL1}}$ *mánasā* *dāśema*
INT.INS.SG.N =or you.DAT.SG mind.INS.SG.N honour.OPT.1PL
'Or in what mind should we honour you?' (RV 1.76.1d)

(4.26) τὰς =δ'$_{\text{WL2}}$ ἄλλοι =με$_{\text{WL1}}$ κέλονται ἀγινέμεναί =σφισιν
DEM.ACC.PL.F =PTC other.NOM.PL.M =me.ACC.SG tell.PRES.3PL bring.INF =them.DAT.PL
αὐτοῖς ἔδμεναι
themselves.DAT.PL.M eat.INF
'And others tell me now to bring these out for themselves to eat' (Hom. *Od.* 20.213)

In sum: the surface-level observations of Wackernagel (1892) have quite remarkably stood the test of time. Indeed, his hypothesis that 2P effects were inherited from the proto-language was all but certified with the decipherment of Hittite (on which see §4.4.3 below). However, it is eminently clear that "Wackernagel's Law" cannot be described with a singular generalisation ("enclitics move to second position"), but rather requires a detailed grammatical analysis in the context of a finer understanding of the left periphery in PIE.

[97] Sanskrit ex. from Keydana (2011: 108).

In what follows, I take the patterns of clitics and left periphery as attested in Vedic as a case study, establishing the various phonological and syntactic factors that conspire to yield the attested distributions. From this model I move onto a comparative analysis and reconstruction of the PIE left periphery in §4.4.

4.3 The Vedic initial string: A case study

Traditional literature on the topic of the left periphery in ancient IE languages has often referred to it as the **initial string**. Authors have noted the lexical elements it may contain (cf. §4.1 above), and have observed that every element within the left periphery is optional, to the extent that it is possible not to have an initial string at all. This is not surprising, since we understand the initial string to be comprised primarily of the Topic-Focus complex of Rizzi (1997), which is optional (see §2.3). Perhaps the most elaborate iteration of the initial string among the attested IE languages is that found in the language of the Rig Veda; this is reflected in the volume of research that has been dedicated to mapping and accounting for its patterns. As such, it provides an apt case study to being to understand how clitics and lexical items are organised in the left periphery.

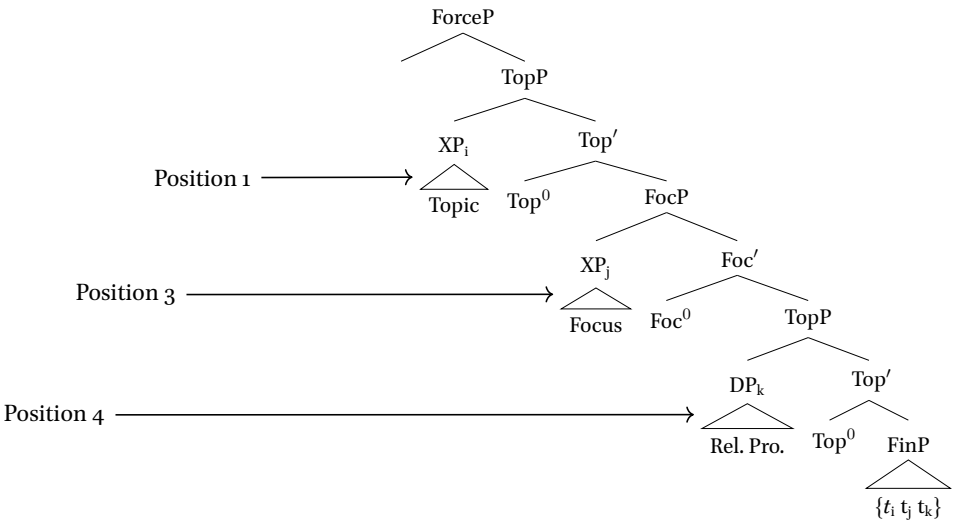
The basic schema of the Vedic initial string can be found in Table 5.1 below: a synthesis of Hock (1989, 1996), Hale (1987, 1996), Keydana (2011), and Lowe (2014), together with my own observations.

Position	0	1	2	3	4	5	6
Type	Conj.	Topic	WL2	Focus	Rel. Pro.	WL1	Dem. Pro.
E.g.	*sá, átha*	[any XP]	*ca, vā*	*mā́, káḥ*, P-word etc.	*yáḥ*	*me, enam* etc.	*sáḥ*

Table 4.1: The Vedic Initial String

Under the theoretical assumptions of this thesis, Positions 1, 3 and 4 are accounted for straightforwardly by the Topic-Focus complex posited in Rizzi (1997):

(4.27)



The main point of contention, therefore, is where the clitics fit into this structure; in what follows, I establish a model for the Vedic left periphery and the positions of clitics within this.

Traditionally, the accounts fall into two theoretical categories: prosody-dominant approaches (Hock, 1989, 1993, 1996; Krisch, 1990; Keydana, 2011), and syntax-dominant approaches (Hale, 1987, 1996, 2007, 2017; Lowe, 2011, 2014).[98] With the exception of the earlier work of Hale (1987) and Hock (1996), all authors accept that both phonology and syntax are involved to some degree; the question is to what extent. On the whole, these approaches are divided more along theoretical lines than empirical ones. Advocates of prosody-dominant approaches tend to view clisis as an **inherently** phonological phenomenon, while syntax-dominant approaches tend to assume the null hypothesis that all word-order (including clitics) is primarily accounted for in the syntax.

4.3.1 Prosody-dominant approaches to Vedic clisis

Hock (1982) was perhaps the earliest work that analysed the Vedic initial string as an ordered sequence of positions, which he summarises in a 'taxonomic form', later referred to as the **template** of the initial string. The template is revisited and revised in Hock (1989, 1993, 1996); reproduced in Fig. 5.1 overleaf is the final iteration.

"NEXUS"	1	2	3	4	5
	{ Ḍ́ / Ẋ́ }	(P)	(Ṕ / < Ḍ́ >)	(E)	(Ḍ́)

Where:	NEXUS	=	Quasi-conjunctional elements such as Ved. Pr. *átho*
	Ṕ	=	accented sentential particle
	P	=	unaccented sentential particle
	Ḍ́	=	accented deictic (including demonstrative *tád*, *etád*, relative *yá-*, interrogative *ká-* etc.); in the Rig-Veda this category includes preposition/ adverbs (always accented)
	E	=	unaccented pronominal (both deictic and personal)
	Ẋ́	=	other accented elements

Figure 4.1: Template of the Vedic initial string (Hock, 1996: 291).

All elements are optional, and all but Position 1 can be doubled.[99] Note that positions 0, 1, 2 & 3 correspond partially with my own labelling in Table 4.1. However, Hock's position 3 incorporates my positions 3 and 4. This is a byproduct of Hock's use of the category Ḍ́, which collapses several important patterns concerning the orderings of P-words, interrogative and relative pronouns (Lowe, 2014 to be discussed below). Setting that aside, however, Hock's position 5 does correspond with my position 6, housing "late" demonstratives, usually in correlative clauses.

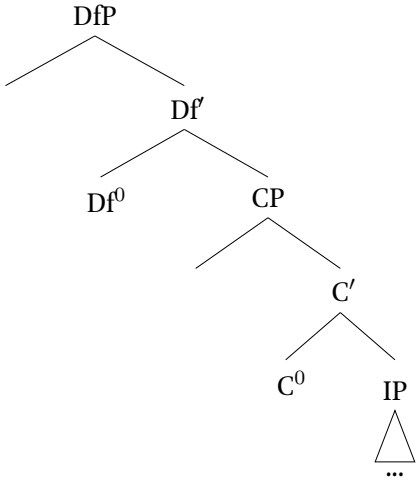
[98] I adopt this terminology from Goldstein (2010).

[99] This is what is implied by the {curly brackets}. The <angled brackets> around Ḍ́ in Position 3 indicate that this word order is only acceptable in RV and not in Vedic prose.

Hock's template was a vital stage in mapping the Vedic initial string, but lacks explanatory power. This criticism is made at length in Hale (1996), Keydana (2011) and Lowe (2014); in the interest of space I do not revisit their arguments here.[100] The central issue is that the template is conceived of as strictly phonological; yet while several authors have argued that clitic movement can be motivated by phonology,[101] the same cannot be said of the full lexical items that also form part of Hock's template. Rather, the pragmatic fronting of full lexical items is best accounted for under the Rizzian model of the left periphery as adopted in this thesis.

Thus in a bid to rectify the deficiencies of the templatic approach while still treating clisis as inherently phonological, Keydana (2011) admits that full lexical items are placed in the initial string/left periphery by the syntax. He assumes the following structure of the left periphery, where DfP = 'Discourse-function Phrase' (Keydana, 2011: 112):

(4.28)



This model is a partial implementation of Rizzi (1997): Keydana (2011) admits one position higher than [Spec, CP], to account for both Topics and Foci ([Spec, DfP]). He does not adopt the Force°-Fin° split, instead modelling a unitary C° at the bottom of the left periphery. The author models "*wh*-words" (both relative and interrogative) as occupying [Spec, CP]. In the context of this model, Keydana proposes the following two rules:

For any given intonational phrase:

(i) WL2 (clausal clitics) follow the first phonological word (ω)

(ii) WL1 (pronoun clitics) follow the first phonological phrase (φ)

Rule (i) is fairly uncontroversial: enclitics, lacking an inherent pitch accent, are prosodically infelicitous in clause-initial position. When they reach PF, they cannot be spelled out unless they have a suitable, accent-bearing host to their left: namely, ω. This is formalised in Halpern's "prosodic flip", and is broadly accepted in both prosody- and syntax-dominant explanations of Vedic clisis (see §4.3.2 below).

[100] See Ram-Prasad (2020) for a brief summary incorporating some of my own criticism.

[101] See, e.g., Radanović-Kocić (1988, 1996) and Aissen (1992), whom Hock (1996) cites in support of this model.

The question, however, is: how do we establish the domain of φ in Vedic? As Keydana himself notes (2011: 123), and Lowe (2014: 12) re-iterates, we have no comprehensive proof of what constitutes a φ in this language. Keydana therefore opts to follow Nespor & Vogel (2007), who define φ in syntactic terms, based on the distribution of *raddoppamiento sintattico*, a connected speech process that occurs within φ but not across φ-boundaries in some varieties of Italian. Their definition is as follows (2007: 168, emphasis mine):

> *Phonological Phrase Formation*
> *φ domain*
> The domain of φ consists of a *C*[litic Group][102] which contains a **lexical** head (X) and all *C*s on its nonrecursive side up to the *C* that contains another head outside of the maximal projection of X.

Nespor & Vogel are rather explicit in stating that '[t]he intended interpretation of [this definition] is that in which only V, N, and A are considered lexical heads' (Nespor & Vogel, 2007: 168). Keydana pays some heed to this consideration in suggesting that Df° cannot be the starting point for φ-construal, because it is 'not only a functional head, but moreover one which is never filled with lexical material' (Keydana, 2011: 123). Yet he finds no issue with C° being the starting point for φ-construal, despite it also being a functional head, and often phonologically null. Nevertheless, these issues aside, Keydana argues that the first φ of the Vedic sentence is construed as the entire left periphery, up to and including C°. On this basis then, since WL1 clitics are placed after C° but before the rest of the clause, they must follow the first φ.

Thus, even though Keydana asserts that clisis is an inherently phonological phenomenon, he must resort to a purely syntactic definition of the phonological phrase in Vedic. If one accepts his premise that 'the null hypothesis [is] that clitic placement is a PF phenomenon' (Keydana, 2011: 122), one can plausibly argue that the syntax diagnoses the phonology, and the latter is the basis for clisis. But I do not accept this premise: Keyadana's prosody-dominant model makes the same predictions as a syntax-dominant one in which WL1 raise to C°—this happens to be precisely the model advocated by Hale (1996) and Lowe (2014), to whom I turn presently.

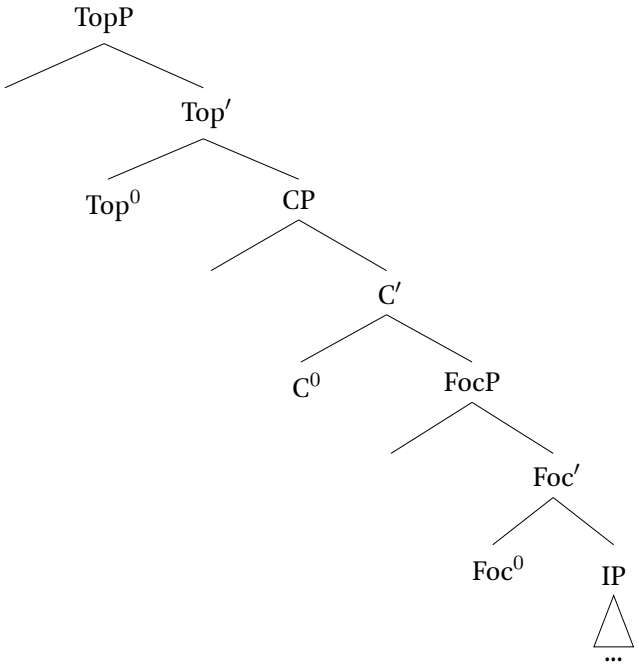
4.3.2 Syntax-dominant approaches to Vedic clisis

Over three decades, Hale (1987, 1996, 2007, 2017) develops an account of the Vedic initial string within, broadly speaking, a Chomskyan framework. The most detailed account is given in Hale (1996), which is written in (a somewhat abrasive) dialogue with Hock (1996). Hale broadly accepts the descriptive adequacy of Hock's template,[103] but seeks to provide a structural account for Hock's observations. He postulates the following structure for the Vedic initial string (1996: 177):

[102] I.e., ω + clitics.

[103] Many of the empirical criticisms Hale (1996) makes are, as Hock (1996) notes, based on a misinterpretation of the template put forward in Hock (1989).

(4.29)



[Spec, TopP] corresponds to Hock's (and my) position 1. Hock's position 3, (my positions 3–4), is accounted for by [Spec, CP].[104] Hale's FocP, placed below [Spec, CP] is only 'provisional', and is not mentioned in his later writings; it is supposed to account for Hock's position 5 (my position 6).

Hale argues that all the non-clitic elements of the left periphery arrive there by syntactic movement. Hale's left periphery is remarkably similar to Rizzi's (1997) in structure: although he posits a unitary C°, like Rizzi he allows for a proliferation of left-peripheral heads motivated by discourse factors. Where Hale's approach departs from the prosody-dominant ones is how he accounts for the positioning of clitics.

Hale argues convincingly that Vedic clitic placement is best explained as an interaction of syntactic and phonological constraints. The latter is inspired by the work of Halpern (1992, 1995): namely the "prosodic flip", more properly called **Prosodic Inversion** (PI). PI is a late-stage phonological process that Halpern (1995: 17) defines in the following terms:

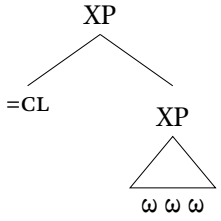
> [T]he position of a clitic in the surface string of a sentence may diverge from what would be expected based on its syntactic position. Specifically, a clitic may "trade places" with a prosodic unit which is adjacent to it... Usually this unit is the prosodic word.

In short: where an enclitic leaves the narrow syntax and enters PF in a prosodically infelicitous position (viz. clause-initially[105]), PI "inverts" the ordering of the clitic and the prosodic word (ω) immediately to its right, such that the enclitic has an appropriate prosodic host. This means that the narrow syntax **can** allow a clitic to be generated in—or moved to—clause-initial position; PI will then "fix" the outcome at PF, yielding the attested word order. This is schematised below:

[104]More precisely, Hale (1996: 173) suggests that 'inflected WH-elements' occupy [Spec, CP], while 'uninflected WH-elements...are usually taken to be in C° itself'. The author does not expand on what he thinks constitutes an 'uninflected WH-element' in Sanskrit. He seems to suggest (1996: 172) that adverbial/conjunctional *yad* occupies C°, but does not explain his choice. In line with other authors (e.g., Lowe, 2014), my analysis will not reflect this distinction.

[105]Or phrase-initially for phrasal clitics.

(4.30) i. Narrow syntax:

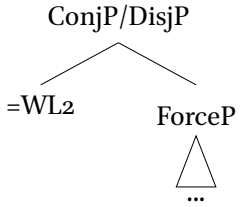


ii. Prosodic Inversion: =CL ω=CL ω ω.

iii. Phonetic Output: ω=CL ω ω

Equipped with the machinery of PI, Hale can then account for WL2 clitics straightforwardly. He argues that they are base-generated where we would expect them to be, i.e. as heads of their own ConjP/DisjP which is **external** to the clause. Thus when they occur in the left periphery, their projection dominates the entire clause:

(4.31)

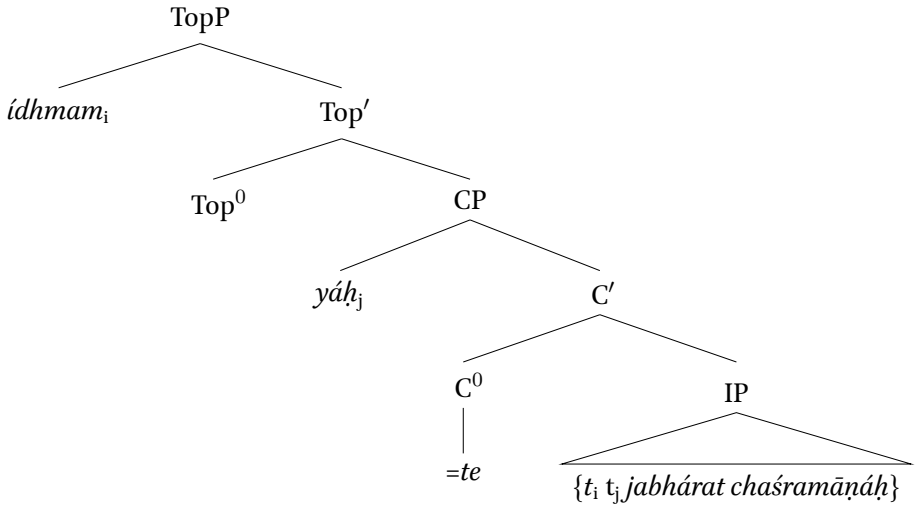


They are then moved by PI into 2P, i.e. following the first ω. Hale uses the combination of external base-generation + PI to yield Hock's (and my) position 2. Descriptively, his claim is the same as that later made by Keydana: WL2 follow the first ω of a phrase (see §4.3.1 above). While the precise mechanisms may differ, there is a point of agreement here.

Hale's account for WL1 is where the syntax-dominant approach diverges most sharply from the prosody-dominant one. In line with other syntactic approaches to clisis, and reminiscent of Delbrück (1878), Hale assumes that WL1 are base-generated in the same position as their non-clitic counterparts, i.e. somewhere within the VP-complex. For them to appear in the left periphery at all then, they must first **raise to C°**. This movement takes place in the narrow syntax, with no prosodic involvement, and supposedly accounts for Hock's position 4 (my position 5). This is how Hale accounts for examples such as the following (1996: 168):

(4.32) *ídhmam* *yáḥ* *=te* *jabhárat* *chaśramāṇáḥ*
kindling.ACC.SG REL.NOM.SG.M =you.DAT.SG bear.PERF.3SG exert.PRESP.NOM.SG.M
'Who, exerting himself, bore the kindling to you' (RV 4.12.2a)

(4.33)



In this way, Hale argues that where WL1 appear following an element in either [Spec, TopP] or [Spec, CP] (or both), the phonology is not involved at all; the movement is entirely in the narrow syntax. However, if the left periphery is empty, PI will kick in (exactly as with WL2) and move WL1 such that they appear after the first ω. Such an example is provided by Keydana (2011: 124):

(4.34) *gandharváḥ* *=asya* *raśanā́m* *agṛbhṇāt*
gandharva.NOM.SG =DEM.GEN.SG bridle.ACC.SG sieze.IMPF.3SG
'The Gandharva seized his bridle' (RV 1.163.2c)

Under Hale's model, *gandharváḥ* would likely occupy the canonical subject position, [Spec, IP]. When *asya* moves to C°, it leaves the narrow syntax in (unacceptable) first position; PI then pushes it behind the first ω, *gandharvaḥ*.

To summarise, Hale posits two mechanisms to account for clitic placement:

1. Prosodic Inversion: Both WL1 (if necessary) and WL2 (always)
2. Movement to C°: WL1 only

Hale uses his twofold explanation to account for a pattern which Hock's original template somewhat glossed over: the fact that position 1 can either be a single word, or a single constituent. Since PI does not see the syntax, it can only "push" the enclitic behind the first ω, regardless of constituent structure.

(4.35) ***devéna*** *=naḥ* ***mánasā*** *deva* *soma* *rāyáḥ*
divine.INS.SG =us.DAT mind.INS.SG divine.VOC.SG soma.VOC.SG wealth.GEN.SG
bhāgám *sahasāvan* *abhí* *yudhya*
share.ACC.SG mighty.VOC.SG upon fight.IMPV.2SG
'**With your divine mind**, oh divine Soma, oh mighty one, fight for a share of the wealth for us'
(RV 1.91.23a)

For the above sentence, Hale's hypothesis would lead us to posit that the constituent *devéna mánasā* does not occupy a left-peripheral position. As such, *naḥ*, having raised to C°, is then moved by PI to follow the first ω, and **not** the first syntactic constituent. This contrasts with an example such as the following (repeated from ex. (4.22) above):

(4.36) ***amṛtatvám*** ***rákṣamāṇāsaḥ*** *=enam* *devā́ḥ* *agním*
immortality.ACC.SG protecting.PPL.NOM.PL =him.ACC god.NOM.PL agni.ACC.SG
dhārayan *draviṇodā́m*
preserve.IMPF.3PL giver-of-goods.ACC.SG
'**protecting (their) immortality**, the gods preserved him as Agni, giver of goods' (RV 1.96.6cd)

Hale hypothesises that in ex. (4.36), and others where WL1 follow a whole constituent consisting of more than one ω, it is because that constituent has been topicalised. Since [Spec, TopP] precedes C°, *enam* is not moved in PF: it simply surfaces precisely where we would expect C°. I re-iterate here that WL2 clitics, since they are always outside the left-most edge of the clause, and are only moved to 2P by PI, will always pattern with ex. (4.35): they follow the first ω even if it splits up a fronted constituent.[106]

Hale's model thus captures a significant portion of the data—but not quite all of it. The central flaw is a failure to distinguish between the position of relative and interrogative pronouns, collapsing them under the category of "*wh*-words" undergoing "*wh*-movement" (see §1.3 and §2.3). However, under the Rizzian model of the left periphery, we understand relative and interrogative pronouns to occupy different positions within the left periphery. It should come as little surprise, therefore, that interrogative and relative pronouns show different distributions in Vedic. This is demonstrated for Vedic by Lowe (2014), who notes that while local particles (P-words) regularly precede relative pronouns in the Vedic left periphery,[107] P-words **never** precede interrogative pronouns. Hale (1996) treats P-words as adverbs adjoined to CP: this predicts that they follow whatever is in [Spec, TopP], but precede whatever is in [Spec, CP]. This makes the correct prediction if we assume **relative** pronouns occupy [Spec, CP]: Hale (1996: 185) lists several examples that back up his point. No such examples, however, can be found with **interrogative** pronouns.

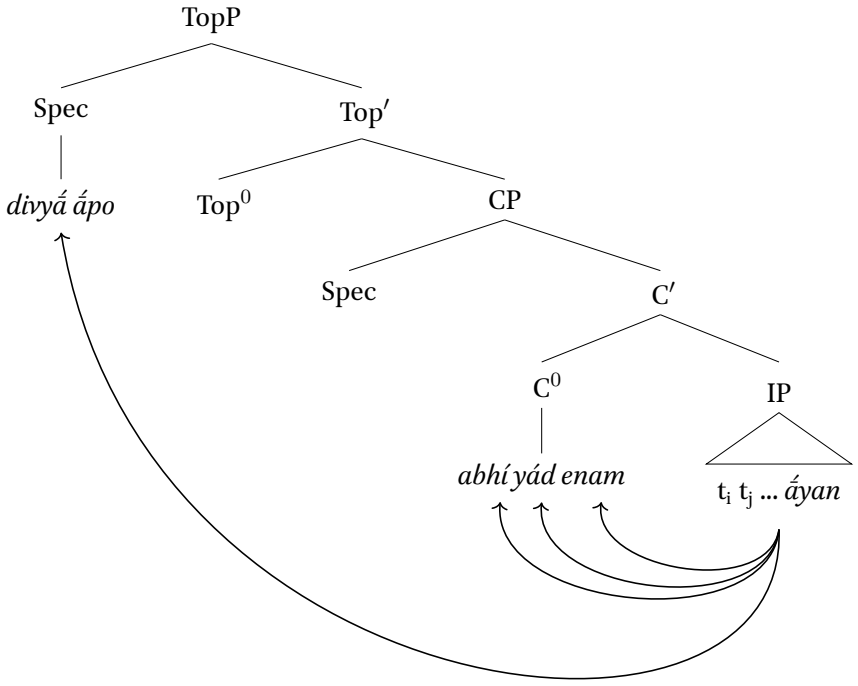
Lowe (2014) presents a rather ingenious solution that accounts for these patterns. He maintains that interrogative pronouns move to [Spec, CP], but that relative pronouns are **optionally** enclitic: in which case they raise to C°, forming a clitic cluster with WL1 (which raise to C° per Hale, 1996). To generate the order P-word – RelP, he also hypothesises that P-words are optionally proclitic; they too raise to C°, but must occur at the beginning of the clitic cluster. The following is an illustrative example, reproduced from (Lowe, 2014: 34):

(4.37) *divyā́ḥ* *ā́paḥ* *abhí* *yád* *=enam* *ā́yan*
divine.NOM.PL waters.NOM.PL toward.PW REL him.ACC come.IMPF.3PL
'When the divine waters came upon him...' (RV 7.103.2a)

[106] One exception to this is 'subordinating' *ca*, which Hale (2017) hypothesises to be generated in C°; see therein for the details, which are beyond the scope of the current discussion.

[107] He notes four exceptions in RV where the relative pronoun precedes the P-word: 4.34.3; 8.50.8; 7.38.3; 10.126.2.

(4.38)



As I mentioned above (n.95), Lowe maintains that the presence or absence of an inherent accent is not a defining feature of clisis; it is the clustering in C° that defines the category here. Thus, Lowe argues that in the majority of cases, *yá-* is not enclitic: its most common position is absolutely clause-initially, in which case it must be a full lexical word. In such instances, Lowe hypothesises *yá-* does not occupy the clitic cluster in C° but rather [Spec, TopP]. It is only in instances where [Spec, TopP] is filled by another constituent, as in ex. (4.37) above, that *yá-* can be understood as enclitic.

In support of this hypothesis, Lowe lists phonological evidence from the demonstrative pronoun *sá-*(/*syá-*) which he also argues is optionally enclitic. In the same way as *yá-*, *sá-* most usually occurs in absolute initial position, and so cannot be consistently enclitic. Nevertheless, Lowe argues that, eschewing accentuation as a phonological diagnostic of clisis, we can use sandhi phenomena instead when *sá-* occurs non-initially. He argues that in examples such as the following, where initial *t* or *s* is retroflexed by a preceding word according to the rules of word-internal sandhi, it is because the demonstrative is actually enclitic on the preceding word.[108]

(4.39) *pári* ***ṣyá*** *suvānó* *akṣā* *índur*
around DEM.NOM.SG.M pressing.PPL.NOM.SG.M flow.PERF.3SG drop.NOM.SG.M
ávye *mádacyutaḥ*
sheep.LOC.SG moving-ecstatically.NOM.SG.M
'That drop, having been pressed, flows through the sheep's wool, moved in ecstasy'
(RV 9.98.3ab)

(4.40) *agníṣ* ***ṭá*** *víśvā* *bhúvāni* *veda*
agni.NOM.SG DEM.ACC.PL.N all.ACC.PL.N world.ACC.PL.N know.PERF.3SG

[108] Examples from Lowe (2014: 21–2). Sandhi is not dissolved in exx. (4.39–41), since it is the point of discussion.

'Agni knows all those worlds' (RV 3.55.10c)

Since the phoneme *y-* is never affected by sandhi, there can be no direct evidence from the relative pronoun *yá-* akin to that of *sá-*. Lowe reaches his conclusion that *yá-* is optionally enclitic by the following steps:

1. Demonstrative *sá-*, when it occurs in non-initial left-peripheral position, is enclitic (witness sandhi phenomena)
2. **Therefore:** non-initial, left-peripheral demonstrative *sá-* occupies C° (the position for WL1 clitics)
3. Demonstrative *sá-* and relative *yá-*, when occuring non-initially in the left periphery, share the same distribution (they are both preceded by P-words, unlike interrogatives)
4. **Therefore:** since demonstrative *sá-* and relative *yá-* share the same distribution, and the former is part of the clitic cluster C°, **non-initial left-peripheral relative *yá-* is enclitic**

I have three core criticisms of Lowe's argument. First, I am not entirely convinced that we can use sandhi as a diagnostic for syntactic clisis. As Lowe himself notes (2014: 23), the lexical verb *stu* ('praise') also undergoes retroflexion in the same way as *sá-*, implying it forms a phonological word with what precedes it. However, Lowe does not seem to suggest the verb in such examples raises to C° here, which would in effect be a reprise of Wackernagel's hypothesis that V2 was part of the same set of phenomena as clitic movement. Indeed, in the case of *stu* there is at least one example where it is retroflexed sentence-finally, well away from the left periphery:

(4.41) *té* *me* *āhur* ... *náro* *máryā*
DEM.NOM.PL.M me.ACC tell.PERF.3PL ... man.NOM.PL.M youth.NOM.PL.M
arepása *imắn* *páśyann* *íti* ***ṣṭuhi***
unblemished.NOM.PL.M DEM.ACC.PL.M seeing.PPL.NOM.SG QUOT praise.IMPV.2SG
'They said to me... the men, the unblemished youths, "when you see them, praise them!"'
(RV 5.53.3ab)

This casts some shade on the claim that *sá-* is **syntactically** enclitic (i.e. it has moved to C°) in the examples Lowe provides.

Second: even if we were to accept that *sá-* is enclitic, there is a key distributional difference between it and *yá-*: the relative position of WL1 clitics. As Hale (1996: 181–2) notes, while *yá-* regularly precedes WL1, *sá-* commonly **follows** them:

(4.42) *prá* ***yáḥ*** *=vām* *mitrā* *varuṇā* *jīráḥ*
forth.PW REL.NOM.SG.M =you.ACC.DU mitra.VOC varuna.VOC quick.NOM.SG.M
dūtáḥ *ádravat*
messenger.NOM.SG.M run.IMPF.3SG
'The quick messenger **who** ran to you, oh Mitra and Varuna' (RV 8.101.3ab)

(4.43) *yáḥ* *mártyaḥ* *śíśīte* *áti* *aktúbhiḥ,* *mā́* *=naḥ*
REL.NOM.SG mortal.NOM.SG sharpen.PRES.3SG over night.INS.PL NEG =us.ACC
sáḥ *ripúḥ* *íśata*
DEM.NOM.SG rogue.NOM.SG rule.INJ.3SG
'The mortal who sharpens his weapons through the night, let not **that** rogue rule over us'
(RV 1.36.16cd)

A model in which *yá-*, *sá-* and WL1 all cluster together in C° does not automatically account for this word order pattern.

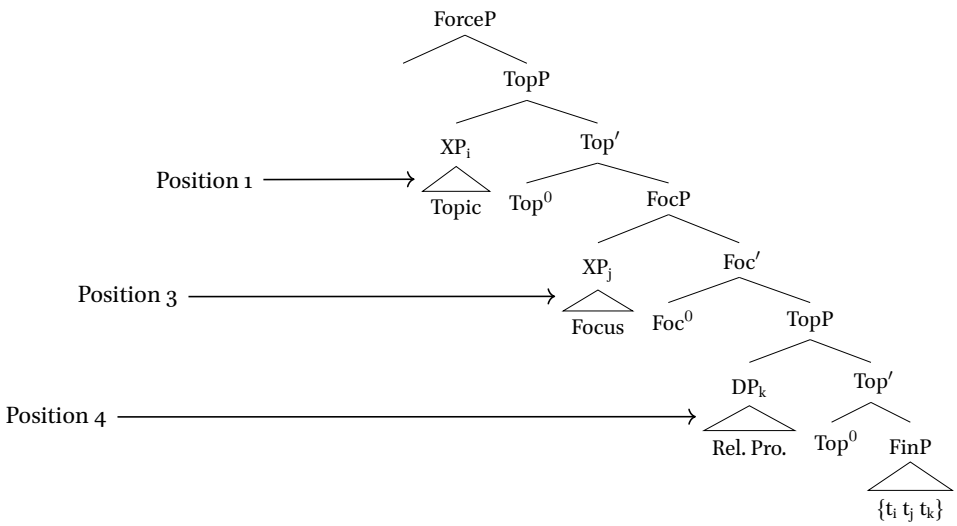
Finally, there is the issue of diachrony: Lowe (2014) must concede that the partial clitic status of Vedic *yá-* is entirely lost in the later stages of Sanskrit. Depending on whether we hypothesise PIE **yó-* was enclitic or not, we must assess the probability of this enclitic behaviour making a brief foray into the feature specification of Skt. *yá-* before disappearing without a trace. I will address this issue from a comparative perspective in §5.1; ultimately, I do not believe PIE *yó-* (*qua* *REL) was enclitic.

In sum, Lowe (2014) synthesises the strongest aspects of Hale's model, and seeks to improve upon its shortcomings. However, his hypothesis that relative pronouns are enclitic still faces several difficulties.

4.3.3 Vedic clisis reconsidered

I reproduce my template for the Vedic left periphery from ex. (4.27).

(4.44)



I account for Position 2 (WL2) in the same way as Hale and Lowe. I take WL2 to be generated outside the clause, and moved behind the first ω by some prosodic mechanism: I take no strong position as to whether we think of this as PI à la Halpern, or something else, such as optimality constraints (Lowe, 2011; Hale, 2017); nonetheless I will continue to use PI as a shorthand for whatever PF process is responsible for the surface order. As for WL1: I must revise the movement hypothesised by Hale and Lowe in accordance with the Rizzian model of the left periphery. Before doing this, however, I will revisit position 3: [Spec, FocP]. This position, nestled between recursive TopPs, follows directly from Rizzi's

template, housing interrogative pronouns, as well as focalised negators such as *mā́* and quantifiers such as *sárva-* ('all'). To this group, I add **P-words**. Lowe observed that P-words never precede interrogative pronouns. In fact, I would argue that the pattern is even stronger than that: **P-words and interrogative pronouns do not co-occur at all in the left periphery.** Lowe (2014: 16, n. 15) concedes that there are no absolutely unambiguous examples of interrogative pronoun + P-word within the left periphery in Vedic. He lists four examples where the P-word could be analysed as occupying a left-peripheral position following the interrogative pronoun; but in every instance, an alternative analysis is possible. In three of them, the P-word directly precedes a verb, with which it may be considered univerbated. In the fourth example, it could be a postposition governing the interrogative pronoun:

(4.45) ***kásmin*** [*ā́* *yatathaḥ*] *jáne*
INT.LOC.SG toward.PW join.PRES.2DU people.LOC.SG
'To **which** people do you join yourselves?' (RV 5.74.2c)

(4.46) ***kám*** [*áchā* *yuñjāthe*] *rátham*
INT.ACC.SG toward.PW yoke.PRES.2DU chartiot.ACC.SG
'**Who** do you yoke your chariot towards?' (RV 5.74.3b)

(4.47) *rátham* ***káḥ*** [*niḥ* *avartayat*]
chariot.ACC.SG INT.NOM.SG out.PW turn.IMPF.3SG
'Who rolled out the chariot?' (RV 10.135.5b)

(4.48) [***kám*** *ā́*] *jánam* *carati*
INT.ACC.SG toward.PW people.ACC.SG go.PRES.3SG
'Towards **which** people does he go?' (RV 6.21.4b)

Lowe cautiously suggests that these four *possible* examples of P-words co-occurring with an interrogative pronoun in the left periphery are promising, given the relative rarity of interrogatives. Yet if we model P-words as occupying [Spec, FocP], the lack of co-occurrence is **predicted**: since FocP is non-recursive, there can only be one focalised element in a clause. Moreover, this model would predict that left-peripheral P-words cannot co-occur in the left periphery with other focalised elements, such as the negator *mā́*. And indeed, this prediction seems to be correct: the Rig Veda contains no clauses in which *mā́* is preceded by a P-word. On the other hand, there is robust evidence for the ordering *mā́* – P-word; in every instance, however, the P-word also directly precedes the verbal stem, as in exx. (4.45–7) above:

(4.49) ***mā́*** [*prá* *gāma*] *patháḥ* *vayám*
NEG forth.PW go.INJ.1PL path.ABL.SG we.NOM
'**Let us not** go away from the path' (RV 10.57.1a)

This evidence further supports an analysis in which fronted P-words occupy [Spec, FocP]. And if we accept this hypothesis, we get another ordering constraint for free: P-words precede relative pronouns, which consistently move to the **lower** [Spec, TopP]. This is not such a departure from Hale's hypothesis that they occupy [Spec, CP], i.e. somewhere lower than the topmost TopP. The central difference is that the lower TopP is also below FocP, yielding the orderings with P-words above. That the

relative pronoun should move to [Spec, TopP] is also supported by the arguments from Bianchi (1999, 2000), discussed in §2.3.

We can now turn to WL1 (pronoun clitics). In sentences where the left periphery is inactive, we can model Force°/Fin° as a unitary C°. In such cases, we may hypothesise (in agreement with Hale and Lowe) that WL1 raise to C°, and are then pushed back behind the first ω in PF. However, when the left periphery is active, WL1 will target one of the left-peripheral functional heads: specifically, **WL1 will target the lowest Top° or Foc°**, yielding the following possibilities. Each word order is attested in the examples I have given above:

i. $[_{\text{FocP}}$ XP $[_{\text{Foc'}}$ Foc°=WL1$_i$ $]$ $[_{\text{FinP}}$ (...) t_i (...) $]]$ ex. (4.9)

ii. $[_{\text{TopP}}$ XP $[_{\text{Top'}}$ Top°=WL1$_i$ $]$ $[_{\text{FinP}}$ (...) t_i (...) $]]$ ex. (4.10)

iii. $[_{\text{TopP}}$ XP $[_{\text{Top'}}$ Top° $[_{\text{FocP}}$ YP $[_{\text{Foc'}}$ Foc°=WL1$_i$ $]$ $[_{\text{FinP}}$ (...) t_i (...) $]]]]$ ex. (4.11)

iv. $[_{\text{TopP}}$ XP $[_{\text{Top'}}$ Top° $[_{\text{TopP}}$ YP $[_{\text{Top'}}$ Top°=WL1$_i$ $]$ $[_{\text{FinP}}$ (...) t_i (...) $]]]]$ ex. (4.32)

v. $[_{\text{FocP}}$ XP $[_{\text{Foc'}}$ Foc° $[_{\text{TopP}}$ YP $[_{\text{Top'}}$ Top°=WL1$_i$ $]$ $[_{\text{FinP}}$ (...) t_i (...) $]]]]$ ex. (4.42)

vi. $[_{\text{TopP}}$ XP $[_{\text{Top'}}$ Top° $[_{\text{FocP}}$ YP $[_{\text{Foc'}}$ Foc° $[_{\text{TopP}}$ ZP $[_{\text{Top'}}$ Top°=WL1$_i$ $[_{\text{FinP}}$ (...) t_i (...) $]]]]]]]$ ex. (4.37)

The final piece of the puzzle is to account for position 6: non-initial left-peripheral demonstratives. Given that they occur **after** WL1, but still appear to be fronted, the distributional evidence suggests they occupy [Spec, FinP]. This is a cross-linguistically common place for so-called 'weak' pronouns (Cardinaletti & Starke, 1999; Roberts, 2021). I will return to the discussion of the 'weakness' of PIE demonstratives in correlative clauses, in §5.2, where I will also investigate the possible enclitic status of *REL in PIE, and its position in the left periphery outside Sanskrit.

4.4 Left-peripheral clitics in PIE

I have established by a detailed case-study that a refined model of the left periphery, employing the basic schema of Rizzi (1997), can be applied fruitfully to the question of fronted elements and clitics in Vedic Sanskrit. I now turn, in a necessarily briefer format, to three other ancient IE languages: Ancient Greek, Latin and Hittite. Each of these has been argued, under various theoretical guises, to have at least one distinct 'Topic' position, usually at the beginning of the clause.[109] Some authors argue that the same position is responsible for hosting 'Foci' (e.g., Spevak, 2006 for Latin), while others suggest a distinct 'Focus' position following 'Topic' (e.g., Salvi, 2005, also for Latin). Under the theoretical assumptions of this thesis, I will start with the premise that we should expect distinct TopP and FocP projections, though it is possible they may not co-occur (and thus be modelled as a single position). However, since I model fronted interrogatives as occupying [Spec, FocP], any evidence for Topics and interrogatives co-occurring will be taken as evidence for distinct projections. As we shall see, such orders are evidenced in Latin and

[109]See *inter alia* Dik (1995), Matić (2003), Goldstein (2016a) on Ancient Greek; Salvi (2005), Devine & Stephens (2006), Danckaert (2012), Bortolussi (2017), Halla-aho (2018) on Latin; Garrett (1994), Luraghi (1998, 2001), Samuels (2005), Hoffner & Melchert (2008), Huggard (2015), Sideltsev & Molina (2015) on Hittite.

Ancient Greek but not in Hittite since it lacks *wh*-movement in interrogatives (see §3.3.3.1). I limit my discussion to left peripheral discourse projections, and so do not consider positions lower in the clause (e.g., preverbal focus).

The primary function of this section is to establish the behaviour of clitics with regard to lexical items that have been pragmatically fronted to the left periphery. I conclude this section with a reconstruction of the PIE left periphery and the behaviour of clitics within this left periphery. I take up the specific position of *REL within this model in §5.

4.4.1 Ancient Greek

Outside Indo-Iranian, Ancient Greek is perhaps closest to Sanskrit with regard to the distribution of left-peripheral clitics. Like Sanskrit, Greek has a large inventory of clitic elements with various functions. There is a distinct class of WL1 pronominal clitics, whose distribution across persons and case varies from Vedic:[110]

	ACCUSATIVE	GENITIVE	DATIVE
1SG	με	μεο, μευ	μοι
2SG	σε	σεο, σευ	σοι
3SG	ἑ, μιν	εὑ, οὑ	οἱ
3DU	σφωε	-	σφωϊν
3PL	σφας, σφεας, σφε	σφεων	σφι(ν), σφισι(ν)

Table 4.2: Enclitic Pronouns (WL1) in Homeric Greek (adapted from Goldstein, 2016a: 6).

To this we may also add the enclitic forms of the **indefinite pronoun**:

	NOMINATIVE	ACCUSATIVE	GENITIVE	DATIVE
SG	τις, τι	τινα, τι	τευ, τεο	τινι, τῳ
PL	τινες, τινα	τινας, τινα	-	-

Table 4.3: Enclitic forms of the indefinite pronoun (WL1) in Homeric Greek.

As mentioned in §3.3.3, there seems to be some evidence from fossilised forms outside Greek that PIE *k^wi- was fronted together with WL1; however, as I will argue below, this was a late-stage innovation that did not run to completion in many branches. The evidence from Hittite is particularly crucial in this regard.

Moving on from pronouns, Ancient Greek also exhibits a set of WL2 clausal clitics, such as the following:

[110] Wackernagel (1892: 365) notes that the barytonic accent pattern of the 1pl pronouns ἥμων (us.GEN) and ἥμιν (us.DAT) suggests they were originally enclitic, but that it is hard to distinguish them from the orthotonic forms—I do not include them here.

(4.50) WL2 clitics in Ancient Greek: ἄρα, αὖ, αὖτε, δέ, δή, γάρ, θην, μάλα, μάν, μέν, μέντοι, μήν, νυ, νυν, οὖν, περ, πω, ῥα, ταρ, τε, τοι, τοίνυν

Many of these are discourse particles, but those which are clearly conjunctions of some sort, and therefore most comparable to the WL2 in Vedic, include δέ ('and', 'but'), γάρ ('for', 'because'), τε ('and'). I do not discuss here the modal particle ἄν/κε(ν),[111] whose grammatical function is clearly innovative, and whose cognates outside Greek are somewhat dubious.[112] Moreover, within Greek it has distinct grammatical functions which affect its syntactic distributions,[113] which obscure the picture further. Clearly a full account of the Ancient Greek left periphery must address this issue, but it is not of immediate comparative importance.

Let us then move onto some simple correspondences. We can model WL2 in Ancient Greek, as in Sanskrit, as being base-generated somewhere to the left of CP and moved behind the first prosodic word at PF. In fact, this was suggested for Homeric Greek before it was for Sanskrit, by Taylor (1990: 118–24):

> ...when a clitic {which} is placed in an untenable phonological position by the clitic syntax, it is "rescued" by some sort of very superficial rule (probably in the PF component) and moved to the right just far enough to satisfy the phonological requirement that it has something to lean on.

If we understand the 'very superficial rule' to be PI, Taylor's account of WL2 in Homeric Greek is essentially the same as Hale's (1996) for Sanskrit. The rule requires some modification for Classical Greek, where μέν and δέ intervene between the innovated definite article and its respective noun; as far as the Homeric corpus is concerned, however, WL2 clitics appear to follow a distribution identical to that observed in Vedic. They can therefore be accounted for by the same mechanism: namely, they are base-generated outside the left periphery and moved behind the first prosodic word (ω) at PF.

Once again, the interest lies with WL1 clitics. As in Vedic, Ancient Greek WL1 strictly follow WL2, both when they occur in clitic clusters and when they are separated by a lexical word:

(4.51)

πολλὰ	=δέ$_{WL2}$	=μιν$_{WL1}$	λιτάνευε	γέρων	ἱππηλάτα
much.ACC.PL.N	=PTC	=him.ACC.SG	entreat.IMPF.3SG	old.NOM.SG.M	horseman.NOM.SG.M
Οἰνεύς					
Oeneus.NOM.SG.M					

'And earnestly the old horseman Oeneus begged him' (Hom. *Il.* 9.580)

(4.52)

ἔγνω	=δὲ$_{WL2}$	ψυχή	=με$_{WL1}$	ποδώκεος
recognise.AOR.3SG	=PTC	spirit.NOM.SG.F	=me.ACC.SG	swift-footed.GEN.SG.M
Αἰακίδαο				
Ajax.GEN.SG.M				

'And the ghost of swift-footed Ajax recognised me' (Hom. *Od.* 11.471)

[111] I treat these as having a shared etymology per Forbes (1958).

[112] See *inter alia* Hoffman (1950: 137), Pokorny (1959: 515–6), Chantraine (1968: 507), and more recently Reece (2009: 73–4), Beekes (2010: 661). I accept in principle the etymological connection of AGk. κεν with Vedic *kam*, but I believe the grammatical functions—irrealis mood vs. emphasis/assertion respectively— are too distinct for any syntactic comparison to be fruitful from a PIE perspective.

[113] See Goldstein (2016a: 92–6; 2016b: 86–7).

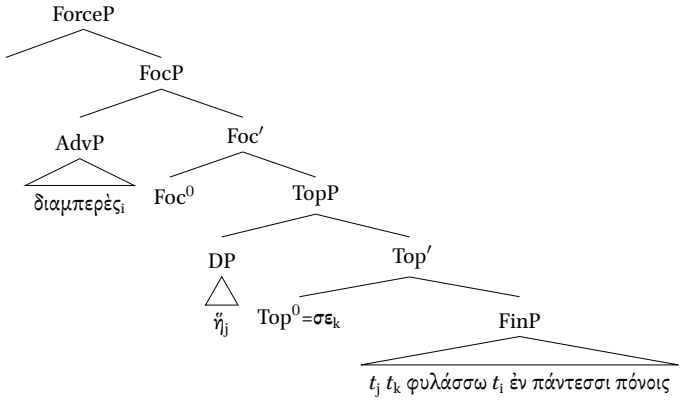
It is worth noting that constructions like ex. (4.52), where WL2 and WL1 are separated, are rare; I return to this matter below. For now, we should note that as yet again, as in Vedic, when WL1 occur later than 2P, in most cases there is a limited set of elements that can precede them. These include relative pronouns, as in (ex. 4.53), and the negator οὔ, as in (ex. 4.54).

(4.53) αὐτὰρ ἐγὼ θεός εἰμι, [$_{CP}$ διαμπερὲς ἥ =σε$_{WL1}$
but I god.NOM.SG.F be.PRES.1SG to-the-end REL.NOM.SG.F =you.ACC.SG
φυλάσσω / ἐν πάντεσσι πόνοις]
guard.PRES.1SG in all.DAT.PL.M toil.DAT.PL.M
'But I am the goddess, who will guard you to the end in all troubles' (Hom. *Od.* 20.47–8)

(4.54) ἐπεὶ οὔ =μιν$_{WL1}$ ὀίομαι οὐδὲ πεπύσθαι / λυγρῆς
since NEG =him.ACC.SG think.PRES.1SG nor learn.PERF.INF terrible.GEN.SG.F
ἀγγελίης
message.GEN.SG.F
'Since I do not think he has learned the terrible news' (Hom. *Il.* 17.641–2)

What we observe for relative pronouns is broadly similar to the patterns attested in Vedic. WL1 cannot appear preceding relative pronouns. If we model relative pronouns as occupying the lower [Spec, Top], this suggests that Greek WL1 follow the same rule as Vedic: target the **lowest Top° or Foc°**:

(4.55)



So far, so similar. But if, as in Vedic, we also model interrogatives as occupying [Spec, FocP], we should find examples corresponding to ex. (4.11) where WL1 raise to Foc° below a topicalised phrase. Topicalisation around the interrogative, however, is rare in Homer; I can find only six examples in the entirety of the *Iliad* and the *Odyssey* where an interrogative is preceded by anything other than a vocative. These are listed below:

(4.56) ἐς τί ἔτι κτείνεσθαι ἐάσετε λαὸν Ἀχαιοῖς;
to INT.ACC.SG.N yet be-killed.INF allow.FUT.2PL people.ACC.SG.M Achaean.DAT.PL.M
'How long will you allow your people to be slain by the Achaeans?'
(Hom. *Il.* 5.465)

(4.57) τῶν =δ' ἄλλων **τίς** =κεν ᾗσι φρεσὶν
DEM.GEN.SG.M =PTC other.GEN.PL.M INT.NOM.SG.M =MOD his.DAT.PL.F mind.DAT.PL.F
οὐνόματ' εἴποι / ὅσσοι =δὴ μετόπισθε μάχην
name.ACC.PL.N say.AOR.OPT.3SG as-many.NOM.PL.M =PTC from-behind battle.ACC.SG.F
ἤγειραν Ἀχαιῶν;
drive.AOR.3PL Achaean.GEN.PL.M
'Out of the others, **who**, from his own mind, could name everyone who came after and roused the Achaeans to battle?' (Hom. *Il.* 17.260–1)

(4.58) ἀνδρῶν =δ' ἐν πολλῷ ὁμάδῳ **πῶς** =κέν =τις
man.GEN.PL.M =PTC in much.DAT.SG.M din.DAT.SG.M how =MOD =INDEF.NOM.SG.M
ἀκούσαι / ἢ εἴποι;
hear.AOR.OPT.3SG or speak.AOR.OPT.3SG
'Among much din of men, **how** could anyone listen or speak?' (Hom. *Il.* 19.81)

(4.59) νῦν =δὲ **τίς** ὧδ' ἤγειρε;
now =PTC INT.NOM.SG.M thus bring.AOR.3SG
'Now **who** has called us together thus?' (Hom. *Od.* 2.28)

(4.60) αὐτὰρ ἐμοὶ **τί** τόδ' ἦδος, ἐπεὶ πόλεμον
but me.DAT.SG INT.NOM.SG.N this.ACC.SG.N pleasure.NOM.SG.N since war.ACC.SG.M
τολύπευσα;
wind.AOR.1SG
'But for me, **what** pleasure will there be in this, that I wound up the war?' (Hom. *Od.* 24.95)

Of these, we can discount ex. (4.56) immediately as ἐς ('to') can be understood as a preposition governing τί ('what') that has been pied-piped to the left periphery.[114] As for the others, although they suggest the possibility of topicalisation around the interrogative—most convincingly in the case of ex. (4.60)—they do not give unambiguous evidence for the positioning of WL1. Exx. (4.59–60) contain no clitic pronouns. Ex. (4.58) may seem the most promising since we see enclitic τις; but it is debatable as to whether we can really consider this an example of topicalisation around the interrogative at all. The PP [ἀνδρῶν δ' ἐν πολλῷ ὁμάδῳ] ('among much shouting of men') is best treated as a **circumstantial adverbial** in the sense of Cinque (1999: 28–30). Following Benincà & Poletto (2004), it may be the case that such adverbials occupy a distinct "scene setting" projection in the left periphery,[115] above [Spec, FocP]; it may equally be the case that they are not raised, but are base-generated there.[116] I take no strong view on the matter here, but note that either way this example alone is not strong evidence that Ancient Greek allows topicalisation around the interrogative in the same way as Vedic. Nevertheless, taken together with the other examples of topicalisation around an interrogative, the rule that WL1 targets the lowest active Top°/Foc° seems to make the correct predictions.[117]

[114] Cf. English [***To whom***]$_{PP}$ *did you give the money?*.

[115] Which Haegeman (2012: 89) identifies with the Modifier Phrase (ModP) proposed by Rizzi (2001).

[116] So Haegeman (2012: 101–2).

[117] On the other hand, if we allow ourselves to look beyond Homer to later texts, we find exactly the Vedic pattern attested in 5th century BCE Attic verse (Goldstein, 2016a: 121).

Thus the Homeric grammar appears to allow a somewhat limited version of the multiple fronting attested in Vedic. Nevertheless, the rules for clitic placement are so far the same: namely, WL2 routinely follow the first prosodic word while WL1 target the lowest active Top°/Foc°, at least in relative and interrogative clauses. A key difference is the behaviour of WL1 concerning the negator οὐ. As noted in ex. (4.54), repeated in ex. (4.61) below, WL1 can follow οὐ even when the latter is not in clause-initial position. However, it is also possible for WL1 to **precede** non-initial οὐ, as in ex. (4.62):

(4.61) ἐπεὶ οὔ =μιν$_{WL1}$ ὀίομαι οὐδὲ πεπύσθαι / λυγρῆς
since NEG =him.ACC.SG think.PRES.1SG nor learn.PERF.INF terrible.GEN.SG.F
ἀγγελίης
message.GEN.SG.F
'Since I do not think he has learned the terrible news' (Hom. *Il.* 17.641–2)

(4.62) οἵ =τέ =μιν$_{WL1}$ οὐκ εἰῶσι βοῶν ἐκ πῖαρ
REL.NOM.PL.M =and =him.ACC.SG NEG allow.PRES.3PL cow.GEN.PL.M from fat.ACC.SG.N
ἑλέσθαι
sieze.AOR.INF
'[Those] who do not allow him to seize the fattest of the cows...' (Hom. *Il.* 11.550, 17.659)

Taylor (1990) explains this by arguing that WL1 **optionally** occupy 'first position' within IP, but that this word order is only attested when something precedes IP (a complementiser, a *wh*-pronoun or a fronted topic). If there is nothing to the left of IP, PI will prevent WL1 being spelled out in initial position. Thus for ex. (4.62), the relative pronoun οἵ ('who') occupies [Spec, TopP] and the WL1 clitic μιν ('him') occupies 'first position' within IP; in ex. (4.61) it occupies 'second position', while οὐ occupies 'first position'. Under my model, in which WL1 have strictly one acceptable 'landing site' (the lowest active Top°/Foc°), this optionality must be captured another way.

A relatively straightforward solution presents itself if we assume that οὐ is base-generated in a functional projection, NegP, immediately dominating IP (Chatzopoulou, 2018: 63–4).[118] This predicts that, unless anything else is fronted to the left periphery, οὐ will routinely occupy clause-initial position, which is the case in Homeric Greek.[119] Trivially, in these structures WL1 will raise to C° and then be moved by PI to follow οὐ:

(4.63) οὔ =σε ἔολπα κακὸν καὶ ἄναλκιν ἔσεσθαι
NEG =you.ACC.SG.M expect.PERF.1SG bad.ACC.SG.M and feeble.ACC.SG.M be.FUT.INF
'I do not expect you will prove to be evil or feeble' (Hom. *Od.* 3.375)

If, however, something is moved to either [Spec, TopP] or [Spec, FocP]—above NegP—WL1 will surface above οὐ, attached to the relevant Top° or Foc°. This generates the order attested in ex. (4.62),

[118] See Willmott (2013) for a more precise analysis of the relative positions of οὐ and μή following Cinque's (1999; 2004) model of the IP. For our purposes it is sufficient to note that οὐ surfaces at its left-most edge, immediately following C°. It is likely, per Chatzopoulou (2018: 58–64) and Gianollo (2021: 11), that οὐ is itself a phrasal category rather than a head.

[119] Out of the 248 instances of οὐ acting as a sentential negator in books 1–12 of the Iliad, Moorhouse (1959: 89) counts 175 (71%) instances of clause-intial οὐ, and a further 59 (24%) of οὐ in second position. In the remaining 14 tokens (5%), οὐ is found immediately preceding the verb.

where WL1 precedes οὐ. Where WL1 **follow** non-initial οὐ, as in ex. (4.61), we must assume that οὐ itself has been raised to the left periphery, possibly to [Spec, FocP]. In this scenario, the lowest active left peripheral head would be Foc°, and this is where the WL1 are found. We know that negative focus preposing is possible in later Greek; Goldstein argues that it has the pragmatic effect of removing 'contextual restrictions on their interpretation' (2016a: 196–200). Further studies may seek to establish whether this interpretative effect can be meaningfully diagnosed in the Homeric data, but it seems clear that it is syntactically acceptable. A further set of apparent 'exceptions' to Wackernagel's Law present themselves in Greek in a way that is less visible than in Sanskrit, namely that enclitics sometimes follow complementisers directly, and at other times follow the first prosodic word **after** the complementiser or conjunction:

(4.64) ὅτε =μοι μένος ἄσχετος ἤσθιε
when =me.DAT.SG might.ACC.SG.N uncontrollable.NOM.SG.M eat.IMPF.3SG
Κύκλωψ / ἰφθίμους ἑτάρους
Cylcops.NOM.SG.M strong.ACC.PL.M companions.ACC.PL.M
'When the Cylcops, resistless in might, ate my strong comrades' (Hom. *Od.* 20.19–20)

(4.65) ἀλλ' ὅτε =δὴ δεκάτη =μοι ἐπήλυθε νὺξ
but when =PTC tenth.NOM.SG.F =me.DAT.SG come-on.AOR.3SG night.NOM.SG.F
ἐρεβεννή
dark.NOM.SG.F
'But when the tenth dark night came upon me...' (Hom. *Il.* 9.474)

These examples are straightforwardly accounted for under the model I have proposed. When WL1 attach directly to the complementiser, this is usually because the left periphery is inactive. As such, we model the complementiser as base-generated in C°: the model predicts that WL1 will then raise to C° and be spelled out immediately after the complementiser itself. When the Topic-Focus complex is activated, we model the complementiser as base-generated in Force°—the very top of the left periphery—and the prosodic host as occupying some lower left peripheral projection, either [Spec, TopP] or [Spec, FocP]. WL1 then raise to the lowest active Top°/Foc°, and so the complementiser appears not to 'count' for second position. In this regard, the only difference between Ancient Greek and Vedic is that the former has innovated overt complementisers, while this category is only nascent in the latter (see §3.2.1).

Thus, the model of the left periphery I have developed on the basis of Vedic seems by-and-large to account for the attested patterns in Homer. However, there are patterns in the Vedic data that are unattested or appear extremely rarely in Homer. Some of these we might predict: I cannot find an example of the order [TopP – FocP – TopP – WL1] in the Homeric corpus. Yet such sentences are hardly common in Vedic either, and rely on the hypothesis that P-words occupy [Spec, FocP], a hypothesis which requires more robust examination for Ancient Greek than space allows. A more pressing issue for our purposes is that in Homer, WL1 rarely follow a syntactic constituent of two or more prosodic words. Goldstein (2016a: 91) quotes the following:

(4.66) τοὺς =δ' ἄλλους =μοι ἔτικτον ἐνὶ μεγάροισι
DEM.ACC.PL =PTC other.ACC.PL.M =me.DAT.SG birth.AOR.3PL in palace.DAT.PL.N
γυναῖκες
woman.NOM.PL.F
'The others, women bore **for me** in the palaces' (Hom. *Il.* 24.497)

There are other examples of this construction, generally involving the demonstrative pronoun ὁ, ἡ, τό, possibly behaving in a fashion more similar to its innovative use as a definite article:[120]

(4.67) τῶν πάντων =οἱ ἕκαστος ὄϊν δώσουσι
DEM.GEN.PL.M all.GEN.PL.M =him.DAT.SG each.NOM.SG.M sheep.ACC.SG.M give.FUT.3PL
μέλαιναν / θῆλυν ὑπόρρηνον
black.ACC.SG.M female.ACC.SG.M with-suckling-lamb.ACC.SG.M
'Each of all these men will give him a black ewe with a lamb underneath' (Hom. *Il.* 215–6)

(4.68) τῶι =δ' ἑτέρωι =μιν πῆχυν ἐπιγράβδην
DEM.DAT.SG.M =PTC other.DAT.SG.M =him.ACC.SG forearm.ACC.SG.M grazing-the-surface
βάλε χειρός / δεξιτερῆς
throw.AOR.3SG hand.GEN.SG.F right.GEN.SG.F
'And with the other [spear] he struck his right forearm with a grazing blow' (Hom. *Il.* 21.166–7)

As Goldstein (2016b: 80–2) notes, there are no fronted participial phrases (of the sort found in Vedic in ex. 4.36) followed by WL1 in the Homeric corpus. Additionally, I cannot find any examples of the type attested in Vedic in ex. (4.37) where a DP containing an adjective and a noun is fronted, followed by WL1. There are, however, a number of instances of WL1 **interrupting** complex syntactic constituents at the front of the clause, as in the following (Taylor, 1990: 114):

(4.69) πολλοὶ =δέ =μιν ἄνδρες ἴσασιν
many.NOM.PL.M =PTC =him.ACC.SG.M man.NOM.PL.M know.PERF.3PL
'And many men know **him**' (Hom. *Il.* 6.151)

(4.70) χόλος =δέ =μιν ἄγριος ἧρει
anger.NOM.SG.M =PTC =him.ACC.SG great.NOM.SG.M sieze.IMPF.3SG
'And great anger seized **him**' (Hom. *Il.* 4.23)

Recall that these structures occur in Vedic:

(4.71) ***devéna*** *=naḥ* ***mánasā*** *deva* *soma* *rāyáḥ*
divine.INS.SG.N =us.DAT mind.INS.SG.N divine.VOC.SG soma.VOC.SG wealth.GEN.SG
bhāgám *sahasāvan* *abhí* *yudhya*
share.ACC.SG mighty.VOC.SG upon fight.IMPV.2SG
'**With your divine mind**, o divine Soma, o mighty one, fight for a share of the wealth for us'
(RV 1.91.23a)

[120] Cf. Manolessou & Horrocks (2007: 228, emphasis mine) 'the optional use of [ὁ, ἡ, τό] with common nouns in Homeric Greek is in fact analogous to the role of the article with proper names and generics in later Greek...namely that it serves a pragmatic role in **re-topicalising** a constituent or in facilitating referent tracking.' It is of note that the authors suggest the interpretative function of '(re)topicalisation' here, which would corroborate the movement to [Spec, TopP] proposed in my model.

Hale's argument is that these structures are diagnostic of the fact that *devéna mánasā* has not been fronted to [Spec, TopP], but remains somewhere within IP. As such, enclitic *naḥ*, having moved to C°, is pushed behind *devéna* by PI. On the other hand, when WL1 follow a whole constituent without interrupting it, this is because the constituent has moved to [Spec, TopP] and WL1 surface in C°. Although the latter construction is exceedingly rare—potentially ungrammatical—in Homer, we have good reason to believe that Hale's account of the former (i.e., WL1 interrupting a constituent NP) applies to Ancient Greek as well. Taylor (1990: 117) notes 21 instances where a clause-initial NP is interrupted by WL1 in the first 12 books of the *Iliad*. Remarkably, in 20/21 cases, that NP is the subject of the clause, as in exx. (4.69–70) above. Taylor's (1990) conclusion, with which I concur, is that these NPs have not been topicalised (or scrambled), but straightforwardly occupy the canonical subject position, probably [Spec, IP]. Thus, if WL1 raise to C°, they will be moved by PI yielding the attested word order.

The question remains, however: why do we not see more instances where WL1 follows a complex syntactic constituent, as in exx. (4.66–8)? I believe a full answer to this question extends beyond the scope of this thesis, but I offer some speculation. The rarity of this construction is not a feature of clitics, but a feature of the way Topic and Focus seem to operate in Homer. Namely, **complex NPs are not fronted in their entirety**. Rather, only part of the NP, either the adjective or the noun, is topicalised; the other part remains within IP.[121]

(4.72) [$_{\text{TopP}}$ θεῖός$_{\text{i}}$] =μοι ἐνύπνιον ἦλθεν [$_{\text{NP}}$ t_{i} ὄνειρος]
divine.NOM.SG.M =me.DAT.SG in-sleep come.AOR.3SG dream.NOM.SG.M
'A **divine dream** came to me as I slept' (Hom. *Od.* 14.495)

This pattern of partial fronting of complex constituents is attested in RV too:

(4.73) [$_{\text{TopP}}$ ***ubháyam***$_{\text{i}}$] *=te* *ná* *kṣīyate* [$_{\text{NP}}$ t_{i} ***vasavyám***]
both-kinds.NOM.SG.N =you.GEN.SG NEG diminish.PRES.3SG goods.NOM.SG.N
'Your **wealth of both kinds** is not diminished' (RV 2.9.5a)

I limit my speculation to the syntactic pattern observed here, and I do not offer any comment as to the possibility of different interpretive effects of partial vs. full fronting, nor the role meter may play in the differing patterns. Interestingly, the fronting of complex constituents is clearly attested in the 5$^{\text{th}}$ century prose of Herodotus (Goldstein, 2016a,b); it may be the case, as Goldstein argues, that it was possible during the Homeric period, just not attested. There is certainly scope for further research in this area.

To summarise: the patterns of the Ancient Greek periphery appear to be a slightly more limited set of those attested in Vedic. WL1 and WL2 follow the same strict ordering constraints. Multiple fronting, in which both [Spec, TopP] and [Spec, FocP] are filled is marginally possible. The fronting of full XPs to [Spec, TopP] is vanishingly rare, but partial fronting of the sort also attested in Vedic is acceptable.

[121] Unlike Taylor (1990), I do not take a strong view on the internal structure of the Ancient Greek DP here. My annotations therefore do not reflect her hypothesis that in cases such as ex. (4.72) the N° ὄνειρος has "escaped" its phrase via [Spec, DP], leaving the remnant DP (containing only θεῖος) to be fronted. See Taylor (1990: 82–112) for more detail on this topic; it is almost certainly the case, given the persistent verb-medial pattern in these clauses, that there are multiple instances of movement implicated in these constructions.

I return to the matter of the placement of relative and correlative demonstrative pronouns in the left periphery in §5.2 and §5.3 respectively.

4.4.2 Latin

The diagnosis of clisis is slightly more complicated in Latin than in either Sanskrit or Ancient Greek. First, we have no direct orthographic evidence for the presence or lack of a phonological accent. This means that we are largely reduced to distributional patterns (i.e., diagnostics of syntactic clisis) in order to contextualise the behaviour of Latin in this discussion. The notable exception to this are the inherited WL2 elements, e.g. *-que* ('and'), *-ve* ('or'), with the potential addition of the Q-particle *-ne*. Beyond these, other examples of 'strict' WL2 clitics are less forthcoming. Traditional examples may be *enim* ('for'), *autem* ('but'), and *vero* ('indeed'). There are, however, instances of clause-initial *enim* in old comedy (Spevak, 2012: 336):[122]

(4.74) ***enim*** *me* *nominat*
yet me.ACC.SG name.PRES.3SG
'**Yet** he calls me by name' (Plaut. *Trin.* 1134)

(4.75) ***enim*** *non* *sinam*
of-course NEG allow.FUT.1SG
'**Of course**, I won't allow it' (Ter. *Ad.* 168)

Clause-initial instances of *autem* are not directly attested, but are discussed as archaic possibilities by ancient grammarians (Spevak, 2012: 336). On the basis of this evidence, Spevak (2012: 337) suggests that these forms were not inherited as enclitic, but were grammaticalised as such within the history of Latin. Within Classical Latin, however, they frequently appear following the first prosodic word (Spevak, 2006). This suggests that they can be accounted for by the WL2 mechanism posited for Sanskrit and Ancient Greek, i.e. base-generation in first position + PI. Although the behaviour of *enim*, *autem* etc. may have developed within the history Latin, that a mechanism of this sort was inherited from PIE is confirmed by *-que* and *-ve*, which are enclitic at every stage of the language and have clear cognates elsewhere in IE.

Setting aside WL2, the status of WL1 in Latin is yet harder to diagnose. Aside from the aforementioned lack of orthographic accents, the 'strong' (i.e., accented, topicalised/focalised) and 'weak' (i.e., unaccented, pragmatically unmarked) forms of the personal pronouns are formally identical. Wackernagel (1892: 406) reassures us that this does not matter, because, discounting the clause-initial 'strong' personal pronouns, the Latin pronouns exhibit the same syntactic distribution as the unambiguously enclitic forms of the personal pronouns (WL1) in Ancient Greek. In other words, we may use the distributional properties of the personal pronouns as evidence for their status as clitics. As such, where a personal pronoun occurs in initial position, we may deduce that this is its lexically accented, non-clitic form:

[122] Such instances are referred to as the *enim affirmativum*, which lacks the causal sense of enclitic *enim*.

(4.76) ***mihi*** *istunc* *vellem* *hominem* *dari* *ut* *ego*
me.DAT.SG DEM.ACC.SG.M wish.IMPF.SUBJ.1SG man.ACC.SG.M be-given.INF that I
illum *vorsarem!*
DEM.ACC.SG.M twist.IMPF.SUBJ.1SG
'I wish that man were given to **me**, so I could be the one to twist him!' (Plaut. *Cis.* 92–3)

On the other hand, where the personal pronouns are 'weak' and appear in 2P, the fact that their syntactic distribution seems to overlap to a significant extent with the unambiguously enclitic forms attested in Ancient Greek and Sanskrit suggests some form of clisis is at play. And indeed, as in Sanskrit and Ancient Greek, there are a significant number of "exceptions" to Wackernagel's law, i.e., where 'weak' pronouns occur later than 2P.

A cogent and influential explanation for these "exceptions" in Latin is found in Adams (1994), whose argument I summarise here. In the first place, he endorses the notion expounded by Fraenkel (1964) that we should analyse Latin sentences in terms of prosodic units called *cola*, broadly equivalant to the phonological phrase (φ). This allows for 2P in purely prosodic terms to be defined with respect to a boundary other than the clausal one. For example (Adams, 1994: 106):

(4.77) *de* *triumpho* *=autem* // *nulla* *=**me*** *cupiditas*
about triumph.ABL.SG.M =but none.NOM.SG.F =me.ACC.SG desire.NOM.SG.F
umquam *tenuit*
ever hold.PERF.3SG
'But as far as the triumph is concerned, no desire ever held **me**' (Cic. *Att.* 7.2.6)

In this example, despite being far from 2P with respect to the clause boundary, enclitic *me* occurs in second position within the *colon* (marked with //). Moreover, the *colon* boundary is not a purely phonological construct; indeed, concerning the grammatical status of *colon* boundaries, Adams (1994: 106) notes that:

> *Cola* might consist, for example, of clause-equivalent participial constructions such as the ablative absolute, prepositional expressions, extended subjects and objects, and various types of headings which are detached from the main body of the clause.

In Minimalist terms, therefore, Adams' syntactic description of *cola* may be rephrased to include topicalised or left-dislocated phrases, as well as parenthetic or extra-syntactic interjections. While *cola* are strictly conceived of as prosodic units, separated by intonational breaks, they overlap to a large extent with syntactic boundaries including, crucially, the left periphery vs. the main clause. This is the case for ex. (4.77) above, to which we **could** ascribe the following structure:[123]

(4.78) [$_{\text{TopP}}$ *de triumpho =autem*] [$_{\text{IP}}$ *nulla =**me** cupditas umquam tenuit*]

Adams (1994) notes that dividing sentences into such *cola* may explain already explain a number of apparent "exceptions" to Wackernagel's Law. Yet even taking *colon* boundaries into account, Adams

[123] As we shall see below, it is likely that *nulla* actually occupies [Spec, FocP] here.

(1994: 110) notes that 'one is left with masses of unemphatic (unstressed?) pronouns which, on a reasonable *colon* division, are not placed second in their *colon*, but later.'[124]

To account for such exceptions, Adams hypothesises that WL1 attach to a set of **preferential hosts**, most of which correspond to elements that are **focalised**. These preferential hosts include 'antithetical terms, dimensional terms, including intensifiers, demonstratives, including temporal adverbs, the relative *qui*, interrogatives, negatives and imperatival verbs' (Adams, 1994: 154–5). Each of these terms, Adams' argues, are focalised, and thus "attract" WL1—in our model, where this focalised item is in the left periphery, we model it as occupying [Spec, FocP]. This can account for the observed 2P effects, according to Adams, because focalised constituents very often occur in first position within their respective *cola*. Adams' model for "Wackernagel's law" in Latin therefore bears a striking resemblance to the model I have posited so far for Vedic and Ancient Greek: WL1 target a functional head (in this case, Foc^{o}); the constituent occupying the specifier of that functional head (in this case, [Spec, FocP]) is very often but not **necessarily** in first position.

Accordingly, I am inclined to accept Adams' account for WL1 placement in Latin, though some adjustments must be made. The first point on which I differ from Adams' account is whether relative pronouns are focalised. So far I have suggested that fronted relative pronouns occupy [Spec, TopP]; in Vedic and Ancient Greek, I was able to establish that this was specifically the lower [Spec, TopP] (i.e., lower than FocP). For this reason, I am not inclined to treat relative pronouns as focalised in Latin to account for their acting as 'preferential hosts' in Adams' terms. From an interpretive point of view, Adams (1994: 146) comes to the same conclusion:

> The relative pronoun, being anaphoric, is by definition thematic, but in Classical Latin at least **it is rarely the focus** in the sense that the various hosts of enclitics identified earlier in this paper can be ascribed that characteristic.

To account for its belonging to the set of 'preferential hosts', Adams invokes the argument that, at an earlier stage of the language, in correlatives where the RC precedes the main clause, the relative pronoun was indeed focalised. From this stage, Adams suggests it is possible that the 'anaphoric' (i.e., non-focalised) forms of the relative pronoun gained their attractive power, on par with other genuinely focalised elements. Nevertheless, he leaves the matter somewhat open, suggesting that in Classical Latin, *qui* 'can be treated as another of the elements which might attract clitic pronouns (and the copula), though its ability to do so might historically have a **different origin** from that same ability as manifested, say, by antithetical terms.' (Adams, 1994: 146–7, emphasis mine). How would this 'diffence in origin' be encoded in my model?

Let us start with the assumption that Latin inherited the same rule of WL1 placement as we see in Sanskrit and Ancient Greek (i.e., WL1 obligatory raise to C^{o}; or to the lowest active Top^{o}/Foc^{o}) with the same prosodic repair strategy (viz. PI) to prevent their occurring clause-initially. At this stage, the propensity of relative pronouns to host WL1 is explained by the fact that they regularly move to (the lower) [Spec, TopP], and are never followed by any other left peripheral elements. On the other hand,

[124] For a more quantitative approach to these exceptions, see Spevak (2006: 269–73); I have not included her data here only because it does not use the *colon* division espoused by Adams (1994).

antithetical terms occupying [Spec, FocP]—absent any relative pronouns—would round off the left periphery. Thus the 'difference in origin' is encoded in the fact that relative pronouns attracted WL1 because they occupy [Spec, TopP], whereas antithetical terms and other 'preferential hosts' occupy [Spec, FocP]. Yet in another sense, their ability to host WL1 does share an origin insofar as the original rule would originally have targeted Top° and Foc° indiscriminately, attaching to the lowest one.

Setting aside relative pronouns, however, in order to derive the situation suggested by Adams, we must hypothesise a change where WL1 was reanalysed as primarily targeting Foc° (where present), regardless of whether it was the lowest active left peripheral head or not. This reanalysis is quite plausible in light of the fact that, with the exception of relative pronouns, topicalised phrases regularly occur in the **higher** [Spec, TopP]. Thus, if something occupies [Spec, FocP], it will very often be the last element in the left periphery and so, incidentally, will act as the host for WL1. In fact, in his conclusion Adams (1994: 172) himself suggests a reanalysis along these lines, where the more general "mechanical" rule of Wackernagel's Law, because of its coinciding with focalised phrases, led to WL1 themselves being reanalysed as focus markers.

With no further changes to the grammar, this would make the prediction that, on the rare occasion that we do see a focalised phrase in [Spec, FocP] co-occurring with a relative pronoun (occupying the **lower** [Spec, TopP]), WL1 should raise above the lower Top°—an illicit movement in the earlier model—and attach directly to Foc°. There may be some instances of this very ordering in Plautus:

(4.79) *satis iam audivi tuas aerumnas, ad portum*
enough now hear.PERF.1SG your.ACC.PL.F tribulation.ACC.PL.F at harbour.ACC.SG.M
*=**mihi** quas memorasti*
=me.DAT.SG REL.ACC.PL.F remind.PERF.2SG
'I've heard enough of your tribulations, which you told **me** about at the harbour'
(Plaut. *Capt.* 929)

(4.80) *quin =**tibi** qui bene volunt, bene vis item?*
why-not =you.DAT.SG REL.NOM.PL.M well wish.PRES.3PL well wish.PRES.2SG likewise
'Why not, for those who wish **you** well, wish them well in return?' (Plaut. *Poen.* 165)

(4.81) *pallam ad phrygionem cum corona ebrius ferebat,*
mantle.ACC.SG.F to embroider.ACC.SG.M with crown.ABL.SG.F drunkenly carry.IMPF.3SG
*hodie =**tibi** quam surrupuit domo*
today =you.DAT.SG REL.ACC.SG.F take.PERF.3SG house.ABL.SG.F
'With a garland, drunk, he was carrying the mantle to the embroider, which he took from **your** house today' (Plaut. *Men.* 563–4)

It must be conceded that the evidence on this front is not particularly strong. In the first place, we could explain away each of these examples by suggesting the pronoun here is not enclitic, but itself fronted to a position in the left periphery. Moreover, there is little in the way of interpretive function that should suggest the fronted phrase *ad portum* is focalised; *quin* and *hodie*, on the other hand, the former being an interrogative and the latter being being a temporal adverb, are contained within Adams' set of 'preferential hosts'. In any event, if we accept at the very least that the pronouns are enclitic, the fact that

they occur **preceding** a fronted relative pronoun, attaching to a phrase higher in the clause, suggests that Latin at this stage may allow for a pattern that is infelicitous in Vedic and Homer Greek, whereby WL1 raise above the relative pronoun. The Plautine structures above could therefore represent a transitional stage in the status of "Wackernagel's Law", where there is still regularly movement of WL1 to the left periphery, but the constraint on WL1 targeting the lowest active Top°/Foc° has been lost—rather, where Foc° is present, it is the most preferred of the 'preferential hosts'. The final stage is that elucidated in Adams (1994), where the focalised phrase itself need not appear in the left periphery:[125]

(4.82) *quoniam neque consili tui neque consolationis*
since neither advice.GEN.SG.N your.GEN.SG.N nor consolation.GEN.SG.F
*cuiusquam spes ULLA =**mihi** ostenditur*
any.GEN.SG.F hope.NOM.SG.F any.NOM.SG.F =me.DAT.SG is-offered.PRES.3SG
'Since not ANY hope is offered **to me** either of your advice or of any consolation whatsoever'
(Cic. *Att.* 11.25.1)

The question remains, however: in the absence of a focalised phrase, why do relative pronouns continue to act as a 'preferential host' for WL1 clitics? A possible answer to this question would be to argue that relative pronouns are somehow lexically specified as 'attracting WL1'; in other words, their ability to host clitics becomes a part of their feature specification rather than their role in the narrow syntax. Such a change is not entirely outlandish. The sequence *si quis...* ('if anyone'), was presumably generated in the first place by some sort of movement process (enclitic indefinite *quis* raises to C°); yet long after the demise of independent indefinite *quis* in Latin, the sequence *siquis* survives as a fossil, bearing witness to the earlier syntactic rule but ultimately modelled as a peculiar behaviour of the conjunction *si* ('if'), and a few other functional items. It is not unreasonable, therefore, that the same be true of relative *qui*. Setting aside relative pronouns, there does not seem to be evidence for Top° attracting WL1; Adams' criteria quite clearly point to Foc°.

Another syntactic change is at play which complicates the matter: the loss of fronting around the relative pronoun. Salvi (2005: 453) suggests that in Classical Latin, in contrast to earlier stages of the language, 'relative *wh*-phrases are always the first constituent in embedded clauses'.[126] He uses this to argue that the position of relative pronouns in Classical Latin is the same as that suggested by Rizzi (1997: 298) for Italian: [Spec, ForceP]. If this is indeed the case, and relative pronouns in Latin do not occupy [Spec, TopP], it is perhaps less surprising that WL1 are regularly hosted by relative pronouns but not by fronted topics. This matter merits further attention for the purposes of understanding the innovative patterns of Classical Latin.

A residual issue, which I will not address in detail here, is the fact that enclitic pronouns very commonly split the focalised constituent that hosts them, as in the following (Adams, 1994: 134–5):

[125] I do not address here the question of whether a non-left-peripheral focus position, directly preceding the verb (i.e., Foc*v*P, see Belletti, 2001, 2004) existed in PIE; it is quite possible that they did, and such an analysis has been explicitly adopted in analyses of Latin such as Devine & Stephens (2006) and Danckaert (2012).

[126] Salvi's hypothesis is based only on prose, and does not hold for Classical Latin poetry, where non-initial *qui* is attested. This weakens the argument that *qui* obligatorily occupies [Spec, ForceP].

(4.83) *equitatum* *in* *Italiam* *quam primum* *mitterent;*
cavalry.ACC.SG.M into Italy.ACC.SG.F as-soon-as-possible send.IMPF.SUBJ.3PL
[*pedestris* ***=sibi*** *copias*] *non* *defuturas*
infantry.ACC.PL.F =them.DAT.PL forces.ACC.PL.F NEG be-lacking.FUT.PPL.ACC.SG.F
'They should send the cavalry into Italy as soon as possible; [infantry forces] would not be lacking **to them**' (Cic. *Cat.* 3.9)

(4.84) [*tantam* ***=me*** *inimicorum* *multitudinem*] *suscepisse*
so-great.ACC.SG.F =me.ACC.SF enemy.GEN.PL.M number.ACC.SG.F take-up.PERF.INF
uideo
see.PRES.1SG
'I see that **I** have made [so great a number of enemies]' (Cic. *Cat.* 4.20)

Constituents split in this way have so far been explained in two ways: (a) with recourse to PI, assuming the split constituent has not been fronted at all, or (b) partial fronting. Neither of these seems appropriate here: Adams' overarching analysis, which I accept, hinges on the very fact that these constituents are focalised, which directly contradicts (a). Option (b) is almost certainly possible, and could be invoked to account for instances of long-distance split-constituents such as *tantam...multitudinem* in ex. (4.84) above, but seems unsuited to capture examples such as (4.83) in which the constituent is **only** interrupted by WL1. On the other hand, Adams (1994: 132–41) notes that where WL1 interrupt a complex constituent, it is most commonly the 'preferential host' (i.e., focalised/antithetical adjective or pronoun) that hosts the clitics, with the rest of the constituent following. This could support a partial-fronting (or, at least, partial-focalisation) analysis. More speculatively, if such focalised phrases were marked by a particular intonation, this pattern could be accounted with recourse only to prosody (i.e., in a prosody-dominant model of clisis in Latin).

Setting aside the innovative patterns of the Classical period, Latin presents a situation that can be derived from the model posited for Sanskrit and Ancient Greek by positing a reanalysis whereby WL1 target Foc° specifically, with the additional, perhaps lexically specified exception of the relative pronoun. Where Foc° competes with a relative pronoun, it is possible, at least, for Foc° to win the attraction of WL1, even when a relative pronoun occupies the lower [Spec, TopP]. These rules clearly generate surface orders that are very similar to those observed in Sanskrit and Greek (hence Wackernagel's original observations) but are accounted for synchronically by a different mechanism.

4.4.3 Hittite

Clitic behaviour in Hittite presents quite a different picture of the left periphery to that displayed in Sanskrit, Ancient Greek and Latin (hence forth 'Nuclear Indo-European', NIE), and has received much interest in the last decade (Huggard, 2011, 2015; Goedegebuure, 2014; Kloekhorst, 2014; Sideltsev & Molina, 2015; Sideltsev, 2017). The basic situation in is as follows. Hittite contains a class of elements that we may call 'conventional' clitics. These routinely occur following either (a) a clause-initial, proclitic sentence connective *nu*, *ša*, *ta*, or (b) following the first lexically accented word in a sentence. When they co-occur, they follow a strict ordering constraint as enumerated in the table below:

Host	1	2	3
Accented Word (+ *-a/-ya* 'and') or (+ *-a/-ma* 'but')	Quotative particle: *-wa(r)*	*-naš* (PRO.DAT-ACC.1PL) *-šmaš* (PRO.DAT.3PL)	*-a-* (PRO.NOM.3SG/ACC.3SG) e.g., *-aš, -an -at, -e, -uš*
sentence connectives: *nu, šu, ta*			

4	5	6
-mu (PRO.DAT-ACC.1SG) *-ta/-du* (PRO.DAT-ACC.2SG) *-še/i* (PRO.DAT-ACC.3SG)	*-za* (REFL)	Local particles: *-an, -apa -ašta, -kan -šan*

Table 4.4: The Hittite Clitic Chain, adapted from Hoffner & Melchert (2008: 410).

To explain briefly: the host of the clitic chain may either be **one** of the grammaticalised, proclitic sentence connectives (*nu, šu, ta*)[127] or, in their absence, the first accented word of the clause. If and only if the clitic chain is hosted by a lexically accented word, the first member of the clitic chain will be the adversative enclitic *-(m)a* or coordinative *-(y)a* if present (on which more below). Absent a *-(y)a* or *-(m)a*, position 1 within the chain is reserved for the quotative particle *-wa(r)* which introduces indirect speech. Positions 2–4 are reserved for various forms of the enclitic personal pronouns; Hoffner & Melchert (2008: 411) note that 'slots 2 and 4 are mutually exclusive and never co-occur.' Slot 5 hosts the reflexive particle *-za*, and slot 6 the local particles.[128]

The Hittite clitic chain clearly contains WL1 (pronominal) enclitics,[129] but it is not limited to them; functional elements such as quotative *-wa(r)* have more in common with WL2 elsewhere. As such, I will eschew these categories in the following discussion, and refer to all members of the clitics chain collectively as **WLH**(ittite).[130] As noted by Huggard (2015: 24–5), topicalised and focalised phrases consistently **follow** the sentence connectives, on which basis he concludes the connectors occupy [Spec, ForceP]. It follows from this that the clitic cluster for WLH surfaces in Force°. When [Spec, ForceP] is unoccupied (i.e., in the absence of *nu*/*šu*/*ta*), PI will push the clitic cluster behind the first lexically accented word. The two possibilities are exemplified below (Luraghi, 1998: 191):

(4.85) *n=**aš**=**mu**=**kan*** *ḫuwaiš*
CONN=he.NOM=me.DAT-ACC=LP escape.PRET.3SG
'He ran away from me' (AM 50.30)

(4.86) *piran=**ma**=**at**=**mu*** ^{mD}XXX.^{D}U-*aš* DUMU ^{m}*zida* *maniyaḫḫiškit*
before='but'=it..ACC=me.DAT-ACC Armadatta.NOM.SG child Zida adminstrate.PRET.3SG
'Before me, Armadatta, the son of Zida, had adminstrated it' (*StBoT* 24 i 28)

Alongside the collapse of WL1 and WL2 into WLH, we now have our second major departure from NIE. Accepting, for the time-being, Huggard's location of *nu* in [Spec, ForceP], it seems that pronominal

[127] Written as *n-*, *š-*, *t-* prevocalically.

[128] This category does not correspond directly to the "local particles" or "preverbs" (i.e., P-words) found elsewhere in IE, though they are likely of a shared origin; see Luraghi (2001).

[129] Notably, to the exclusion of indefinite *kuišk*i, which is nevertheless enclitic (see §3.3.3.2).

[130] I will not treat the internal ordering of this clitic chain here (as I have not treated cluster-internal orderings elsewhere). This is an unfortunate gap in my account of the Hittite left periphery, but a necessary one for reasons of space.

clitics do not attach to the functional left peripheral heads Top° or Foc°, but raise beyond them to Force°. I return to the implications of this behaviour shortly.

This fairly clear picture of left-peripheral clitics in Hittite is exponentially complicated by the behaviour of the two enclitic discourse particles: -(*y*)*a* ('and') and -(*m*)*a* ('but'). There is a fairly extensive literature on the function, both pragmatic and syntactic, of these, a full discussion of which is beyond the scope of this thesis. In short, both seem to have **at least** two distinct uses. Each can be used as effectively a conjunction, as reflected in the translations I have given above (Hoffner & Melchert, 2008: 395). However, it is well-established that they have pragmatic functions beyond clause-linking (Sideltsev, 2017: 181):

> ...-(*m*)*a* marks contrastive as well as new/shifted topic in the left periphery, and identificational focus in the preverbal position; -(*y*)*a* marks additive focus in the left periphery and scalar additive focus 'even' in the preverbal position...

Focusing therefore on their left-peripheral uses, what sets -(*m*)*a* and -(*y*)*a* apart is that their distribution differs systematically from WLH. Although -(*y*)*a* and -(*m*)*a* commonly occupy 2P—i.e., following the first prosodic word—there is a set of lexical items that are "ignored" for the purpose of calculating second position. Sideltsev (2017: 181) lists them as 'connective *nu*, subordinators *takku* 'if', *mān* 'if/when', adverb *namma* 'then', the conjunction *našma* 'or' and (less commonly) the relative pronoun *kui-*'. The members belonging to this category are 'diachronically unstable' (Sideltsev, 2017: 181, n.11). Although the details of this variability are generally beyond the scope of this discussion, it is worth noting that 'indeterminate' *kuiš* (see §3.3.3.1) at some stages does seem capable of hosting -(*y*)*a*/-(*m*)*a*, which Huggard (2015: 147–52) takes as the usual ordering, but Kloekhorst (2014: 615, n.2200) and Sideltsev & Molina (2015) treat as aberrant. To my knowledge, the ordering is not attested in Old Hittite, but this does not necessarily mean it is an innovation; indeed, Kloekhorst explains the aberrant ordering by suggesting synchronically they are univerbations, but that 'these pronouns historically seem indeed to consist of *kui-* + =(*m*)*a* and *kui-* + =(*y*)*a*' (2014: 616, n.2200). I will return to the implications of the two possibilities below.

Diachronic instability aside, with the exception of the sentence connective *nu* and the complementiser *takku*,[131] I will refer to this group collectively as **delayers**. The effect of these delayers is demonstrated in the following, where -(*m*)*a* occurs later in the sentence than WLH (Sideltsev & Molina, 2015: 3):

(4.87) *mān=**aš*** *nāwi=**ma*** *paizzi*
if=heNOM not-yet=but go.PRES.3SG
'But if he has not already gone...' (HKM 66 obv. 26)

Kloekhorst (2014: 599–617) accounts for this distribution in prosodic terms. Under his account, the delayers are categorically unaccented; as such, he argues that we can model left-peripheral -(*m*)*a* and -(*y*)*a* as following the first lexically accented word. If we were to accept Kloekhorst's phonological hypothesis concerning the status of the delayers, the syntactic account would be quite straightforward:

[131] These appear in complementary distribution, and require a different syntactic analysis from others in this class (to be discussed below).

-(*y*)*a* and -(*m*)*a*, as with WL2 elsewhere, are generated externally to the clause, i.e., in syntactic first position, and moved by PI. However, Kloekhorst's claims in this connection have received some robust criticism from Sideltsev & Molina (2015) and Sideltsev (2017). While the evidence that the sentence connective *nu* is proclitic, and hence unaccented, is quite clear, the same cannot be said of the delayers. Indeed, Sideltsev & Molina (2015: 6) go so far as to say that '[a]ll the tests that unambiguously show that *nu* is proclitic...show exactly the opposite for *mān* and other members of the class'. As such, the authors argue that a more complex syntactic approach is warranted to account for the distribution of -(*m*)*a* and -(*y*)*a*; their position cannot explained as a purely PF phenomenon. Before summarising their account (which is developed further in Sideltsev, 2017), it is worth noting that there is another set of elements that are ostensibly left-peripheral, according to Sideltsev (2017), but **follow** -(*m*)*a* and -(*y*)*a*. These include the conjunctions *maḫḫan* ('if', 'when'), *kuit* 'that', *kuwapi* ('when') and, most interestingly for the purposes of this thesis, the relative pronoun *kuiš* in 'determinate' relatives (cf. §3.3.2.1). I will revisit the last of these in §5.1.

This yields the following set of orderings for Hittite, akin to the 'initial string' of Vedic but differing substantially in its structure (Sideltsev & Molina, 2015: 41):

1. Sentence connectives: *nu*/*šu*/*ta*
2. WLH (see Table 4.4 above). In the absence of a sentence connective these will attach to the first prosodic word of the clause
3. Delayers: *kāša/kāšma* (perfecitiviser), *našma* ('or'), *namma* ('but'), *mān* ('if'), 'indeterminate' *kuiš* (REL)
4. Topicalised/Focalised XP; non-delaying subordinators (e.g., *maḫḫan* 'if'), verbs or preverbs in the absence of a fronted topic
5. ***-(m)a, -(y)a***
6. *kuiš* in 'determinate' relatives; indefinite *kuišk*i; non-delaying subordinators (in the present of a fronted topic)

Continuing with Huggard's analysis that position 1 is [Spec, ForceP], we may assume the following. Position 2 is Force°. Skipping over Position 3 for now, Position 4 looks like either [Spec, TopP] or [Spec, FocP], which would suggest that position 5 is Top°—in other words, -(*m*)*a* and -(*y*)*a* are construed as markers of Topic and Focus respectively (Samuels, 2005; Huggard, 2011; Yates, 2014).[132] Recall that Hittite lacks *wh*-movement of interrogatives, which have so far been our touchstone for diagnosing the position of FocP. Nevertheless, the evidence presented in Goedegebuure (2014: 463) seems to suggest that that we can distinguish between TopP and FocP, with the former preceding the latter (as expected):

(4.88) ^{mD}LAMMA-*aš*=***ma***$_{Top}$=*mu* *apēdani*=***ya***$_{Foc}$ *mēḫuni* *šer* *akta*
Kurunta.NOM.SG.C=**but**=me.DAT-ACC DEM.LOC.SG=**and** time.LOC.SG up die.PRET.3SG
'Even at that time Kurunta put his life on the line for me' (Bo 86/299 obv. ii 54–55)

[132] These analyses are in turn dependent on Garrett (1994); they primarily address the use of -(*m*)*a* as a topic marker and not -(*y*)*a* as a focus marker, but the analogy is clear.

In this example, -(*m*)*a* marks the (contrastive) topic, while -(*y*)*a* marks (additive) focus. Beyond this and one other putative example, however, evidence is quite scant. This is perhaps to be expected; I note once again that the co-occurrence of TopP and FocP in the left periphery is generally quite rare; without the help of interrogatives routinely occupying [Spec, FocP], the diagnosis of such a position is especially subjective. To this end Sideltsev & Molina (2015: 20–1) disregard Goedegebuure's evidence, and posit a single position for Topics and Foci. On the basis of the comparative evidence, I would rather model them as two separate projections; however, pending further investigation of the Anatolian data, I make no strong claims in this regard.

This leaves position 5, which could be a lower left-peripheral position, perhaps [Spec, FinP], or an IP-internal (i.e., non-left-peripheral) position; I do not investigate it further here.

Presently, we may return to Position 3. Given the nature of the elements that occur there, this position does not seem to have either a clear correlate in the formalism of Rizzi (1997), nor a comparative equivalent in the languages surveyed so far. Nevertheless, the elements that occupy this position—while diachronically unstable—appear to share some similarity in their interpretive functions. Recall that, according to Huggard (2015), 'indeterminate' *kuiš* is fronted to [Spec, ContrFocP], yielding its conditional semantics (see §3.3.3.1). It is not implausible that in this context, it is in a complementary distribution with the overt conditional marker *mān* ('if'). Note that when *mān* and (indefinite) *kuiš* co-occur, the latter is not fronted to the left periphery (Huggard, 2015: 132):

(4.89)	*nu=wa=mu*	*mān*	*idālun*	*memian*	***kuiš***	[*memai*]
	CONN=QUOT=me.DAT-ACC	if	evil.ACC.SG	word.ACC.SG	INDEF.NOM.SG	tell.PRES.3SG
	'If someone tells me a bad word'					(KUB 14.1+ rev.45)

One could unify the left-peripheral position of *mān* and 'indeterminate' (i.e., conditional) *kuiš* by arguing that both occupy [Spec, ContrFoc]. This could then in turn be used to account, at least partially, for Position 3 as identified by Sideltsev & Molina (2015). It must be conceded, however, that this solution does not neatly account for the other delayers identified above. One could extend the contrastive semantics of *mān* and *kuiš* to include *našma* ('or'); in fact, *našma* can also be used as a conjunction meaning 'or **if**', in which case it looks quite a lot like *mān* and 'indeterminate' *kuiš*. It is not clear, however, that the same could be said for an element such as *kašma*. It is likely, therefore, as Sideltsev & Molina (2015: 41, n. 131) suggest, that Position 3 is not a singular projection, and requires further subdivision. It is also worth noting that Huggard (2015: 147–50) uses precisely the surface position of -(*m*)*a* directly after 'indeterminate' *kuiš*—the ordering which is treated as aberrant by Kloekhorst (2014) and Sideltsev & Molina (2015)—as support for his hypothesis that *kuiš* occupies [Spec, ContrFocP]; he sees -(*m*)*a* as base-generated in ContrFoc°.[133] A more thorough review of the distributional data may shed some light on the matter, but the variability of the members of the delaying set, including *kuiš* presents some challenges to developing a more detailed theory. In any event, what is clear is that the conventional Topic-Focus complex of Rizzi (1997), which has accounted for the data so far, is insufficient to account

[133]This in turn raises the question as to whether -(*m*)*a* is a marker of Topic, as suggested by Sideltsev & Molina (2015), or Focus, as suggested by Huggard (2015). Such a discussion is beyond the scope of the current discussion, and is somewhat irrelevant if we cannot establish clearly distinct TopP/FocP projections.

for the Hittite left-peripheral patterns; rather, one must posit **at least** one more projection directly under ForceP for these delayers: I will continue to call this ContrFocP.

There is a deficiency to this model, in that -(*m*)*a* cannot consistently be modelled as occupying Top°. In its adversative use, for example, -(*m*)*a* can be enclitic on a conjunction such as *maḫḫan* (Sideltsev & Molina, 2015: 12):

(4.90) *maḫḫan*=***ma***=[*ka*]*n*? GIŠ*ḫulugganniš* *parā* *iyannai*
when=**but**=LP carriage.NOM.SG out move.PRES.3SG
'**But** when the carriage moves out...' (IBoT 1.3 obv.ii 23–4)

Thus, an alternative suggestion is put forward by Sideltsev & Molina (2015), who argue that -(*m*)*a* must occupy a higher left peripheral position; they decide on Force°. This decision is in turn dependent on their stipulation, *contra* Huggard (2015), that sentence connectives like *nu* occupy 'unconventional' left-peripheral positions that are **even higher** than [Spec, ForceP] (Sideltsev & Molina, 2015: 39). They do not expand on this suggestion, which also entails that Force° bears (presumably optional) strong features for 'contrastive topic, additive focus, contrast' (2015: 39). Thus, where -(*m*)*a* appears to mark a contrastive topic (i.e., where one might expect Top°), it actually resides in Force°. As for what occupies [Spec, ForceP], absent a topicalised/focalised constituent, Sideltsev & Molina (2015: 39–40) seem to suggest quite fleetingly that it is occupied by *wh*-interrogatives, but in their absence it is not clear what their hypothesis is for the positioning of -(*m*)*a*. There is some suggestion that it could be possible that 'the subordinator raises [from Top°] to Force to provide the host for -(*m*)*a*' (Sideltsev & Molina, 2015: 39); for our purposes, we could envisage this not as syntactic movement but as PI taking place at PF.

As far as I can tell, Sideltsev & Molina (2015) do not offer a specific hypothesis as to where the delayers go—they are presumably housed in one of the 'unconventional' projections above ForceP. I am unsure, however, as to how this captures their 'delaying' effects. If we imagine that -(*m*)*a*/-(*y*)*a* are base-generated in Force°, it should be possible for them to follow a Delayer in, say, [Spec, YP] (where Y° is an 'unconventional' pre-Force° head) directly. The only mechanism that can rule out this word order is to stipulate that -(*m*)*a*/-(*y*)*a* must be hosted by a lexically-accented word **within** ForceP.[134] Such a stipulation is not made explicit in Sideltsev & Molina (2015), but seems a necessary part of their model.

However, I am inclined to agree with Huggard (2015) that *nu* occupies [Spec, ForceP], with the WLH clitic chain in Force°. Setting aside the deviance of -(*m*)*a*/-(*y*)*a*, there is no **distributional** evidence to the contrary; as noted above, Topics and Foci **follow** *nu* + WLH. Moreover, it is worth considering the 'unconventional' pre-ForceP projections that have been suggested by authors such as Hill (2007) and Haegeman (2014).[135] Hill (2007) posits the existence of a S(peech)A(ct)P(hrase) that selects ForceP as its complement; Haegeman (2014) further divides this into a higher projection, hosting discourse markers, and a lower one hosting vocatives. Now, clearly *nu* is not a vocative; could it be considered a discourse marker in the sense of Hill (2007) and Haegeman (2014)? The answer is almost certainly no. According to Haegeman (2014), discourse marker express a speaker's **attitude**. She list four properties that all discourse markers share (Haegeman, 2014: 120):

[134] Alternatively, one could stipulate that [Spec, ForceP] is obligatorily filled.

[135] Both of whom are cited by Sideltsev & Molina (2015: 39) in support of their analysis.

(a) They are not truth-functional, and lie outside the scope of negation and tense

(b) They are 'conversational' or 'interactional', and imply "the obligatory (and largely implicit) presence of the entities involved in the **specific communicative situation (speaker and, especially, hearer**" (Munaro, 2006: 6, emphasis mine)

(c) They are 'expressive' (Kratzer, 1999) or 'illocutionary'/'interpersonal'. They may express "**the mental state of the speaker**, which can be surprise, curiosity, desire, disappointment, anger and so on" (Munaro, 2006: 7, emphasis mine)

(d) 'They are deictic...they may express a response to a linguistic event or to a non-linguistic event which is manifest in the speech situation'

With the exception of property (a), *nu*/*šu*/*ta* clearly do not pattern with the discourse markers hypothesised to occupy SAP. While their precise function is somewhat elusive, it has been widely held since at least J. Friedrich (1960: 155) that the pragmatic function of the Hittite sentence connectives is 'progression'. They act as a marker of discursive cohesion, indicating 'that the events described in subsequent clauses follow each other in a natural way, without actually stressing the temporal sequence itself' (Luraghi, 1990: 49). As such, *nu* very rarely occurs at the beginning of a discourse; *šu* and *ta* never occur in this position (Hoffner & Melchert, 2008: 390). On interpretive grounds then, *nu* etc. do not seem to fit the profile of elements occurring in SAP.

Furthermore, the syntactic distribution of the sentence connectives differs from discourse makers in one crucial way: they are used to introduce subordinate clauses as well as main clauses (Hoffner & Melchert, 2008: 390–5). The discourse markers in Hill (2007) Haegeman (2014) are categorically excluded from this position. Finally, as noted by Huggard (2015: 158, n.25), *nu* 'does not co-occur with other C-related lexemes such as *našma* 'or if', *namma* 'furthermore', *takku* 'if'.' To me, this suggests that Huggard (2015) is right to put the sentence connectives in [Spec, ForceP], with WLH clustering in Force°.Yet if we combine this with the hypothesis that (at least some of) the delayers occupy [Spec, ContrFocP] which is **below** ForceP, we cannot maintain an analysis in which -(*m*)*a*/-(*y*)*a* are consistently generated in Force°; an alternative must be sought.

As far as **sentence-connecting** -(*y*)*a* and -(*m*)*a* are concerned, we could fall back on the WL2 placement hypothesis for Vedic, i.e. -(*y*)*a*/-(*m*)*a* are generated clause externally and moved by PI to follow the first lexically accented word (i.e., ignoring *nu*/*šu*/*ta* + WLH). However, this approach does not account for behaviour of -(*y*)*a*/-(*m*)*a* with respect to the delayers. Assuming they are accented (*pace* Kloekhorst, 2014), they are perfectly acceptable prosodic hosts. As such, PI seems not entirely satisfactory as a solution, though it may account for the possible exceptions to 'indeterminate' *kuiš* (*inter alia*) acting as a delayer in Middle Hittite.

Alternatively, we may argue that the position of -(*y*)*a*/-(*m*)*a* is entirely syntactic; i.e., they are base-generated within the left periphery itself. Since they do not attach to the delayers—modelled as occupying [Spec, ContrFocP]—they must be base-generated in a lower functional head. This could be Top°, as traditionally conceived for -(*m*)*a*; analogously, -(*y*)*a* may occupy Foc°. Since TopP (and FocP)

follow ContrFocP, any of the delayers that reside in [Spec, ContrFocP] will fail to 'count' for its positioning. Instead, it will follow whatever is in [Spec, TopP/FocP]: namely, a topicalised/focalised XP, a non-delaying subordinator etc. The weakness of this account is that while it is apt to capture the use of -(*m*)*a*/-(*y*)*a* as topic/focus markers, it seems a counter-intuitive place for a clause-linking conjunction to be placed.

I do not propose to resolve the issue here. There is a somewhat cyclical logic between where we assume *nu* is base-generated and where we think -(*y*)*a* and -(*m*)*a* are base-generated. If *nu* occupies [Spec, ForceP], -(*y*)*a*/-(*m*)*a* cannot be base-generated in Force°, and must reside somewhere lower in the left periphery, and lower than the projection occupied by the delayers, at least some of which occupy [Spec, ContrFocP]. On the other hand, if *nu* is generated outside [Spec, ForceP], -(*y*)*a*/-(*m*)*a* can occupy Force° but the delayers must also be moved to projections outside ForceP, and we must find a novel explanation as to what happens when [Spec, ForceP] is ostensibly empty. Both models demand a level of stipulation that suggests neither is immediately preferable to the other as accounting for -(*y*)*a*/-(*m*)*a*; this is compounded by the fact that, for -(*m*)*a* at least, it seems at least possible that it can occur variably in ContrFoc° (per Huggard, 2015), or later (per Sideltsev & Molina, 2015). Overall, I am more convinced by Huggard (2015) that *nu* occupies [Spec, ForceP], and that (at least some of the) delayers occupy a unique [Spec, ContrFocP] projection above [Spec, TopP].

4.4.4 Summary

In this section, I have presented an overview of left-peripheral clitics in the ancient IE languages. The systems of Vedic Sanskrit and Homeric Greek are strikingly, if predictably, similar. The Latin system deviates slightly more, but is ultimately reconcilable with the other NIE languages. Hittite shows an altogether different state of affairs. There is no distinction between WL1 and WL2 akin to what we find in the other ancient IE languages. Rather, all inherited pronominal clitics occupy the same left peripheral position, which I have modelled as Force°. Notably, the indefinite pronoun does not belong to this category at all, but occupies the second position within its phrase according to scope requirements (see §3.3.3.2) and does not participate in routine fronting to the left periphery.

On the basis of this evidence, I offer the following reconstruction of clisis and the left periphery in PIE. I will adopt the term PIE_1 to refer to PIE "proper", before the Anatolian split, and PIE_2 to represent "nuclear" PIE, i.e. post-Anatolian Split.

Clisis in PIE_1 was implemented by two mechanisms: PI and left-peripheral movement. The conspiracy of these two processes accounts for what has traditionally been called "Wackernagel's Law". PI operated at PF as a last resort mechanism; where a prosodically deficient element would leave the syntax in a phonologically infelicitous position, PI "pushed" the clitic behind the first prosodic word to its right. Prosodic deficiency largely overlapped with a lack of lexical pitch accent, but some elements could have been prosodically deficient while maintaining a pitch accent. PI worked at the phrase level; clitics can therefore be moved by PI to fulfil a requirement that they have a suitable prosodic host within their clitic domain. This is the behaviour of certain WL2 clitics such as **k^we* 'and' when coordinating DPs.

PIE_1 had a set of enclitic conjunctions (WL2). These were base-generated outside the clause, but were moved by PI behind the first prosodic word of the clause. Separately, PIE had a set of enclitic pronouns (WL1) which were base-generated within the clause and moved to the left periphery. When the Topic-Focus complex was inactive, they targeted C°; when it was active, they targeted the lowest active Top° or Foc°. Depending on the utterance, this could be either Top° or Foc° in the basic model of Rizzi (1997).

The clitics belonging to the categories WL1 and WL2 vary between PIE_1 and PIE_2. There are a couple of secure WL2 cognates for PIE_2 (e.g., *k^we 'and', *we ~*$w\bar{e}$ 'or'), but beyond that different languages seem to attest different members of this set. There are no securely cognate WL2 such that we could reconstruct them confidently for PIE_1.

On the other hand, we can be confident that WL1 in PIE_1 contained a number of (mostly) monosyllabic oblique personal pronouns. The evidence suggests, however, that in PIE_1, **WL1 did not include the indefinite *k^wi-**. Rather, enclitic *k^wi- patterned more with phrasal clitics, taking second position within a phrase lower than CP. The strongest evidence for this is that Hit. indefinite *kuiš*(*ki*) does not form part of the clitic cluster in Force°, but rather takes as its clisis domain, the DP or *v*P containing it (Huggard, 2015). The fact that most branches appear show morphologically augmented forms of indefinite *k^wi-, with 'bare' k^wi- surviving only in fossilised forms suggests that that at the time *k^wi- was being "attracted" into WL1 (where it ends up in Greek and possibly PIE_2), it was also subject to morphological renewal that directly counteracted its enclitic behaviour.

To derive the Anatolian[136] situation from PIE_1 as outlined above, two catastrophic reanalyses are required. These may be situated within the broader picture described by Luraghi (1998), who concludes that 'the rigid structure of the left sentence boundary in Hittite was the result of a number of converging processes of grammaticalization', involving discourse particles, enclitic pronouns, local particles and sentence connectives (1998: 207). It is the last of these that I believe is most pivotal to the reworking of the syntactic processes underlying Wackernagel's law from PIE_1 to (Proto-)Anatolian. I will take *nu* as an illustrative example, but an equivalent argument can be made for other sentence connectives. It is uncontroversial to assume that Hit. *nu* is cognate with other IE words for 'now', cf. AGk. νῦν, Skt. *nū́* < PIE *$nú$ ~*$nū́$.[137] If PIE_1 *$nú$ had the original adverbial value of 'now', its left-peripheral position would be somewhere in the Topic-Focus complex. Absent anything else in the left periphery, it will be followed by WL1 and WL2 alike. It is only in situations where another fronted element co-occurs with fronted *$nú$ that WL1 and WL2 are disambiguated.

In the Anatolian languages, various lexemes such as *$nú$ are grammaticalised as a sentence connectives (reanalysis 1). In this process, rather than its moving to a left-peripheral position within the

[136] Although space does not allow a detailed comparative analysis of Anatolian languages other than Hittite, I take Hittite as being broadly representative of Proto-Anatolian for the purposes of this discussion. This follows from work on RCs that compares Hittite to Lycian (Garrett, 1994; Samuels, 2005) and Luwian (Yates, 2014). As for the left periphery more generally, see Melchert (2017: 192), who notes that 'the beginnings of this system [viz. sentence connectives + clitic string] are certainly Proto-Anatolian but...each of the languages has elaborated it in its own fashion.' Note in particular that while the precise orderings of the clitic string and the lexical material for the sentence connectives vary, all Anatolian languages appear to lack a distributional distinction comparable to that of WL1 vs. WL2 found in NIE.

[137] It is even attested as such a value within Anatolian, cf. Palaic *nū*.

Topic-Focus complex, what was once an adverb becomes associated with (and ultimately base-generated within) the ForceP layer. In tandem with this reanalysis, **all** enclitics—both WL1 and WL2—are interpreted not as targeting the lowest active Top°/Foc°, but as targeting **nú* (or another newly-minted sentence connective) specifically (reanalysis 2). Thus, once **nú* (vel sim.) is base-generated in [Spec, ForceP] and regularly precedes Topics and Foci, the clitics move with it, attaching to Force°. The actualisation of this change is witnessed in two ways: (i) enclitics raise over any Topics or Foci, and (ii) in the absence of anything in [Spec, ForceP], the clitics still raise to Force° and are moved by PI to follow the first prosodic word in the sentence. This leads to the collapse of WL1 and WL2 as categories in Anatolian. It is likely that the syntactic development of -(*m*)*a* and -(*y*)*a* leading to the distribution discussed above post-dates this change precisely because they do not attach to proclitic Hit. *nu*; the set of clitics that can take proclitic *nu* as a prosodic host is a closed set; new clitics such as -(*m*)*a* and -(*y*)*a* require a full prosodic word as a host. The development of a dedicated ContrFocP is likewise an innovation that likely took place after the grammaticalisation of sentence connectives, though it has a crucial role in explaining the development of 'indeterminate' *kuiš* in Anatolian, to which I return in §5.2.

While these reanalyses are taking place in Proto-Anatolian, the PIE_2 left periphery begins to crystallise, with the development of WL2 enclitics **k*^{w}*e* and **-we*, as well as the aforementioned attraction of indefinite **-k*^{w}*i* towards WL1, a change which is incomplete at the end of the PIE_2 period. The left periphery of PIE_2 is preserved most clearly in Vedic and Homeric Greek, while Latin undergoes a subsequent reanalyses of WL1 placement as outlined in§4.4 above. This model of the left periphery and Wackernagel's law provides the grammatical context within which we can establish the precise syntactic behaviour of *REL in PIE.

Chapter 5

Relative Clauses in Proto-Indo-European

In this chapter, I reconstruct the principal features of relative clause syntax in Proto-Indo-European. According to the methodology of Minimalist reconstruction (§2.1), of central importance will be ascertaining the feature values of the PIE relative pronoun, *REL (§3), and its position in the PIE left periphery (§4). Moving away from *REL, I turn to the more general structures of, and relations between, 'plain' relative and correlative clauses in Proto-Indo-European. The chapter is rounded off by a discussion regarding the semantic types of relative clauses in PIE and their interaction with syntax.

5.1 Position of the relative pronoun

Many authors cited so far in this thesis have either tacitly assumed—or explicitly argued—that PIE had *wh*-movement in RCs, a movement also hypothesised for interrogative clauses.[138] From two different perspectives, I have suggested that we should be cautious when using the term *wh*-movement, as it collapses what I consider to be two distinct types of movement: the fronting of interrogatives, and the fronting of relative pronouns.

One set of arguments comes from the theoretical syntactic literature. *Wh*-movement is traditionally understood to target [Spec, CP] as a landing site. Once we acknowledge that [Spec, CP] is only shorthand for a distinct set of structural positions (Rizzi, 1997), including the Topic-Focus complex, the picture changes. In the first place, if fronting is understood to have some pragmatic motivation, it is not immediately obvious that the relative and interrogative pronouns should be fronted for the same reasons, at least in the narrow syntax.[139] Indeed, throughout this thesis, I have held that fronted interrogative pronouns are focalised, and hence occupy [Spec, FocP].[140] On the other hand, fronted relative pronouns are topicalised, and hence occupy [Spec, TopP]. .

This brings us onto the second set of arguments that I have made at length concerning the status of *wh*-movement in ancient IE languages, which arises from the distributional evidence from the attested languages. Of the four archaic branches of IE I have investigated in some detail (Hittite, Sanskrit, Ancient

[138] See, e.g., Hale (1987, 1996), Hock (1989), Garrett (1994), Kiparsky (1995).

[139] I do not address here the notion of 'covert raising' or 'LF raising', on which see *inter alia* Cole (1987) and Williamson (1987). I also set aside the motivations for *wh*-movement that involve scope, on which see Lutz *et al.* (2000).

[140] I have not considered the IntP projection of Rizzi (2001), though further studies aimed specifically at interrogative syntax in PIE may shed light on whether such as position is necessary to account for the attested patterns.

Greek, Latin) **there is a clear distributional difference between interrogative and relative pronouns** (see §4). A compounding factor in this distributional separation, which I have touched on only briefly from a comparative perspective, is the possible status of PIE *REL as a **clitic**. An enclitic *REL would not only occupy a left-peripheral position different from that of an accented interrogative, it would presumably raise as part of "Wackernagel's Law", in the same process as other enclitic pronouns.

In the first part of this chapter, therefore, I address to the question of clisis, first for the languages in which a purportedly enclitic *REL surfaces as **yó-* (Sanskrit, Celtic) and then those in which is surfaces as **k*^{w}*ó-* (Hittite). I will argue that there is insufficient motivation to reconstruct **yó-* as enclitic. The evidence from Hittite is more complex, but I will ultimately argue that the clitic behaviour of 'determinate' *kuiš* is also likely an innovation, which does not provide any clear evidence for an enclitic *REL in PIE.

With clisis excluded as a possibility, I will make the straightforward step of arguing that left-peripheral PIE *REL occupied [Spec, TopP]. More speculatively, I will conclude that it was the **lower** [Spec, TopP] of Rizzi (1997), though the evidence in this regard is less clear.

5.1.1 Relative clisis in Proto-Indo-European: A negative appraisal

5.1.1.1 Sanskrit

I argued in §4.3 that Sanskrit *yá-* is not enclitic. This was in the context of criticising Lowe's (2014) theory of left-peripheral clitics in Vedic. However, the tradition of reconstructing non-initial **yó-* in PIE as enclitic has a longer history than that, dating back to at least Watkins (1963) and Hettrich (1988), whose hypotheses I discuss below.

As noted briefly in §4.2, discussions of clisis in the vein of Wackernagel (1892) often lack a theoretical context in which clisis is defined **syntactically**. Although Watkins saw the possible value of using generativist syntactic analysis in reconstruction (1963: 3), his understanding of clisis was—syntactically speaking—quite atheoretical. He uses the term 'enclitic position' several times, including in relation to the relative pronoun in Vedic; yet what this position constitutes does not clearly correspond to the way I have defined clisis so far.[141] Rather, working on the Old Irish verbal complex (to be discussed in detail below), Watkins (1963) establishes the following pre-historic word orders, supposedly inherited from PIE:

1. # Verb – Enclitic(s)...
2. # Preverb – Enclitic(s) – Verb...

On this basis, he notes that 'that the position of the enclitic element is constant: E[nclitics occupy] the **second position in the sentence**, whether preceded by V[erb] or by P[reverb]' (Watkins, 1963: 12, emphasis mine). Thus, he is able to use 'enclitic **position**' to mean simply 'second position'.[142] On this basis, then, Watkins (1963: 29) suggests that 'the...enclitic position of the relative stem **yo-* occurs in the

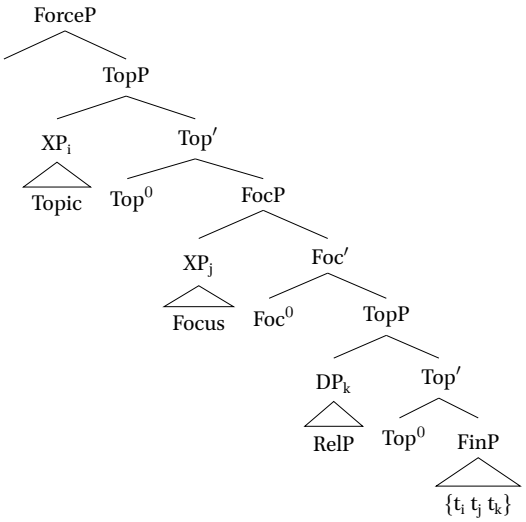
[141]The same point is made by Lowe (2014: 24) with respect to his own definition of syntactic clisis, which broadly corresponds with my own.

[142]It is unclear whether this is calculated by word or constituent in Watkins (1963); the former seems more likely given the evidence he quotes.

Rig Veda, though it is not mandatory.' Hettrich (1988: 758–62) makes broadly the same points; unlike Watkins (1963) he notes that in marked word orders, a constituent bearing 'contrastive emphasis' (*Kontrasthervorhebung*) may precede Vedic *yá-* without the latter being enclitic; however, he suggests that *yá-* could be considered enclitic in cases where he does not judge the fronted constituent to be sufficiently emphatic. He also cites the evidence from Celtic in turn as comparative support for the fact that Vedic *yá-* in enclitic 'second position' after either a verb or a preverb could be inherited (Hettrich, 1988: 760).

In the theoretical framework adopted in this thesis, Skt. **yá-* is not syntactically enclitic, in the sense that it does not raise to C° (or the lowest active Top°/Foc°) like WL1 (see §4.3–4). Rather, I have argued that we should understand both the relative pronoun and the constituent preceding it as occupying distinct specifier positions within the Topic-Focus complex: relative *yá-* occupies the lower [Spec, TopP] in the model of Rizzi (1997). This is represented in the following syntactic representation, repeated from §4.3.3:

(5.1)



My model is an alternative to that offered by Lowe (2014), in which non-initial relative *yá-* (and demonstrative *sá-*/*syá-*) are optionally syntactically enclitic, raising to a unitary C° (see §4.3). It is worth noting, however, that Lowe (2014) is primarily a synchronic account of the facts in Vedic: indeed, the author notes quite explicitly that (2014: 26, emphasis mine):

> ...the existence of an enclitic variant of the relative pronoun in the Ṛgveda is a reasonable proposition...on comparative grounds **(whether those comparative grounds support inheritance or parallel development).**

In other words, even if we were to accept Lowe's hypothesis for Vedic *yá-*, the author does not make any rigid claims as to whether this would support the existence of an optionally enclitic **yo-* in PIE. Yet it is worth exploring the situation in which this could be possible. In the first place, we have strong evidence that PIE contained a **non**-clitic form **yó-*, as evidenced in the more dominant non-clitic behaviour of Skt. *yá-*, AGk. ὅς etc. (§3.2.1). If we were to posit a putative enclitic relative pronoun **yo-*, it

would seem reasonable to suppose that it would have developed from its non-clitic counterpart, following well-established principles of grammaticalisation (Hopper & Traugott, 1993).[143] However, as Lowe (2014) notes, in later Sanskrit 'enclitic' *yá-* is entirely lost; the usual position for *yá-* is in clause-initial position, as it is in the majority of Vedic tokens. Lowe's argument is that *yá-* was only ever **optionally** enclitic, so his suggestion is not that 'enclitic' *yá-* was somehow decliticised. Yet if we accept the existence of optionally enclitic **yá-* in Vedic, and that this grammatical behaviour was inherited, we still have to posit that a change towards the cliticisation of **yó-* (*qua* *REL) > **yo-* began in PIE, was maintained briefly in Sanskrit, but was ultimately abandoned. Such a set of changes is not beyond the realms of possibility,[144] but if we have independent evidence to suggest that Vedic *yá-* was not at all clitic—as I have argued in §4.3—Sanskrit provides no evidence in favour of this reconstruction.

To summarise: as far as Vedic is concerned, I am confident that the relative pronoun *yá-* is never enclitic. With the rule for WL1 placement established (target the lowest active Top°/Foc°), a unified analysis in which **yá-* consistently occupies (the lower) [Spec, TopP] can account for every instance of left-peripheral *yá-*, both initial and non-initial. To posit that non-initial *yá-* is (optionally) enclitic therefore adds a level of syntactic complexity to Sanskrit *yá-*, and an abstruse grammaticalisation process that lacks sufficient motivation. I conclude that Sanskrit offers no support to the hypothesis that **yó-* (and, implicitly, *REL) was enclitic in PIE.

5.1.1.2 Celtic

On the other hand, there is incontrovertible evidence for enclitic relative **yó-* in another branch of IE: Celtic.[145] Recall that in §2 I used Celtiberian, with its non-clitic form of **yó-* as representative of Proto-Celtic. This is not an uncontroversial position: I provide some motivation for the claim now, which has significant implications for how we construct the syntactic behaviour of PIE **yó-*.

It is worth mentioning in the first place that much of the scholarship on RCs in Celtic predates the publication of Celtiberian texts. Watkins (1963), for example, did not have access to Celtiberian; he therefore focuses on Old Irish and, to a lesser extent, Gaulish, both of which support the existence of some form of enclitic relativiser. Nevertheless, since the publication of the first Botorrita Bronze by Beltrán & Tovar (1982), it has become clear that Celtiberian exhibits an inflected, non-enclitic form of the relative pronoun, as described by Beltrán & Jordán (2019: 260):

> The forms believed to belong to the relative pronoun paradigm are: **ios** [Z.09.01, A-10], [Z.09.24, A-7], nom. sg. masc.; **ia** [Z.09.03, 01], nom. sg. fem. or nom. acc. sg. neut.; **iom** [Z.09.01, A-5, A-7, A-10], [Z.09.24, B-4], possibly acc. sg. masc.; **ias** [Z.09.01, A-8], acc. pl. fem.; **iomui** [Z.09.01, A-7], dat. sg. masc.?

There are two very general arguments that could lead us to take the situation Celtiberian situation as representative of Proto-Celtic, without any further consideration of the other Celtic languages.

[143] See further Hendery (2012: 155–7) for a discussion of this grammaticalisation pathway specifically for relativisers.

[144] See further Haspelmath (2004: 33–5); this could be considered a case of 'retraction'.

[145] I am deeply indebted to Mark Darling for his invaluable input to the following section, including his generous support in navigating the Old Irish corpus and enlightening discussion on the topic of Celtic relativisation strategies. Needless to say, any inaccuracies are of my own doing.

One is its age: with the Botorrita bronze dating to the first century BCE, it predates the oldest Irish texts by over half a millennium. Yet the same claim could arguably made for Gaulish, which has an enclitic relative pronoun (see ex. 5.7 below) and is attested around the same time as Celtiberian. On the other hand, there is also the aforementioned tendency for lexical words to grammaticalise into clitics (and then affixes), rather than *vice versa*. Such generalisations, however, do not consitute proof; we need to take a look at the comparative data.

Starting with Old Irish, then: Thurneysen (1946: 312–20) describes three types of RC. These are summarised by Jasanoff (1999: 205, see therein for details):

1. Prepositional RCs
2. Leniting RCs
3. Nasalising RCs

The first of these is formed by the addition of an indeclinable 'relative particle', -(s)a^n,[146] to the preposition that would govern the relative pronoun in its translational equivalent. Thus, e.g., OIr. *ar-* 'because of', *ar-a*^{n} 'because of **which**', 'because of **whom**'. The 'relative particle' here is formally identical to the neuter accusative singular of the definite article, *in*, which is ultimately derived from PIE **só-*/**tó-*, and thus irrelevant to the current discussion.

The two other forms of relativisation, however, show specialised verb forms that reflect a descendant of PIE **yó-*. The phonological pathways from **yó-* to the attested forms are complex and thoroughly discussed elsewhere.[147] For our purposes, the phonetic residue of **yo-* surfaces in two ways. First, as a specialised ending when following the verbal stem directly:[148]

(5.2) **Simple form:** *carait* (Ml. 124d13) 'they love' < **kár-onti*
Relative form: *cartae* (Strachan, 1907: 7, l.14) 'whom they love', 'who love' < **kár-onti-**yo***

Second, as either a lenition or nasalisation[149] of the **following** onset when preceding the verbal stem:

(5.3) **Simple form:** *adciam* (Wb. 6a30) 'we behold' < **ad-k*^{w}*ísomos*(*i*)
Relative form: *adchiam* (Ml. 42b22) 'which we behold' < **ad-**yo**-k*^{w}*ísomos*(*i*)

With this maximally brief summary of the phonetic facts, I now turn to the syntactic analysis put forward in Watkins (1963). In the same vein as the formulae listed above for the position of enclitics, the

[146] Superscript <^{n}> is used hear to indicate nasalisation of the following syllable.

[147] See, e.g. Thurneysen (1946: 312ff.), Lewis & Pedersen (1937: 243ff.), Watkins (1963: 24–9), Jasanoff (1999: 205–8). I follow McCone (1997: 254) in treating nasalising RCs as essentially an Irish innovation that follows the same syntactic pattern as leniting RC. In my reconstructions I write **yo*$^{(n)}$ to represent **yo* when it is attested subsequently in a nasalising relative construction; this is not a comment on the reconstructed phonetic reality in Proto-Celtic.

[148] A specialised form of this type also appears to be attested in Middle Welsh *yssyd* 'which is' < **esti-yo* (Lewis & Pedersen, 1937: 243), but is not attested with any other verbs. See further Roma (2007: 257–64).

[149] "Nasalisation" is the standard term for this process, though it does not always result in a nasal outcome phonetically; depending on the sounds involved, it can also manifest synchronically in other ways, such as voicing or gemination (Thurneysen, 1946: 147).

author posits the following reconstruction for the distribution of the relative pronoun in Proto-Celtic (Watkins, 1963: 31, adapted):

(i) #V[erb] **yo-* ... #

(ii) #P[article] **yo-* (P_x ...) V... #

The first thing to note is the form of **yo* that is used in relativisation strategies. We can be confident that whatever form of **yo-* is implicated in OIr. relativisation, it was indeclinable; this is clear in the fact that relative verb forms do not reflect case relations: as Watkins (1963: 24) notes, '*ind fir* ***bertae*** is both *uiri qui ferunt* [men who carry] and *uiri quos ferunt* [men whom they carry]'. Thurneysen (1946) and Lewis & Pedersen (1937) argue the form was in fact the neuter singular **yo-d* (cf. the use of uninflected *yád* as a quasi-complementiser in Sanskrit, §3.2.1), while Watkins (1963) argues for the bare stem **yo*, which he prefers on a phonological basis, drawing an etymological connection to the Hittite sentential connective -(*y*)*a*- (see §3.2.2). I return to questions of form below, but for now I continue to write **yo* without any strong commitment to its phonetic form.

Formula (i) yields specialised relative verb forms such as in ex. (5.2) above. Formula (ii) is somewhat more complex. First, we must understand what is meant by P. This category includes the inherited preverbs/prepositions (P-words). As Thurneysen (1946: 495) notes, these occur variably in compounds (both nominal and verbal), as well as prepositionally, i.e. governing a noun in a dependent case or with a suffixed pronoun.[150] To this we add what are traditionally referred to as **conjunct particles** (Thurneysen, 1946: §38). The conjunct particles are a grammatically heterogeneous category including:

a. Negators: *ní, na, nad* etc.

b. Complementisers: *ara*^{n} 'in order that', *dia*^{n} 'if', *co*(*n*)^n 'so that'

c. The interrogative particle *in*

Finally, P includes *no-*, a semantically vacuous conjunct particle that serves the same syntactic function as a preverb but occurs only with simple verbs and in the absence of any other conjunct particle. *no-* is particularly important to the formation of RCs, because the specialised verb forms such as ex. (5.2) only exist in the 3sg, 1pl and 3pl. In all other persons, therefore, in the absence of any other conjunct particles, *no-* is obligatory in the relative forms of simplex verbs. This yields the following idealised paradigm:

[150] See Thurneysen (1946: 495–539) for a full overview.

	Non-relative	Relative
1sg	*caraim* 'I love'	*no-**ch**araim* 'which I love'
2sg	*carai* 'thou lovest'	*no-**ch**arai* 'which thou lovest'
3sg	*caraid* 'loves'	*car**as*** 'who/which he loves'
1pl	*carmai* 'we love'	*car**mae*** 'which we love'
2pl	*carthae* 'you love'	*no-**ch**araid* 'which you love'
3pl	*carait* 'they love'	*car**tae*** 'who/which they love'

Table 5.1: Relative and Non-relative forms of OIr. *carait*, reproduced from McCone (1997: 14).

Formula (ii) denotes that when there are multiple P-elements, which there often are, **yo*- occurs after the **first** P, and is separated from the verbal stem by any additional P-elements. Synchronically, this is evidenced by the fact that in multi-preverbal relative verbs, the lenition or nasalisation affects not the verbal stem, but the second preverb.

(5.4) *for-**n**-diuclannar* (Ml. 104b5) 'which is devoured' < **wor-**yo**$^{(n)}$-dī-uss-glann-*

Watkins (1963) argues that the distribution of **yo*-, as summarised in these two formulae, is identical to that of the enclitic pronouns (as in the first set of formulae mentioned in §5.1.1.1 above).

(5.5) *berith**i*** (Ml. 42b7) 'he refers (lit. "bears") it' < **bereti-**i**(**d**)*
(Suffixed pronoun ≈ Formula 1)

(5.6) *immi-**m**-thabarthar* (Ml. 41c2) 'if I be surrounded' < **imbi-**me**-do-beretor*
(Infixed pronoun ≈ Formula 2)

With this analogy drawn, he comes to the following conclusion: the relative pronoun **yo* falls into the same syntactic category as the enclitic pronouns, and they are placed in 'clitic' second position according to Wackernagel's law. That this is an inherited word order, Watkins argues, is verified by what he considers to be an equivalent 'clitic' position of the relative pronoun in both Hittite connective -(*y*)*a*- and Vedic Sanskrit *yá*- when it appears in second position (see §3.2.2.2 and §5.1.1.1 above respectively; I do not accept either as convincing *comparanda*).

There seems to be broad consensus that Watkins (1963) was correct about an uninflected enclitic relative **yo* being a feature of all the ancient Celtic languages available to him at the time. Indeed, well before Watkins, Thurneysen observed that enclitic **yo* is attested directly in Gaulish:[151]

(5.7) *martialis* *dannotali* *ieuru* *ucuete* *sosin*
Martial.NOM.SG Dannatolos.GEN.SG. dedicate.PRET.3SG Ucuetis.DAT.SG DEM.ACC.SG
celicnon *etic* *gobedbi* ***dugiíonti=ío*** *ucuetin* *in* *alisiía*
building.ACC.SG and blacksmith.INS.PL honour.PRES.3PL=REL Ucetis.ACC.SG in Alisia
'Martial (son) of Dannotalus dedicated this building to Ucuetis, and with the blacksmiths who honour Ucuetis in Alisia' (RIG II 1 L-13)

[151]See Lambert (1994: 98–102).

Watkins' analysis of Old Irish demonstrated that this was a shared inheritance. Yet the argument that an enclitic relative **yo* was a feature of Proto-Celtic, ultimately inherited from PIE, faces a severe challenge in the form of the Celtiberian evidence cited above. Setting aside any prejudgement based on age or grammaticalisation, there are two ways we can square this circle:[152]

1. Proto-Celtic inherited both accented (+ inflected) **yó-* and enclitic (+ uninflected) **yo* from PIE as relativisation strategies: the latter is reflected in Gaulish and Old Irish, the former in Celtiberian
2. Proto-Celtic inherited only accented (+ inflected) **yó-* as a relativisation strategy, and enclitic (+ uninflected) relative **yo* was an innovation that took place after the separation of Celtiberian from the rest of the language family

As Wodtko (2000: 138) notes, it is currently unclear as to which of these hypotheses is correct; I do not believe it is possible to prove either comprehensively on the basis of the evidence available to us. However, I would argue that the second scenario is the more likely one, both on the basis of the Celtic evidence and with regard to the *comparanda* from outside Celtic, especially Sanskrit.

I will start with the issue of the syntax of enclitic relative **yo*, as attested in Old Irish and discussed in Watkins (1963). As I noted in §5.1.1.1 above, Watkins (1963) does not have a clear definition of syntactic clisis. Thus, he does not consider the possibility that the occurrence of **yo* in "second position" could be due to either to clisis (i.e., raising to a functional head, C^{o}/Top^{o}/Foc^{o}) or some other, non-clitic form of movement (i.e., to a left-peripheral specifier, i.e. [Spec, TopP]/[Spec, FocP]). Thus, the "second position" placement of (Pre-)Old-Irish **yo* (and, according to Watkins (1963), non-initial Vedic *yá-*) is treated by default as a result of the same process as the appearance of enclitic pronouns in "second position", viz. "Wackernagel's law". I would argue, however, that once we have established that Vedic non-initial *yá-* is definitely **non**-clitic (see §5.1.1.1 above), we may flip Watkins' argument on its head: since it is possible to generate the attested word order with **yó-* in second position **without** clisis, we can argue that the attested position of **yo* in Old Irish has its roots in a **non-enclitic** form **yó-*: specifically, when it occupies the lower [Spec, TopP], preceded by another TopP and/or FocP. In short, the enclitic **yo* he reconstructs from Old Irish continues a word order—albeit marginal—that we have reason to reconstruct for non-clitic *REL in PIE.

As for how this marginal word order became generalised as the norm, I believe the answer is intrinsically tied to the shift from SOV to VSO. A full discussion of this reconfiguration is far beyond the scope of this thesis, but a tentative outline of the logic is as follows.[153] Numerous studies, employing various methodologies, have come to the conclusion that PIE was quite likely underlyingly SOV.[154]

[152] According to the analysis I presented in §3.2, I do not consider here the possibility that non-clitic relative **yó-* is some sort of Celtiberian innovation (*pace*, e.g., Szemerényi, 1996: 210–11).

[153] See Hickey (2002) and the references therein for a (relatively) recent overview of the literature on this topic. Many authors, not least Watkins (1963) and Eska (1995), have argued for a link between the positioning of clitics and the shift from SOV > VSO, though their accounts differ. My account stands apart from theirs insofar as I do not believe **yó-* must be enclitic in order to occur in second position, and thus clisis, in the form of Wackernagel's law and Vendryes' restriction, does not 'drive' syntactic change in the way these authors envisage, though it may be implicated at a later stage.

[154] See most recently, e.g., Hock (2015), Krisch (2017), Keydana (2018), Windhearn (2021).

Within the Celtic family, Celtiberian appears to maintain this word order. On the other hand, Old Irish has unmarked VSO word order. This leaves Gaulish, which seems to occupy a sort of middle-ground, with unmarked SVO word order. It is reasonable to hypothesise that Proto-Celtic was SOV, and instantiated the marked patterns of non-initial **yó-* that are also attested in Vedic. We may further posit that this pattern was inherited and maintained in Celtiberian, which remains SOV and keeps the non-clitic, inflected paradigm of the relative pronoun *ios*.

However, once Celtiberian splits off from the rest of the family, things start to change. The fronting of constituents, which was pragmatically conditioned in PIE, begins to become grammaticalised. In the insular languages, the fronting of the preverb (for compound verbs) or the verb itself (for simplex verbs), becomes the unmarked order. At this point, sentence-initial **yó-* is altogether lost; it is found most commonly in "second position", following either P or V (cf. Watkins' formulae above). With this distribution, it is quite reasonable that **yó-* is reanalysed as enclitic; its distribution now overlaps almost entirely with the inherited enclitic pronouns. The sole difference is that **yó-* precedes other enclitic pronouns:

(5.8) *dondrón* (Wb. 5d18) '(in) which I do it' < **de-**yo**$^{(n)}$-**di**(**d**)-ro-gn-*

But whereas once this was because **yó-* occupied [Spec, TopP] and the enclitics raised to Top°, this can now be reanalysed as simply an ordering constraint within a clitic cluster occupying the same syntactic position.

Once **yó-* is reanalysed as syntactically enclitic, the loss of accent and inflection follow from the well-observed phonological concomitants of grammaticalisation (Hopper & Traugott, 1993: 154–9). It is possible that at the earliest stages of its development into a clitic **yo-* remained inflected; over time, however, through the usual processes of erosion, the endings were lost, the result being an uninflected monosyllable. It is for this reason that I believe neither **yo* nor **yod* meaningfully represents the "original" form of the enclitic relative pronoun in Insular Celtic and Gaulish, as what we are really looking at here is a case of paradigm atrophy rather than the selection of, say, a neuter singular form that supplants the rest of the paradigm. Nevertheless, that **yod* may have been generalised as the enclitic relative cannot be ruled out. Sound changes related to grammaticalisation are often idiosyncratic and thus need not be part of a Neo-Grammarian rule affecting the entire language.[155] As such, the fact that word-final /d/ after a short vowel is not regularly lost in Gaulish is not definitive proof either way, *pace* McCone (1997: 250). In any case, it is evident that the relative marker underwent a serious level of phonological reduction.

To conclude this section: it is perfectly plausible that enclitic relative **yo* as attested in Insular Celtic and Gaulish, is straightforwardly a grammaticalised form of non-clitic **yó-* as attested in Celtiberian. Without any support from Vedic for the presence of an enclitic relative **yó-* in PIE, and without a secure etymological link to Hittite connective -(*y*)*a*, it seems far more plausible that Proto-Celtic simply inherited the full lexical form of relative **yó-* from PIE than that it inherited both a full lexical form and an enclitic form.

[155] On this topic see Mańczak (1980), Shield (1980), Picard (1997), Enrique-Arias (2005).

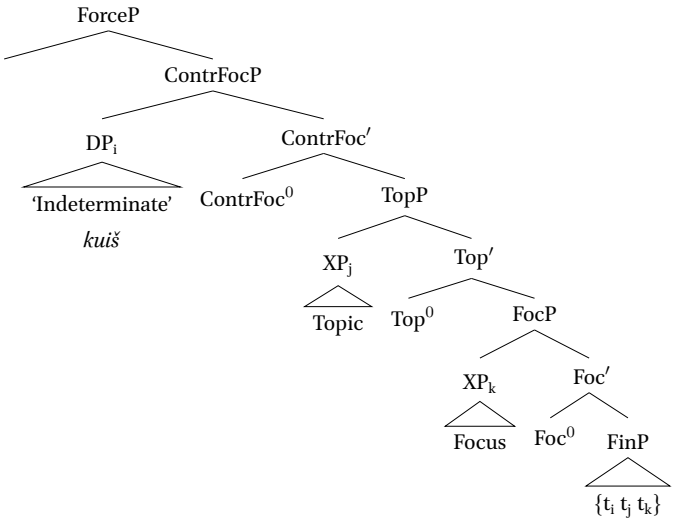
5.1.1.3 Hittite

I noted in §3.3.2 that the Hittite relative pronoun *kuiš* in 'determinate' relatives was prosodically deficient. The phonological evidence for this is presented in Kloekhorst (2014: 624–6), and discussed in relation to the syntactic behaviour of *kuiš* in Huggard (2015: 158). From a distributional perspective, 'determinate' *kuiš* is **always** preceded by at least one lexically accented word in the clause, such that if there is only one constituent other than *kuiš* (i.e., the verb) *kuiš* will **follow** the verb.

(5.9) *paprizzi* *kuiš* 3 GÌN KUBBABAR *pāi*
sully.PRES.3SG REL.NOM.SG.C 3 half-shekels-of-silver give.PRES.3SG
'The one who sullies gives three shekels of silver' (KBo 6.2 I 57)

However, relative *kuiš* does not occupy the Hittite clitic string (§4.4.3). On the contrary, according to Huggard (2015), 'determinate' *kuiš* must remain within the *v*P domain and cannot be fronted to the left periphery; its cliticisation domain is the DP when there is an overt NP, or the *v*P in its absence' (see also §3.3.2). On the other hand, *kuiš* in 'indeterminate' relatives is unambiguously non-clitic: it usually occurs in initial position (discounting *nu*+clitics), and exceptionlessly precedes its head noun (Huggard, 2015: 29). In §3.3.2.1, I accepted the argument from Huggard (2015) that *kuiš* in these clauses occupies [Spec, ContrFocP], yielding *wh*-conditional correlatives. In §4.3, I established that ContrFocP is immediately dominated by ForceP and dominates TopP/FocP, yielding a left periphery of the following form:

(5.10)



I argued further in §3.3–4 that *kuiš* in both 'indeterminate' and 'determinate' relatives should be understood as a grammatical descendant of *REL; I held that the feature specification argued by Arsenijević (2009) of *wh*-words being treated as bearing 'weak existential commitment' did not imply that *kuiš* under these circumstances should be treated as an indefinite rather than a relative pronoun. In short, both 'determinate' and 'indeterminate' relatives in Hittite should factor into our reconstruction of PIE *REL.

How, therefore, should we reconcile the distinct behaviours of 'determinate' and 'indeterminate' *kuiš* in our reconstruction? If we take the former as representative of the 'original' *REL, we must argue

for some form of productive decliticisation to yield the 'indeterminate' word order; by extension, this decliticisation would have to have occurred separately for *REL in PIE_2, given the abundant evidence for non-clitic *REL from NIE, whether of the form **yó-* or **k^wó-*. On the other hand, if the 'indeterminate' pattern is the original, we have to account for how it gained its unique semantics, and how 'determinate' *kuiš* became enclitic. I will argue that the second option is more plausible, given the unique developments of the Anatolian left periphery.

Let us start with the premise that PIE_1 had a non-clitic *REL that was routinely fronted to the left periphery. The evidence from NIE suggests that its landing site was [Spec, TopP]—more specifically, the **lower** [Spec, TopP], such that it could be preceded by an element in either [Spec, FocP] or the higher [Spec, TopP] (marginally, both). In this scenario, there is no grammaticalised positional rule regarding the determinacy of the RC. Correlatives may therefore be either **specific** in that they only apply to real situations, or **conditional**, in the sense that they abstract over all possible worlds. Such an ambiguity is attested elsewhere: Arsenijević (2009: 147) argues that we see it in Serbo-Croatian *wh*-conditional correlative such as the following.

(5.11) *Koji student je prvi ušao, taj je položio ispit*
which student AUX first entered, that AUX passed exam
Specific: 'The student who entered first passed the exam'
Conditional 'Any student who entered first passed the exam'

Commenting on this example, Arsenijević (2009: 147) notes that:

> It has one reading [viz. **specific**] that refers to the actual world, and states that a definite and topical eventuality of passing a definite and topical exam involves as its theme a particular definite and topical (or possibly focalized) student who entered first. The other reading [viz. **conditional**] involves possible worlds, and states that for any student that passed the exam, the possible world in which this happened is characterized by the same student that entered first.

Semantically, this ambiguity is strikingly similar to the distinction between 'determinate' and 'indeterminate' correlatives in Hittite. Arsenijević (2009: 146–7) argues that the two are distinguished by virtue of the adjunction site of the correlative RC; for our purposes, it is sufficient to note that the surface forms are identical.

At some point after its split from PIE, Proto-Anatolian sees the development of a unique ContrFocP projection at the top of the left periphery, occupied by certain conditional adverbs such as *mān* ('if').[156] The precise motivations for the development of this position remain obscure, but they are part of a larger set of Anatolian left-peripheral innovations as outlined in §4.4. Where *REL is fronted, in the absence of any other topics, it can be analysed as occupying either [Spec, ContrFocP] or [Spec, TopP]. A complementary proposal to Arsenijević's suggestion that conditional vs. specific readings are dependent on the adjunction site is that the position of *REL in the left periphery is also implicated: [Spec,

[156] It is possible that such a projection existed in PIE_1 itself, but I have not established any clear evidence for it as distinct from FocP except in Anatolian. Moreover, in NIE languages the difference between a conditional ('indeterminate') and a specific ('determinate') RC can be encoded in the mood of the verb, with the subjunctive/optative for the former and the indicative for the latter. This presents an alternative means of expressing conditionality, possibly present in PIE_2, that would make any putative ContrFocP redundant in RCs.

ContrFocP] generates the conditional ('indeterminate') reading, while the [Spec, TopP] gives the specific ('determinate') meaning. If this is the case, and ContrFocP sits at the top of the left periphery, the only instance in which there is no ambiguity between specific and conditional meanings is when *REL is **non-initial**. Here, it is clear that *REL does not occupy [Spec, ContrFocP], but rather [Spec, TopP], since the former cannot be preceded by any other lexical items. On the other hand, if *REL is initial, it could occupy **either** [Spec, ContrFocP] or [Spec, TopP].

This proposal makes the same distributional predictions as Garrett (1994): namely, that there are (at least) two left-peripheral positions, and the relative pronoun (in 'determinate' relatives) is only fronted to the lower projection. Given the criticism of this model, including Huggard (2011, 2015) and Yates (2014), I have not argued that Garrett's model can account for the Hittite data synchronically. Nevertheless, from a diachronic perspective, Garrett's analysis provides a crucial link between the situation as observed in NIE and the patterns that arise in the attested Anatolian languages. What Garrett's model does not account for satisfactorily—namely, the **prohibition** of 'determinate' relative pronouns in initial position—requires some further explanation.

To derive the attested distribution in which 'determinate' *REL is **never** clause-initial, the next stage is the deaccentuation of **unambiguously** 'determinate' (i.e., non-initial) *REL in Proto-Anatolian. This change is mirrored by the developments in Celtic. Interestingly, the Celtic data seem to reflect cliticisation of *REL via multiple fronting into the left periphery, with the relative pronoun occupying the lower [Spec, TopP]. Such patterns may also have been prevalent in Proto-Anatolian, per hypotheses such as those of Garrett (1994), Samuels (2005) Hoffner & Melchert (2008) *et al.*, who assume *wh*-movement (i.e., movement of the relative pronoun to [Spec, TopP]) in Anatolian. On the other hand, given the observations of Goedegebuure (2009), Huggard (2011, 2015) and Yates (2014), it may be the case that in Proto-Anatolian the *wh-in-situ* strategy, marginally attested elsewhere (see §5.1.2 below), was more prevalent. This would have further increased the number of tokens of non-initial *REL, which may have catalysed its grammaticalisation as a clitic. One could further argue, as implied by Huggard (2015), that 'indeterminate' *REL could never appear *in-situ* in Proto-Anatolian, but obligatorily moved to [Spec, ContrFocP], for reasons of scope: this in turn could have strengthened the association between non-initial (+enclitic) *REL and 'determinate' semantics.

Clearly the most striking way in which the grammatical outcome for *REL in Anatolian differs from the analogous process in Celtic is the maintenance of non-clitic 'indeterminate' *REL in [Spec, ContrFocP], a position which was not grammaticalised in this way outside Anatolian. Where *REL occupies [Spec, ContrFocP], it necessarily stands in clause-initial position and so cannot be enclitic; on the contrary, it may well bear a specific intonational morpheme associated with focus (Huggard, 2015: 149).

The final stage is a reanalysis in which de-accentuation (clisis) is taken to be part of the feature specification of 'determinate' *REL. This is the situation we see in Hittite. It is not the case that the grammar 'requires' some higher left-peripheral projection to be filled in order to get a 'determinate' reading of *REL; it is simply the case that *$\text{REL}_{+\text{DET}}$ is enclitic, while *$\text{REL}_{+\text{INDET}}$ is not.[157] It could be that in certain

[157] Synchronically, this could be captured by a model such as Huggard's (2015), where the various functions of *kuiš* (indefinite, interrogative, relatives of all types) are collapsed into a single underlying lexical item, which is either enclitic or not

instances it was still occupy a left-peripheral position, per Garrett's (1994) analysis, or that it became strictly *wh-in-situ*, as suggested by Huggard (2011, 2015) and Yates (2014), since both options were available in PIE.

This proposal is more attractive than the alternative in several ways. In the first place, if we were to start from opposite premise—that PIE_1 *REL was enclitic—we have to posit two separate instances of its decliticisation: one within Proto-Anatolian, for the purposes of deriving 'indeterminate' relatives; and one in PIE_2 (see §4.4 for terminology), to account for the fact that it does not have an enclitic relative pronoun. Setting aside the typological arguments against positing such a change (twice),[158] it is notable that it is not attested with any other clitic pronouns. If PIE_1 *REL was originally enclitic but was subsequently decliticised, we would have to explain why this possibility was uniquely available to *REL and not to other members of the clitic pronouns, such as WL1 or the indefinite pronoun *k^wo- (without morphological augmentation). Furthermore, although I have argued at length that PIE_2 allowed for fronting around the relative pronoun while it was still unambiguously **non**-clitic, there remains the possibility of a *wh-in-situ* strategy, albeit only marginal outside Anatolian (§5.1.2 below). It is parsimonious to suggest that these behaviours were inherited from PIE_1, as long as we can posit plausible reanalyses to account for the distinct patterns in Anatolian.

So far, I have meticulously avoided this issue of the morphological form of Proto-Anatolian *REL. I argued in §3 that for the purposes of syntactic reconstruction we should treat *REL as a unitary category and remain agnostic on its form in PIE. Clearly, however, at least by the end of the period of Anatolian unity, it was undoubtedly *k^wi-, given its outcomes across the attested Anatolian languages. One could also argue that the developments that distinguish Anatolian from NIE may best be explained where there is identity between interrogative, indefinite, and relative pronouns, understood collectively as "*wh*-words"; such an argument has been made synchronically for Hittite by Huggard (2015), and further studies may seek to investigate the diachronic implications of this hypothesis within Proto-Anatolian.

I still do not see this as comprehensive proof that *REL in PIE_1 was of the form *$k^wó$-. Clearly Proto-Anatolian presents a set of innovations in RC syntax that are quite distinct form those found in NIE. If we accept that these changes were made possible by having a relative pronoun of the form *$k^wí$-, this alone does not tell us whether it was an innovation or an archaism. It is also worth noting that the changes attested in Anatolian are not found in Latin, which also has a relative pronoun in the form *$k^wí$-/*$k^wó$-, which suggests that the form of the pronoun is, at the very least, not a sufficient condition for the syntactic changes attested in Anatolian; clearly other grammatical factors are at play that cannot be explained by the presence or absence of this morphological form. Finally, I do not believe that the enclitic status of 'determinate' *kuiš* is definitive proof that it is of the same grammatical origin as indefinite *k^wi-, since the Celtic languages also show the development of an enclitic relative pronoun which happens to be of the form **yo*-. I thus continue with the hypothesis developed in §3 that PIE had one relative pronoun, *REL, whose surface form I remain agnostic towards, but which was a full lexical word.

depending on where it ends up in the derivation.

[158] I am not aware of any enclitic relative pronouns that have become full lexical items in any other languages. Within PIE, where decliticisation of pronouns is attested (e.g., various developments of indefinite *k^wo-), it requires morphological embellishment of the sort that is not usually found for relative pronouns, whether from *$k^wó$- or **yó*-

To conclude this section: I find no compelling evidence from either Vedic, or Celtic, or Anatolian, that PIE *REL was a clitic. I therefore proceed with the assumption that *REL was a full lexical item, occupying a specifier position when fronted to the left periphery.

5.1.2 Fronted *REL

Given the conclusion in §5.1.1 that PIE *REL was a full lexical item, its appearance in the left periphery cannot be accounted for under any variation of "Wackernagel's Law" (see §4.2). The question remains, when *REL was fronted to the left periphery, where did it land? All the evidence I have cited so far across §2–4 point to one position: [Spec, TopP]. A subset of the evidence points quite specifically to the lower [Spec, TopP] in Rizzi's (1997) model. I will summarise the evidence here.

The first set of arguments come from the Minimalist theory of RCs discussed in §2.2. There is clear cross-linguistic evidence, as noted by Bianchi (1999) that fronted elements in some languages can precede left-peripheral relative pronouns, *contra* the original observations made by Rizzi (1997) for Italian. It seems possible, therefore, that the landing site of the relative pronoun varies cross-linguistically, and so comparative evidence from the relevant ancient IE languages is crucial. I have argued that there is clear distributional evidence from several archaic IE languages that fronted *REL occupied the **lower** [Spec, TopP]. I have made the case most thoroughly for Vedic Sanskrit (§4.3). In the first place, Skt. relative *yá-* can be preceded by other Topics and Foci; it therefore cannot occupy [Spec FocP], nor the higher [Spec, TopP]. However, its position in the left periphery is secured by the fact that *yá-* regularly precedes WL1. I established that when they appear in "second position", WL1 strictly target C^{o} or the lowest active Top^{o}/Foc^{o}; if *yá-* were not fronted at all, or occupied some lower position, we would expect it to be preceded by WL1, as is potentially the case for correlative demonstratives (to be discussed in §5.2.3 below).[159] The comparative evidence suggests that this was the case in (at least) PIE_2. The evidence from Homeric Greek is more limited, but suggests relative ὅς occupies the lower [Spec, TopP] (see §4.4.1): as in Sanskrit, it can be preceded by fronted topics. There is no unambiguous evidence that it can be preceded by Foci. Ancient Greek does seem, however, to show the same rule for WL1 placement as Sanskrit. Conversely, while Latin shows a different rule for WL1 placement (target Foc^{o} if present), this allows us to establish some evidence from archaic texts that relative *qui* occupies the lower [Spec, TopP] (see §4.4.2). However, from Classical Latin onward this pattern is recessive, with *qui* occupying either [Spec, ForceP] (as argued by Rizzi, 1997 for Italian) or strictly the **higher** [Spec, TopP].

To these *comparanda*, I have added evidence from Celtic and Anatolian that may obliquely support the lower [Spec, TopP] hypothesis. Both these branches suggest evidence for extensive fronting around the relative pronoun at some stage, suggesting at the very least that a Classical Latin-style [Spec, ForceP] (or the higher [Spec, TopP] for that matter) is untenable; this is evidenced by the cliticisation of the relative pronoun, a process which is innovated within the attested history of Celtic languages and which I have argued similarly to be an innovation in Anatolian. [Spec, ForceP] is also comprehensively excluded for Hittite, as that is where we find the sentence connectors such as *nu*, which freely co-occur

[159] I am aware of only one example of apparently *in-situ yá-* which is preceded by WL1, RV 8.45.14, cited by Hock (1997). In all other examples he lists, where WL1 and apparently *in-situ yá-* co-occur, the former do not precede the latter.

with left-peripheral *kuiš*. I have argued, following Huggard (2015), that 'indeterminate' *kuiš* in Hittite occupied a distinct [Spec, ContrFocP] at the top of the left periphery (see §4.4.3). I have further argued, in §5.1.1.3 above, that a constituent in [Spec, ContrFocP] could occupy a surface position identical to that of clause-initial [Spec, TopP]. Given the lack of distributional evidence for a distinct ContrFocP projection outside Anatolian, I would argue that [Spec, TopP] was the original landing site for fronted *REL, and that its movement to a distinct [Spec, ContrFocP] in 'indeterminate' relatives was an Anatolian innovation resulting from a set of reanalyses detailed in §5.1.1.3. In short, fronted PIE *REL targeted the lower [Spec, TopP].

There is also the likely possibility that PIE allowed for a *REL that was not fronted at all; in such instances, the relative pronoun is hypothesised to remain in its base-generated pre-verbal position. Hock (1997: 114) lists examples such as the following for Sanskrit:

(5.12) *úd* *usríyāḥ* *jánitā* *yáḥ* *jajā́na*
forth.PW ruddy.ACC.PL.M creator.NOM.SG.M REL.NOM.SG.M create.PERF.3SG
'The creator who gave birth to the ruddy [cows]' (RV 3.1.12c)

While it is not impossible to treat *úd*, *usríyāḥ* and *jánitā* as fronted to various left peripheral positions, clearly a more parsimonious analysis is to hold that *yáḥ* remains in its base-generated *v*P-internal position. Hock (1997: 114, n.11) suggests there are at least 20 examples of this word order in the Rig Veda. Some of these may show genuine evidence for *wh-in-situ*, but others are not diagnostic owing to the small number of constituents:

(5.13) *divyā́ḥ* *ā́paḥ* *abhí* *yád* *enam* *ā́yan*
divine.NOM.PL waters.NOM.PL toward.PW REL him.ACC come.IMPF.3PL
'When the waters came upon him' (RV 7.103.2a)

I have dealt with this example already in §4.3.2 (as ex. 4.37). Hock lists it as an example of pre-verbal (i.e., *in-situ*) *yá-*: Lowe (2014: 34) treats it as an example of supposedly enclitic (and therefore moved) *yá-*. I treat it as an example of multiple fronting, with the higher [Spec, TopP] filled by *divyā́ḥ ā́paḥ*, [Spec, FocP] filled by *abhí*, and *yád* occupying the lower [Spec, TopP]. Clearly, at a surface level, each of the analyses could generate the attested order. There is no way of disambiguating *prima facie* between an element that is fronted and just happens to be pre-verbal, and one that it is *in-situ*. The only clues are (a) the number of constituents, and (b) the position of clitics (see §4.3). And this kind of ambiguity is not by any means unique to Sanskrit. Take the example I provided for the possibility of fronting around the relative pronoun in Latin:

(5.14) *satis* *iam* *audivi* *tuas* *aerumnas,* *ad* *portum*
enough now hear.PERF.1SG your.ACC.PL.F tribulation.ACC.PL.F at harbour.ACC.SG.M
=***mihi*** *quas* *memorasti*
=me.DAT.SG REL.ACC.PL.F remind.PERF.2SG
'I've heard enough of your tribulations, which you told **me** about at the harbour'
(Plaut. *Capt.* 929)

It is perfectly plausible, in this instance, to treat *quas* not as fronted to [Spec, TopP], with *ad portum* fronted to [Spec, FocP], but rather with *quas* occupying its base-generated position preverbally. The situation is further complicated by the fact we cannot be certain that *mihi* is enclitic; and even if it were, owing to Latin's Focus-based rules of WL1 placement (see §4.4.2), we would not be able to make any strong claims as to whether the relative pronoun is fronted or not. Such ambiguities abound also in Hittite, where the situation is perhaps uniquely unclear. While Huggard (2015) prefers a *wh-in-situ* analysis of 'determinate' *kuiš*, Hoffner & Melchert (2008: 425), citing Hale (1987: 49) and Garrett (1994: 46), note that 'in the vast majority of determinate RCs, the relative pronoun is preceded by only one syntactic constituent.' This would permit a fronting-around-the-relative-pronoun analysis, i.e. one with *wh*-movement Yet, according to Huggard (2015: 156), 'missing from this...is the observation that the *wh*-phrase in these clauses is within close nexus to the finite verb.' To Huggard (2015), this suggests precisely an absence of *wh*-movement, which he ties into the grander notion that Hit *kuiš* must remain within the *v*P to be bound by the 'Rule of Existential Closure' (2015: 156).

Thus again, the distributional facts do not distinguish between fronted and *in-situ* behaviour in short clauses; our analysis is dependent on our observations and assumptions from elsewhere in the grammar. If we accept that fronting around the relative pronoun is possible, we may explain away all such instances as still instantiating *wh*-movement; on the other hand, if we treat non-initial pre-verbal relative pronouns as occupying their base-generated position, we may lose some of our putative evidence in favour of multiple fronting. Not for the first time, our analysis can be stuck in a cycle where antithetical assumptions result in antithetical conclusions. I would tentatively argue that in a language such as Latin or Sanskrit, which shows fairly unambiguous evidence for *wh*-movement—even multiple *wh*-movement—we are justified in attempting to explain such tokens with recourse to multiple fronting. It is not so clear, however, that the same argument could be made for Hittite 'determinate' relatives, as argued by Huggard (2011, 2015). On the other hand, if we want to derive Hittite 'indeterminate' RCs diachronically from a PIE_1 RC—and I believe that we should—the possible lack of *wh*-movement in 'determinate' relatives does not exclude the possibility of *wh*-movement at an earlier stage of development, before the grammaticalisation of determinacy according to the position of the relative pronoun.

More generally, as I noted in §3.3.3, *wh*-movement is a process that can come and go from a language, both in RCs and interrogative clauses. The presence of *wh*-movement is also not a singular, binary parameter; languages can have multiple strategies when it comes to the placement of *wh*-words (Lutz *et al.*, 2000). And as far as *REL is concerned, in Minimalist terms, *wh*-movement is not strictly a feature of the pronoun itself, but rather a feature of the functional head, conventionally modelled as C^{o}. According to the the theoretical assumptions discussed in §1.3, if the relevant functional head (e.g., C^{o}) bears a **strong** uninterpretable feature (*u*REL, *u*INT etc.), it triggers movement of the corresponding pronoun to the left periphery; if the feature is **weak**, it will not trigger movement. I would suggest that in PIE, we could capture the availability of *wh-in-situ* by stipulating that the feature [*u*REL] on C^{o} (or, more precisely on Top^{o}) can be either weak or strong. I would argue it was predominantly strong, and it is for this reason that I believe the position of *REL within the left periphery is a crucial component of understanding PIE relative syntax; however, it is certainly possible, perhaps even likely, that the

[*u*REL] feature on Top° was optionally weak. This variability is attested elsewhere (e.g., Hindi) without any particularly clear pragmatic implications (Mahajan, 2000). For this reason, I will incorporate both options—strong and weak [*u*REL] on Top°—into my model of PIE. I offer no speculation as to whether a putative pre-PIE was *wh*-in-situ and gained *wh*-movement, as I believe we have sufficient grounds to accept that *wh*-movement is inherited into the attested languages some form.[160] My views on this matter are therefore quite conventional, ultimately aligning with Hale (1987), Hock (1989) Garrett (1994), Kiparsky (1995), Hoffner & Melchert (2008), *inter alia*.

5.2 Correlatives

Given the abundance of correlative clauses in ancient IE languages, cited throughout this thesis, it seems clear that some form of correlative clause was available in PIE; indeed, while many authors have explicitly reconstructed correlative clauses for PIE (Sturtevant, 1930; Haudry, 1973; Hale, 1987; Hock, 1989; Hajnal, 1997; Kiparsky, 1995; Probert, 2015), scarcely any authors explicitly argue that PIE **lacked** a correlative clause, at least in some form. In this section, therefore, I address some of the more specific syntactic questions on the nature of correlatives in PIE.

5.2.1 CP vs. DP?

In §2.2, I suggested that correlative relative could either be DPs or CPs cross-linguistically. Arguing that correlatives where always CPs, de Vries (2002: 40) lists the following criteria:

(a) Correlative RCs do not occur in DP positions
(b) Correlative RCs never have an external determiner
(c) Correlative RCs never have an external Case ending or another nominal marking
(d) Correlative RCs never have an external (affixed) adposition

However, Cinque (2020: 139–41) argues that correlative RCs should instead by analysed as left-dislocated DPs, rather than CPs, going so far as to claim the DP status of correlatives was 'confirmed'. I suggested that this could be something that varies cross-linguistically, and therefore something that must be established for PIE. However, the evidence is all but certain: PIE correlative RCs were CPs. To my knowledge, this has been accepted, either implicitly or explicitly, by all authors working on the topic of correlative clauses in ancient IE languages. Even Garrett (1994), who treats Hittite correlative RCs as occupying a 'Topic' slot in the left periphery of the MC, analyses them as the category S", and notes explicitly that 'relative clauses occupy the Topic position, which NPs also sometimes fill' (Garrett, 1994: 45).

The simple facts are that de Vries' (2002: 40) criteria apply across the board to ancient IE languages. Point (b) is true vacuously insofar as ancient IE languages do not have grammaticalised articles.

[160] I also do not make any claims as to whether PIE had **interrogative** *wh*-movement, as that lies beyond the scope of my reconstruction.

Point (c) is also true, and somewhat vacuously: Cinque's apparent exception is the appearance of a specialised nominal topic marker in Isbukun Bunun (2020: 140). Of all the elements that could be considered overt topic markers in ancient IE languages (e.g., Hit. -(*m*)*a*, AGk. μέν etc.), none are restricted to nominal uses. As for point (d), since it is very likely the case that PIE did not have a properly grammaticalised set of adpositions, but rather 'local particles' or P-words (see, e.g., Dunkel, 1979; Bertrand (2014); Reinöhl (2016); Ram-Prasad, 2018), this diagnostic may also turn out to align with de Vries (2002) vacuously.

Thus the only point on which there is some real contention as far as PIE is concerned is (a): that correlative RCs do not occupy DP-positions. Cinque (2020) argues that the lack of occurence in DP positions is not a useful diagnostic because simple DPs can also be fronted and subsequently resumed by a pronoun, in the same way as a correlative RC. I do not find this a particularly convincing argument, because simple DPs clearly can occupy 'DP-positions', and their left-dislocation (+ resumption) is an optional, pragmatically conditioned process. Cinque's (2020: 140) argument seems to be that correlative RCs could be interpreted as extraposition because extraposition leads elements to occur outside 'DP-positions'. But in such instances, we have positive evidence that the left-dislocated element **can** occupy a DP-position; the problem with correlative RCs is that they can **only** occupy an extraposed (i.e., non-DP) position. In other words, if something **never** occupies a DP-position, on what grounds do we assert it is 'dislocated' at all? For this reason, I see the fact that correlative clauses in IE languages consistently occur at the periphery of the MC and so do not occupy DP positions as consistent with de Vries' first criterion.

There is some further evidence that ancient IE correlative RCs were CPs in the form of case phenomena. In correlatives in archaic IE languages both the relative pronoun and the head noun (if present) in correlative relatives regularly surface with the case assigned **internal to the RC**, regardless of the role of the correlative anaphor in the MC. This is exemplified below with Sanskrit:

(5.15) ***yáyā*** *gắḥ* *ākárāmahe* ***sénayā*** *agne* *táva*
REL.INS.SG.F cow.ACC.PL.M gather.PRES.1PL weapon.INS.SG.F agni.VOC.SG you.DAT.SG
utyắ, ***tắm*** *naḥ* *hinva* *magháttaye*
help.INS.SG.M DEM.ACC.SG.F us.ACC.PL spur-on.IMPV.2SG bounty.DAT.SG.F
'**By which weapon** we make the cows our own with your help, O Agni, spur on **that** [weapon] for the bounty' (RV 10.156.2)

For Cinque (2020: 297–8), this behaviour appears to be subsumed under the category of 'inverse attraction' (discussed above), which he notes is often explained by adopting the HRA-style analysis over the MA analysis; this is still theoretically compatible with Cinque's correlatives-as-DP hypothesis.[161] Yet unlike other 'attraction' phenomena, the assignment of case to the head noun (+ relative pronoun) within the correlative RC is the **regular** pattern. If we were to adopt Cinque's hypothesis that correlative RCs are left-dislocated DPs, we might expect it to be at least possible for the head noun in the correlative RC to bear the case assigned by the MC. Indeed, the regular behaviour for topicalised constituents in ancient IE languages is that their case features are unaffected by fronting to the left periphery. Failing that,

[161] Elsewhere, Cinque (2020: 132) seems to treat these phenomena under the same umbrella as a lack of 'case connectivity' observed in German-style contrastive left-dislocation.

perhaps we would observe the head noun occurring in a case used especially for left-dislocation/hanging topics (e.g., the *nominativus pendens*). This case behaviour is not attested with correlative RCs at the earliest stages of IE languages; as noted above, case is regularly assigned to the head noun and the relative pronoun strictly **within** the RC. If we model the correlative RC as an independent CP, this behaviour is exactly what we expect: the head noun and relative pronoun are straightforwardly assigned the case within the RC, which is adjoined to the MC.

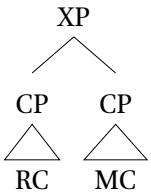
Given all the above, with the additional possibility for multiple-headed correlatives, for which even Cinque (2020: 135–8) admits a CP analysis (see §2.2.2), I do not see any convincing reasons as to why we should abandon the *communis opinio* that correlative RCs in ancient IE languages were CPs rather than DPs. On the basis of identity, therefore, I see it as a fairly simple step of extending this analysis to the proto-language. As such, I will continue with the traditional view that **correlative RCs in PIE were CPs.** This is not necessarily to say that PIE lacked DP-level RCs altogether: that is a question I will return to in §5.4 below.

5.2.2 Adjunction level

Since we have established that PIE correlative RCs were CPs, the next question is: what is the nature of the syntactic relationship between the correlative relative CP and the MC? In theoretical terms, we may rephrase the question as: at what level does the correlative relative CP adjoin to the MC?

The adjunction site for correlatives appears to be something that varies cross-linguistically. Dayal (1996) and Bhatt (2003) demonstrated that Hindi correlative relatives adjoin at the IP level, and no higher. On the other hand, Arsenijević (2009: 146–7) suggests that in Serbo-Croatian and Hungarian, adjunction is also possible at the CP level.[162] As for the ancient IE languages, this question has received sporadic attention over the last three decades; more or less all the facts suggest that correlatives CPs adjoin to the matrix CP, as represented below:

(5.16)



There is a long history of seeing ancient IE languages, especially Sanskrit and Hittite, as somehow more inclined to loosely connected, paratactic structures, rather than the hypotactic ('embedded') structures more common in contemporary IE. I do not make any general claims in this regard; but a central *topos* for this hypothesis is the formation of correlative clauses. By their nature, in a correlative clause the RC sits at the sentence periphery, either left or right of the MC but not within it (this is broadly what is meant by *parataxis*). In generativist terms, this implies some sort of high-level adjunction, but it does not necessarily distinguish between IP-level adjunction and CP-level adjunction. In order to do that, one

[162] See also Izvorski (1996) and Citko (2009) for Slavic languages, and den Dikken (2009) for Dutch.

must first establish some diagnostics, applicable to the relevant ancient IE languages, for differentiating between a CP and an IP.

To address this issue, Hock (1989) takes up the question of adjunction level for correlatives in Sanskrit. Hock suggests that Sanskrit correlatives are formed of two adjoined CPs, by virtue of a syntactic 'mirroring' effect, whereby both the RC and the MC have the same structure.[163] He takes the example:

(5.17) *yám* *u* *ha* *evá* *tát* *paśávaḥ* *manuṣyèṣu* *kắmam*
REL.ACC.SG.M EMPH EMPH EMPH then cattle.NOM.PL.M men.LOC.PL.M desire.ACC.SG.M
árohan *tám* *u* *ha* *evá* *paśuṣu* *kắmam*
attain.IMPF.3PL DEM.ACC.SGM EMPH EMPH EMPH cattle.LOC.PL.M desire.ACC.SG.M
rohati
attain.PRES.3SG
'What very desire the cattle then obtained among men, that very desire he (now) obtains among cattle.' (ŚB 2.1.2.7)

Hock (1989: 94) suggests the other possible forms the sentence could take:

(5.18) [*yám$_i$ ... kắmam$_i$ árohaṁ*s]$_{RC}$ [*tám$_i$... rohati*]$_{MC}$

(5.19) [*yám$_i$ ... árohaṁ*s]$_{RC}$ [*tám$_i$ ... kắmaṃ$_i$ rohati*]$_{MC}$

(5.20) [*yám$_i$ ... árohaṁ*s]$_{RC}$ [*tám$_i$ ... rohati*]$_{MC}$

(5.21) [*tám$_i$ ... (kắmam$_i$) arohan*]$_{MC}$ [*yám$_i$ ... (kắmaṃ$_i$) róhati*]$_{RC}$

(5.22) a. [RP (...) (head)]$_{RC}$ [(CP) (...) (head)]$_{MC}$

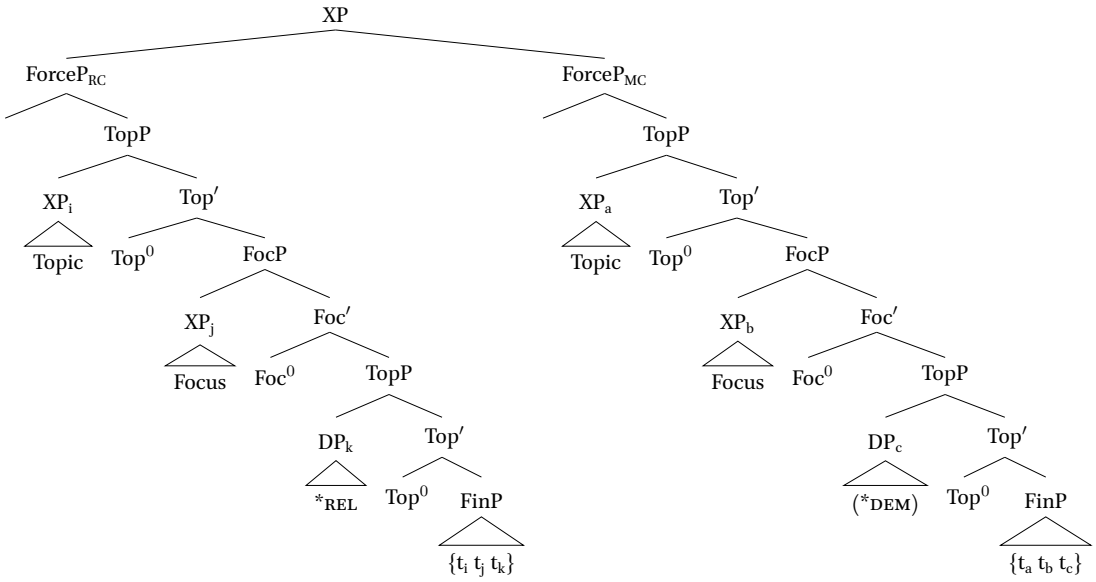
b. [(CP) (...) (head)]$_{MC}$ [RP (...) (head)]$_{RC}$

Hock (1989) observes that in Sanskrit correlatives, both the RC and the MC instantiated a complete **initial string.** The argument follows that because the initial string contains 'sentence-oriented particles', it characterises a CP. As the author later notes in Hock (2015: 7), a certain amount of revision is necessary to interpret his argument within a more contemporary syntactic framework, within which his original argument is made all the more convincing.

In §4, I formalised the initial string in terms of the fine-structure of the left periphery proposed by Rizzi (1997). This means that, beyond the evidence suggested by 'sentence-oriented' particles alone, the entire initial string is located schematically within the CP-layer, spanning the Topic-Focus complex. Clearly, therefore, both the RC and the MC contain their own left periphery and are therefore, minimally, CPs. This is represented below, including the positions of *REL and the correlative demonstrative *DEM:

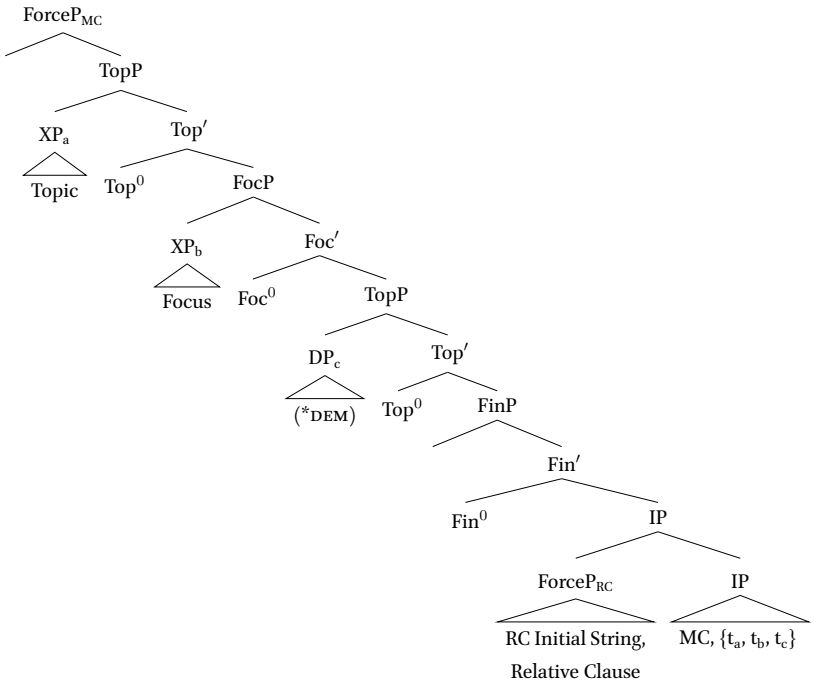
[163] In Hock's paper, he calls this 'conjunction', as opposed to 'adjunction'. This is because he takes 'adjunction' in this case only to refer to theories where the RC is adjoined lower in the structure (probably IP), such as those of Andrews (1975) and K. Hale (1975). In Minimalist terminology, whether the RC attaches to IP or CP, it will always be referred to as 'adjunction'.

(5.23)



Adjunction to IP, as evidence in Hindi is categorically excluded, since this would suggest the possibility for the initial string of the MC to precede the initial string of the RC, with all non-left-peripheral material following:

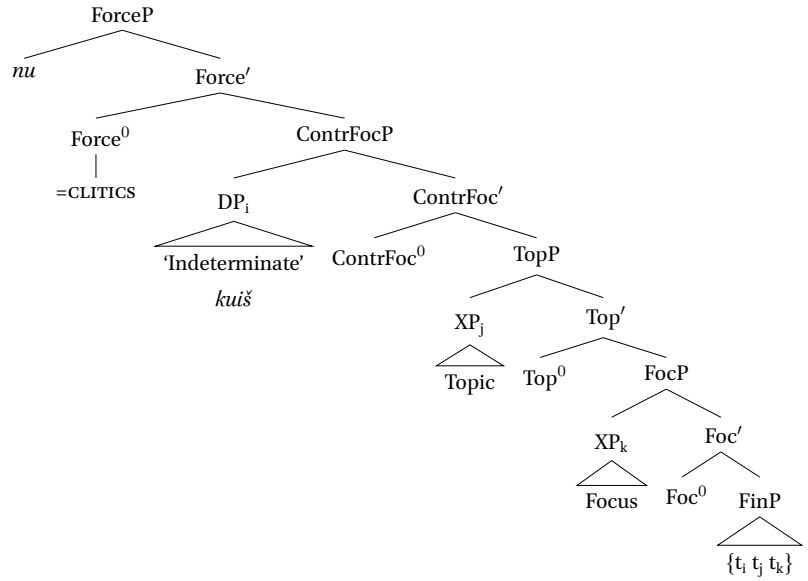
(5.24)



(5.25) *Topic$_{MC}$ – Focus$_{MC}$ – (DEM) – Clitics$_{MC}$ – Topic$_{RC}$ – Focus$_{RC}$ – REL – Clitics$_{RC}$ – Remainder$_{RC}$ – Remainder$_{MC}$

I find this to be conclusive evidence that in Sanskrit, correlative RCs adjoin to (at least) CP. The evidence is equally striking from Hittite. In §4.4 I also modelled Hittite as making extensive use of left-peripheral positions. In contrast to Sanskrit, Hittite very often has elements in [Spec, ForceP], viz. the sentence connectors, and in Force°, viz. the clitic string:

(5.26)



As noted by Hoffner & Melchert (2008: 391–2), sentence-connectors such as *nu* can introduce both correlative RC and MCs from the very earliest stages of the language, as in the following:

(5.27) *nu* *kuit* [(LU)]GAL-*uš* *tezzi* *nu* *apāt* *iyami*
CONN REL.ACC.SG.N king.NOM.SG.M say.PRES.3SG CONN DEM.ACC.SG.N do.PRES.1SG
'Whatever the king says, that I do' (KBo 17.4 ii 12–13)

As such, the Hittite evidence allows us to confirm not only that both halves of a correlative (RC and MC) not only have a Topic-Focus complex, but in fact a complete CP-layer extending up to ForceP.

The evidence from Latin and Greek is consistent with this hypothesis, though it is slightly more difficult to interpret. In the first place, while I established some ordering constraints in §4.4, the left periphery at the earliest stages of these languages does not exhibit the same level of elaborate grammaticalised regularity as either Hittite or Sanskrit. Nevertheless, in Ancient Greek it is also clear that the correlative RC has CP status, given the patterns of WL1 and movement that point to an articulated left periphery. However, the CP-status of the correlative RC is not in itself diagnostic of adjunction. Even embedded, 'plain' relatives have a CP layer (see §2.2). The key diagnostic is actually the left periphery of the correlative **MC**: if all left-peripheral MC elements follow the correlative RC, the RC itself cannot be adjoined lower than CP. There is some evidence for this from constructions such as the following in Homer:

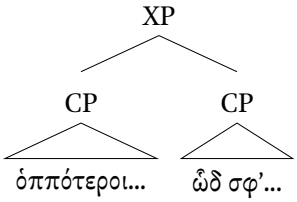
(5.28) [RC ὁππότεροι πρότεροι ὑπὲρ ὅρκια πημήνειαν] / [MC ὣδέ
REL.NOM.PL.M first.NOM.PL.M over oath.ACC.PL.N ruin.AOR.OPT.3PL thus
σφ' ἐγκέφαλος χαμάδις ῥέοι ὡς ὅδε
them.DAT.PL brains.NOM.SG.M on-the-ground flow.PRES.OPT.3SG as DEM.NOM.SG.M
οἶνος]
wine.NOM.SG.M
'**Whoever** will first cause harm, breaking their oaths, may **their** brains stream along the ground thus, just as this wine' (Hom. *Il.* 3.298–9)

This sentence is particularly complex as it instantiates two correlative-relative pairings: one nominal (ὁππότεροι...σφι) and one adverbial (ὣδε...ὡς). Both halves of the adverbial correlative are contained within the MC of the nominal correlative. Crucially, the correlative adverb ὣδε ('thus'), which hosts the WL1 pronoun σφι ('their'), appears to have been fronted to the left periphery of its clause: and since this clause is the MC of the pronominal correlative MC, if the adjunction site were any lower than CP, we might expect ὣδε to appear preceding the pronominal correlative RC, producing the following:

(5.29) *[$_{\text{MC (left periphery)}}$ ὣδέ σφ' [$_{\text{RC}}$ ὁππότεροι πρότεροι ὑπὲρ ὅρκια πημήνειαν], [$_{\text{MC(remainder)}}$ ἐγκέφαλος χαμάδις ῥέοι ὡς ὅδε οἶνος]]

Since this is not the attested order, it seems more straightforward to posit once again that both halves of the correlative are CPs:

(5.30)



The situation in Latin is more obscure: unlike the other languages discussed so far in this section, the correlative demonstrative is always placed initially within the correlative MC; while this pattern is dominant elsewhere, especially in Sanskrit, it is not exceptionless. Given the secured existence of IP-level discourse-oriented positions in Latin (Devine & Stephens, 2006; Danckaert, 2012), it could be the case that IP-level adjunction was allowed. Moreover, Latin correlatives exhibit certain interpretational constraints that do not apply in, e.g., Sanskrit and Hittite, such as the fact that Latin correlatives are always **maximalising** (Pompei, 2011b); yet this is also the case for correlatives with RC-MC order in Ancient Greek (Probert, 2015), and so may not be a useful diagnostic. On the other hand, the behaviour of Latin correlative RCs could tie into the advanced development of the 'plain' relative in Latin, which I will return to in §5.3. In any case, Latin does not appear to provide any proof that correlative RCs did **not** adjoin to CP (e.g., locality effects of the sort found in Hindi), and so does not provide any substantive challenge to the hypothesis of CP-adjunction suggested by the other archaic IE languages, even if it could also be compatible with an IP-adjunction hypothesis.

In short, we appear to have identity across the ancient IE languages that correlative RCs do not adjoin to IP: at a minimum, they adjoin as high as ForceP. The question remains, however: are they adjoined at all? Could it not simply be the case that we have **genuine** parataxis here, i.e., there is no syntactic relation between the two halves of a correlative? A major distinction between ancient IE-style correlative clauses and those in languages such as Hindi is the relative position of correlative RC and MC, which seems to vary (with the exception of Hittite, to which I will return below). This variability would be a trivial outcome of the fact that there is no syntactic dependency between the two halves; their adjacency is determined only by pragmatic factors, and so is not bound by syntax-based rules of linearisation. To me, this seems unteneable in light of the fact that RCs are infelicitous in the absence of a

correlative MC; in other words, RCs cannot occur independently. This suggests there is some form of dependency relation between the two halves. Without locality constraints, however, it is not immediately clear where in the narrow syntax we should encode this dependency. I would argue, in a similar vein to Davison (2009) for Sanskrit, that the answer lies with the feature specification of the relative pronoun (PIE *REL).

Let us start with the less controversial aspects of correlative RC formation. In the first place, we may imagine that Top° bears an uninterpretable feature [*u*REL], while *REL bears the interpretable feature [*i*REL]. When the [*u*REL] feature on Top° is strong, *REL moves to [Spec, TopP] in order to check it (*wh*-movement). This is more or less the model proposed by Rizzi (1990: 67) and subsequent authors. However, *REL also bears a unique syntactic-semantic feature, which I will call **EXTERNAL REFERENCE** [EXREF].[164] The [EXREF] feature on *REL essentially requires that *REL is co-indexed with a DP **external** to its own CP. This can be satisfied by either (a) a correlative demonstrative, or (b) a full DP, with which *REL agrees in φ-features. [EXREF] must be fulfilled under normal pragmatic conditions.

The [EXREF] feature on *REL occupies a position that sits at the edge of syntax and semantics. [EXREF] is clearly sensitive to syntax in two ways: first, it cannot be satisfied by anything within its own CP, and must 'probe' outside. Second, it can only 'probe' as far as one CP to either the left or right; this separates *REL from other anaphoric pronouns, which can be separated from their antecedent by multiple CPs. However we interpret this sensitivity to syntax, [EXREF] does not seem *prima facie* to rely on other conventional syntactic locality relations such as dominance or c-command. These two features—[*u*REL] on Top° which must be checked by [*i*REL] on *REL, and [EXREF] on *REL which must be checked by a co-indexed nominal in a neighbouring CP—conspire to yield the basic form of the PIE correlative clause.

To reduce the complexity of [EXREF] as a feature, we could capture the 'neighbouring CP' constraint by positing a higher projection that dominates both the relative CP and the correlative CP, which I have so far labelled agnostically as XP. Whatever the nature of this XP, it is then possible to stipulate that [EXREF] only operates within it; when this XP is sent to the interfaces, unless [EXREF] has been satisfied the derivation crashes (a gentle crash). This has the advantage of capturing an exception I have not yet mentioned to the neighbouring-CP rule: [EXREF] can in fact be checked across multiple CPs, **as long as all intervening CPs are co-referential RCs.** This is attested in the appositive 'stacking' structures attested in Vedic:

(5.31)	***yáḥ***	*pṛthivím*	*vyáthamānām*	*ádṛṁhat*	***yáḥ***
	REL.NOM.SG.M	earth.ACC.SG.F	trembling.PPL.ACC.SG.F	make-firm.IMPF.3SG	REL.NOM.SG.M
	párvatān	*prákupitān*	*áramṇāt*	***yáḥ***	
	mountain.ACC.PL.M	shaking.PPL.ACC.PL.M	stay.IMPF.3SG	REL.NOM.SG.M	

[164]The precise theoretical status of this feature and its operation merits further investigation beyond the scope of the current discussion, especially in light of the varying understandings of feature theory within the Minimalist program. For the purpose of this thesis, my central claim is that however we model this feature, it is unique to the relative pronoun and responsible for its syntactic behaviour, differentiating it from other types of pronoun.

antárikṣam *vimamé* *váriyaḥ* ***yáḥ*** *dyā́m*
atmosphere.ACC.SG.M measure-out.PERF.3SG more-widely REL.NOM.SG.M sky.ACC.SG.M
ástabhnāt ***sá*** *janāsa* *índraḥ*
support.IMPF.3SG DEM.NOM.SG.M people.VOC.PL.M Indra.NOM.SG.M
'**Who** settled the trembling earth, **who** held fast the shaking mountains, **who** measured out the sky more widely, **who** held up the sky, **he**, O peoples, is Indra' (RV 2.12.2)

This apparent exception to the 'neighbouring CP' rule could be captured if all the RCs, together with the MC, fall under a singular XP projection.[165] Whatever the nature of XP, if present, I will hold that PIE *REL contained a feature, [EXREF], which required it to be co-indexed with either a DP or a correlative demonstrative in the MC. This will be central to my analysis in §5.3.

5.2.3 Correlative pronouns

I turn now to the position of correlative demonstratives in correlative clauses. In the first place, it is worth noting that the correlative pronoun is routinely an accented demonstrative that is fronted within the correlative MC. This option is attested across the languages I have considered:

(5.32) *ā́* ***tát*** *=te* *dasra* *mantumaḥ* *pū́ṣan*
toward.PW DEM.ACC.SG.N you.GEN.SG wondrous.VOC.SG.M wise.VOC.SG.M Pūṣan.VOC.SG.M
ávas *vṛṇīmahe* *yéna* *pitŕ̄n* *ácodayaḥ*
help.ACC.SG.N choose.PRES.1PL REL.INS.SG.N father.ACC.PL.M quicken.IMPF.2SG
'We choose for ourselves **that** support of yours, O wondrous and wise Pushan, by which you quickened our fathers' (RV 1.42.5)

(5.33) αἳ γὰρ ὑπ' ἠελίῳ τε καὶ οὐρανῷ ἀστερόεντι /
DEM.NOM.PL.F for under sun.DAT.SG.M and and sky.DAT.SG.M starry.DAT.SG.M
ναιετάουσι πόληες ἐπιχθονίων ἀνθρώπων / **τάων** =μοι
dwell.PRES.3PL city.NOM.PL.F earthly.GEN.PL.M man.GEN.PL.M DEM.GEN.PL.F =me.DAT.SG
περὶ κῆρ τιέσκετο Ἴλιος ἱρὴ
around heart.ACC.SG.N be-honoured.IMPF.3SG Troy.NOM.SG.F sacred.NOM.SG.F
'For of all cities of earthly men which lie beneath sun and starry heaven, of **these** sacred Troy was most honoured in my heart' (Hom. *Il.* 4.44–6)

(5.34) KUR.KUR.ḪIA *kue* *dannatta* *ammuk* EGIR-*pa* *ašešunanun*
lands.N REL.ACC.PL.N empty.ACC.PL.N I.NOM again settle.PRET.1SG
nu=mu=kan ***apē*****=ya** *ḫūmanda* *arḫa* *dāš*
CONN=me.ABL=LP DEM.ACC.PL.N=and all.ACC.PL.N away take.PRET.3SG
'The lands which I resettled, all **those** he took away from me too' (Ḫatt. iii 57–57)

(5.35) *cui* *testimonium* *defuerit,* ***is*** *tertiis*
REL.DAT.SG.M testimony.NOM.SG.N be-lacking.FUT.PERF.3SG DEM.NOM.SG.M third.ABL.PL.M
diebus *ob* *portum* *obvagulatum* *ito*
day.ABL.PL.M before door.ACC.SG.M demand.SUP go.IMPV.3SG

[165] Indeed, under this theory, the putative functional head X^{o} could itself bear a syntactic feature relevant to the operation of [EXREF].

'Who(ever) lacks testimony, let **him** go to the door (of the witness) every third day to demand it' (Twelve Tables II.3)

Additionally, in Hittite, Ancient Greek and Sanskrit, it is possible for the correlative pronoun to be a clitic (WL1), as in the following:

(5.36) *své* *á́* *yáḥ* *túbhyam* *dáme* *á́* *vibhā́ti*
one's-own.LOC.SG.M to.PW REL.NOM.SG.M for-you house.LOC.SG.M to.PW shine.PRES.3SG
... várdhaḥ *agne* *váyas* *=**asya*** *dvibárhāḥ*
... increase.INJ.2SG Agni.VOC.SG.M age.ACC.SG.N =his doubly-strong.NOM.SG.M
'He who shines forth praises for you in his own house... O Agni the double-mighty, **his** vitality you increase' (RV 1.71.6)

(5.37) [RC ὁππότεροι πρότεροι ὑπὲρ ὅρκια πημήνειαν] / [MC ὣδέ
REL.NOM.PL.M first.NOM.PL.M over oath.ACC.PL.N ruin.AOR.OPT.3PL thus
σφ' ἐγκέφαλος χαμάδις ῥέοι ὡς ὅδε
them.DAT.PL brains.NOM.SG.M on-the-ground flow.PRES.OPT.3SG as DEM.NOM.SG.M
οἶνος]
wine.NOM.SG.M
'**Whoever** will first cause harm, breaking their oaths, may **their** brains stream along the ground thus, just as this wine' (Hom. *Il.* 3.298–9)

(5.38) *nu* DUMU-*an* *kuin* *ḫukkiškemi* *n=**an**=kan* *ŠUM=ŠU*
CONN child.ACC.SG.C REL.ACC.SG.C conjure-spells.PRES.1SG CONN=**him**.ACC=LP name
tēmi
say.PRES.1SG
'The child I am casting a spell over, **him** I say the name. (KUB 7.1 i 7)

That this possibility does not seem to be attested in Latin is probably down to the apparent lack of third person clitic pronouns (with the possible exception of reflexive *se*). Indeed, outside Hittite, this possibility is unavailable in cases where the correlative pronoun functions as the subject of the MC, owing to the lack of nominative clitics. I will return to this matter below.

From a syntactic perspective, the position of fronted enclitic correlative pronouns is accounted for by the rule of WL1 placement established in §4.3: namely, they raise to C° or the lowest active Top° or Foc°, and are subject to PI (i.e., part of "Wackernagel's Law"). As for the accented demonstrative forms of the correlative pronoun, things are somewhat less clear. In the first place, we may imagine that the correlative pronoun is often fronted for pragmatic reasons, moving to [Spec, TopP] or [Spec, FocP]. This would account for their very common location clause-initially. That one of these positions is implicated appears to be secured in Sanskrit by the occurence of correlative *tát* preceding WL1 even when it is non-initial, as in ex. (5.32). In the absence of any other left-peripheral material, WL1 are not diagnostic of a specific left-peripheral syntactic position as even if they raised to a unitary C° they would be pushed behind the first phonological word by PI. Yet the occurence of *=te* following *tát* in **third** position suggests that *tát* itself occupies the spec of the lowest active left peripheral head: in this case, [Spec, TopP]. This supports the arguments of Hock (1989) and Lowe (2014) that distribution of (accented) correlative

pronouns "mirrors" the distribution of relative pronouns. However, as noted by Hale (1996), while left-peripheral relative pronouns regularly precede WL1, correlative demonstratives often follow WL1, as in the following:

(5.39) *yáḥ* *mártyaḥ* *śíśīte* *áti* *aktúbhiḥ,* *mā́* *=naḥ*
REL.NOM.SG mortal.NOM.SG sharpen.PRES.3SG over night.INS.PL NEG =US.ACC
sáḥ *ripúḥ* *ī́śata*
DEM.NOM.SG rogue.NOM.SG rule.INJ.3SG
'The mortal who sharpens his weapons through the night, let not **that** rogue rule over us'
(RV 1.36.16cd)

This suggests that correlative demonstratives may also occupy a position lower than that of fronted relatives, while still moving to the left periphery. I suggested fleetingly that this position might be [Spec, FinP], which is a common position for 'weak' pronouns (in the sense of Cardinaletti & Starke, 1999), as discussed in Roberts (2021). This incidentally would suggest that Fin° does not constitute a viable goal for WL1: as such, we could rephrase the rule for their positioning as "**target the lowest active left-peripheral head above Fin°**". Returning to 'weak' pronouns, Cardinaletti & Starke (1999) note that they may be subject to speech connection processes that 'strong' pronouns are not, but they retain their lexical accent; this could account for the retroflexion discussed by Lowe (2014) without resorting to syntactic clisis (see §4.2.2). This merits further investigation from a Sanskrit perspective, but I do not pursue it here.

Of more interest from a PIE perspective is that the very possibility of correlative pronouns being enclitic supports the hypothesis that correlatives were an unmarked relativisation strategy in the language. Even in traditional terms, clitics, by definition, are not used for any sort of "emphasis"; in Minimalist terms, this is partly encoded in the fact that they cannot occupy [Spec, TopP] or [Spec, FocP] themselves. In such clauses, the clitic pronoun in the correlative clause occurs simply to satisfy the [EXREF] feature on *REL (see §5.2.2 above). On the other hand, the fact that correlative pronouns could also take the form of accented demonstratives occupying [Spec, TopP] or [Spec, FocP] suggests that correlative clauses *could* be in some sense "emphatic". This is perhaps the more intuitive reading for speakers of languages in which correlatives are a marked form of relativisation (e.g., English); it is possible that some branches (e.g., Latin) continued only the "emphatic" style of correlative, preferring 'plain' relatives as the unmarked strategy. One could convincingly argue, therefore, that in Latin the correlative pronoun **must** be topicalised or focalised; without a sufficient clitic inventory, however, this point is not necessarily provable. As such, it seems likely to me that both resumption strategies—a clitic pronoun and an accented demonstrative—were available in PIE correlative clauses.

To end this section, I return to the question of subject clitics. If, as argued by Garrett (1996), nominative clitics are an Anatolian innovation, PIE had a paradigmatic gap in the clitic inventory. If it so happened that the correlative pronoun was a subject in the MC, there was no clitic available; in this scenario, the only option to fulfil [EXREF] overtly was to use a lexically accented demonstrative. One may ask, however: what about cases in which a correlative pronoun occurred in the nominative, but was not to be interpreted as topicalised/focalised? I.e., if the pragmatic conditions were such that an

oblique correlative pronoun would have been enclitic, what of the nominative? This is where the possible existence of a 'weak' pronoun may work. One could argue, for instance, ex. (5.39) shows just these conditions: a correlative pronoun that does not seem unambiguously to require an emphatic reading, but must nevertheless occur in the nominative. This may account for its apparent movement to [Spec, FinP], as diagnosed by its following WL1. However, this hypothesis is almost impossible to test: assuming that the unmarked position for subjects is something like [Spec, IP] anyway, there is absolutely no way to distinguish this from [Spec, FinP] in the absence of an overt complementiser in Fin°. There is also the possibility that [EXREF] could be fulfilled by a null *pro* (cf. §5.4 below), which would would render the correlative unnecessary in such cases. It is possible that an exhaustive analysis of non-initial, post-WL1 occurrences of *sá* may shed further light on the matter. For now, I leave open the possibility of 'weak' correlative pronouns in PIE.

To conclude this section: correlative pronouns in PIE could be either enclitic or accented demonstratives. In the case of the former, the correlative pronoun bore no particular pragmatic function, and surfaced only to satisfy the [EXREF] feature on *REL. When an accented demonstrative was used, it could be fronted to the left periphery in a way similar to that of *REL. Paradigmatic gaps in the clitic inventory may account for the varying employment of correlative clitics in the descendant languages.

5.3 Position of the head noun(s)

As noted for Sanskrit by Hock (1989) and Davison (2009), and Hoffner & Melchert (2008) for Hittite,[166] head nouns in correlative clauses can be generated in two places: within the MC, or within the RC. For clarity, I will adopt the following terms in this discussion: HN_{MC} refers to the head noun in the MC of a correlative; HN_{RC} refers to the head noun of the RC. The latter is what is most commonly referred to as the 'head noun' when discussing RCs. In Sanskrit, the surface structure may reflect either the presence of HN_{MC} or HN_{RC} or both (see exx. 5.18–22 above). A crucial observation made by Hettrich (1988: 578–81), discussed in Hock (1989) and Kiparsky (1995), is that while there is apparent variation between which of HN_{MC} or HN_{RC} is spelled out, there is a preference for the head noun to occur in the **first** clause, whether it is RC or MC. Thus we may schematise the possibilities as follows:[167]

(5.40) $[_{RC}$ *REL HN_{RC}...$]$ $[_{MC}$ *DEM...$]$

(5.41) $[_{MC}$ (*DEM) HN_{MC}...$]$ $[_{RC}$ *REL...$]$

While this is a preference in Sanskrit, in Hittite it appears to be a strict rule. In the first place, the order for correlative clauses (i.e. those which are unambiguously adjoined CPs with overt resumption) is exceptionlessly RC-MC in Hittite. Moreover, according to Hoffner & Melchert (2008: 424), the head noun in preposed correlatives is exceptionlessly found in the RC (i.e., HN_{RC}). We may choose to see the consistent surfacing of HN_{RC} as a feature of the RC type itself; on the other hand, the comparative

[166] See, e.g., Bhatt (2015) for the same point made in a Minimalist context.

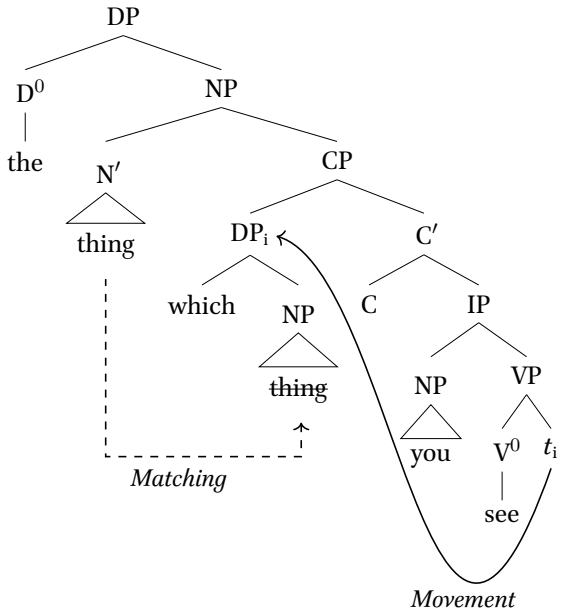
[167] In the following I have represented the HN_{RC} as movement to the left periphery together with *REL, but this movement is not obligatory: alternatively HN_{RC} can remain in its base-generated position, or even move above *REL.

evidence from Sanskrit may suggest that the surfacing of HN_{RC} is incidental to the rigid RC-MC ordering. In the absence of any correlatives with MC-RC ordering, it is hard to test this claim. However, evidence from postposed 'plain' relatives (see §5.3 below), suggested that the linearisation rule works; when the RC is postposed, the head noun surfaces (at least superficially) in the MC.

To summarise what I have said so far: at a base-level, correlative clauses allow for the occurrence of both HN_{RC} and HN_{MC}. These are not syntactic copies of each other, but independently base-generated DPs which happen to be identical. If only one of HN_{RC} or HN_{MC} surfaces, it will preferably be the left-most instance of HN, regardless of whether it is HN_{RC} or HN_{MC}. If there is a strong rule of RC-MC ordering (as in Hittite) the exceptionless occurence of HN_{RC} is **epiphenomenal**; the actual constraint is what I will call the **LEFTMOST HEAD NOUN RESTRICTION** (LHNR). This has some serious repercussions for our theory of 'plain' relatives to be discussed below. However, for now, it is worth dedicating at least some space to considering what the mechanism could be for the LHNR.

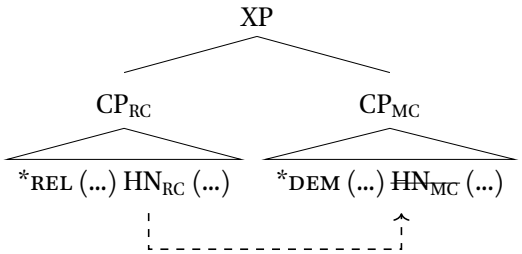
The facts of the LHNR as I have presented them so far bear a striking resemblance to the Matching Analysis (MA) of 'plain' RCs (see §2.2). As in my account for correlatives, the MA posits the existence of two distinct occurrences of the HN: one base-generated within the RC, and one base generated within the MC to which RC is adjoined:

(5.42)

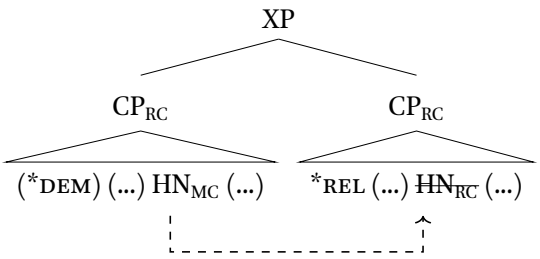


The two HNs 'match', and the lower instance is deleted at PF—that this happens at PF is significant, because it means deletion-under-identity is only sensitive to linear order and not constituent structure. This means that the fact I have modelled correlatives as adjoined CPs, apparently without any relation of syntactic dominance between the two, should not preclude a matching analysis. Moreover, an assumption about matching made explicit by Citko (2001: 138) is that 'deletion is less strict than movement with respect to identity of features, and that total identity is not required for deletion to be possible (van Orman Quine, 1962; Chomsky, 1965).' This means, crucially, that deletion-under-identity can take place even when there is case disagreement between the matches. As such, I propose the following options for correlative clauses:

(5.43) **RC-MC order:**



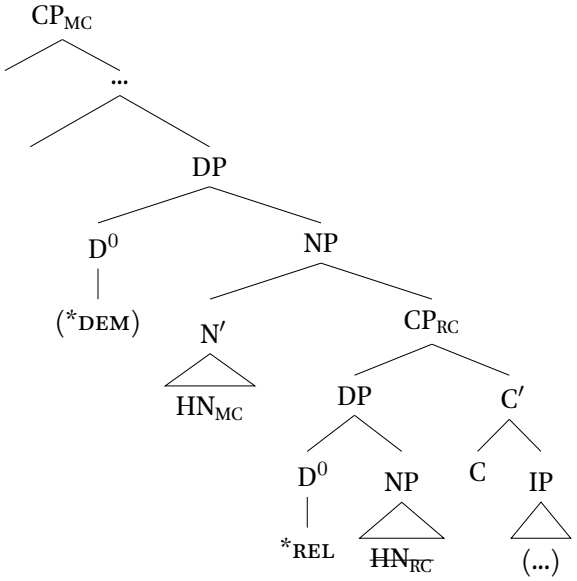
(5.44) **MC-RC order:**



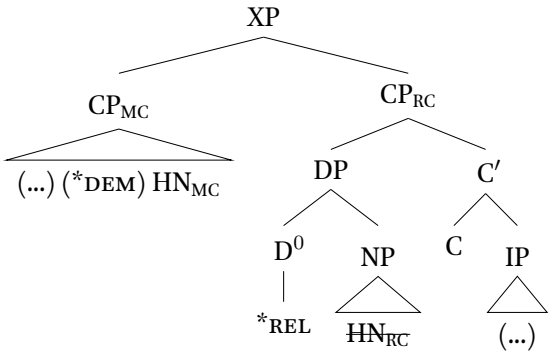
In each instance, the HNs will match (represented by the dashed arrows above) and the rightmost instance will be deleted. If, independently, there is a constraint on the ordering RC-MC (as there is in Hittite), HN_{RC} will always be spelled out and HN_{MC} will always be deleted. I do not believe PIE had such an ordering constraint, a point which I will develop in subsequent discussion.

The LHNR interacts with the [EXREF] feature on *REL in an interesting way. Recall that [EXREF] requires *REL to be co-indexed with **either** (a) a demonstrative pronoun, **or** (b) a full DP, in a neighbouring CP. In a correlative with RC-MC order, the LHNR will delete HN_{MC}. This means that in order for [EXREF] to be fulfilled, **the MC must use a demonstrative pronoun** to resume *REL. However, if the correlative had MC-RC order, HN_{MC} would be spelled-out, providing a sufficient goal for the [EXREF] feature on *REL, making the correlative demonstrative redundant. This is represented in exx. (5.43–44) above: with RC-MC order, *DEM is obligatory, while with MC-RC ordering, it is only optional (as it would be in a simple clause). The conspiracy of [EXREF] and the LHNR therefore generates what looks, superficially, like a 'plain' RC. In fact, when HN_{MC} occurs in clause-final position, the only difference between a correlative of the structure ex. (5.44) and a true 'plain' relative (under the MA) is where the relative CP is generated. Under the MA, it an adjunct (or complement) of HN_{MC}; in a correlative, it is adjoined to the MC at the CP-level. The two options are represented below:

(5.45) **Matching analysis for 'plain' relatives:**



(5.46) **Matching analysis for correlatives:**



In short, surface 'plain' relatives of the postnominal type can be generated by **precisely the same structural processes** as correlative clauses, the only genuine difference being the ordering of MC and RC, from which all else follows. This has significant repercussions for the way we reconstruct 'plain' relatives for PIE.

Before moving on, I note here that the obligatoriness of LHNR is not something that can be precisely established for PIE. As I noted above, LHNR is only a **preference** in Sanskrit. It is certainly possible that the same was true of PIE itself. From a theoretical standpoint, this is not too troubling: deletion-under-identity manifests in various ways cross-linguistically, to various degrees of obligatoriness (see, e.g., Allerton, 1975). It is also possible that LHNR emerges from general pragmatic principles, and is eventually grammaticalised as a key component of RC formation (as in the 'traditional' MA). This process merits further investigation; for now, I will continue to model LHNR as a 'rule', with the understanding that some rules are made to be broken.

5.4 'Plain' relatives

Every branch of IE shows some evidence for what I have so far called 'plain' RCs, where there is no clearly identifiable correlative DP, as in the following:

(5.47) *O Libane, uti miser est homo qui*
O Libanus.VOC.SG how sad.NOM.SG.M be.PRES.3SG man.NOM.SG.M REL.NOM.SG.M
amat
love.PRES.3SG
'O Libanus, how miserable is a man who's in love' (Plaut. *Asin.* 616)

(5.48) τίμησόν μοι υἱὸν ὃς ὠκυμορώτατος
honour.AOR.IMPV.2SG me.DAT.SG son.ACC.SG.M REL.NOM.SG.GM dying-earliest.NOM.SG.M
ἄλλων / ἔπλετ(ο)
other.GEN.PL.M go.IMPF.3SG
'Honour my son, who is destined for the quickest death before the others'
(Hom. *Il.* 1.505–6)

(5.49) *tvám mahā́n indra yáḥ ha śúṣmaiḥ*
you.NOM.SG great.NOM.SG.M Indra.VOC.SG.M REL.NOM.SG.M EMPH storm.INS.PL.M
dyā́vā jajñānáḥ pṛthivī́ḥ áme dhāḥ
sky.ACC.SG.M born.NOM.SG.M earth.ACC.SG.F terror.LOC.SG.M put.AOR.2SG
'You [are] great, O Indra who terrorised Heaven and Earth with your storms when you were just born' (RV 1.63.1ab)

(5.50) *nu* 8 DUMU.MEŠ-*uš uwadanzi* MUNUS-*ni=ššan kuiēš nāwi*
INTRO 8 boy.ACC.PL bring.PRES.3PL woman-towards REL.NOM.PL not-yet
pānzi
go.PRES.3PL
'They bring eight boys who have not yet gone to a woman' (KUB 9.31 ii 9–10)

However, 'plain' relatives include a number of different constructions that are attested to various degrees in the different daughter languages. The main point of variation is their syntactic position with regard to the main clause. 'Plain' RCs appear most usually to the right of the MC with which they are associated, as in the exx. (5.47–50) above. In Latin and Ancient Greek, we see centre-embedded RCs, which occur to the right of head noun but within the MC:[168]

(5.51) [$_{MC}$ *si vidulum illum* [$_{RC}$ *quem ego in navi*
if suitcase.ACC.SG.M DEM.ACC.SG.M REL.ACC.SG.M I in ship.ABL.SG.F
perdidi] / *cum auro atque argento salvom investigavero*]...
lose.PERF.1SG with gold.ABL.SG.N and silver.ABL.SG.N safe.ACC.SG.M find.FUT.PERF.1SG
'If I find that suitcase which I lost on the ship, safe with the gold and silver...'
(Plaut. *Rud.* 1339–40)

(5.52) [$_{MC}$ δαιμόνι' οὐκ ἄν τίς τοι ἀνὴρ [$_{RC}$ ὃς
sir.VOC.SG NEG MOD INDEF.NOM.SG.M you.DAT.SG man.NOM.SG.M REL.NOM.SG.M
ἐναίσιμος εἴη] / ἔργον ἀτιμήσειε
sound-minded.NOM.SG.M be.OPT.3SG deed.ACC.SG.N dishonour.AOR.OPT.3SG
μάχης]
battle.GEN.SG.F

[168] It is possible that centre-embedded RCs existed in Old Hittite too, but the evidence is quite fragmentary, cf. Probert (2006: 52–3) and Hock (2015: 65–6).

'Sir, no man who was in his right mind would fail to honour your work in the battle'

(Hom. *Il.* 6.521–2)

In such instances, the RC is unambiguously **embedded**; in generativist terms, this means the relative CP is adjoined to, or a complement of, the DP containing the head noun, as opposed to adjoined to the matrix CP, cf. the two analyses in ex. (5.45–6) above. As I noted in the previous section, however, when the RC is peripheral (i.e., no elements from the MC follow it), the structure is ambiguous at a surface level: it could either be a correlative clause (with the appropriate deletion of HN_{RC} and no overt correlative demonstrative), or an embedded relative. Without centre-embedding, therefore, there is some debate as to whether 'plain' relatives are genuinely embedded, or whether they are adjoined to the CP. For example, while Sanskrit and Hittite both exhibit 'plain' relatives, neither has clear evidence for centre-embedding. I will return to this matter shortly.

Another type of RC that falls under the general category of 'plain' relatives are what are generally referred to as **free relatives**. Free relatives by definition do not contain a head noun, but they behave syntactically as DPs (see, e.g., Alexopoulou, 2007), meaning they may occupy argument positions and are not resumed by a correlative demonstrative:

(5.53) *iouesat* *deiuos* ***qoi*** *med* *mitat* *nei*
call-to-witness.PRES.3SG god.ACC.PL.M REL.NOM.SG.M me.ACC.SG send.PRES.3SG if-not
ted *endo* *cosmis* *uirco* *sied*
you.ACC.SG in friend.NOM.SG.F maiden.NOM.SG.F be.PRES.SUBJ.3SG
'**The one who** sends me calls the gods to witness: if the young girl is not kind towards you...'

(CIL I² 4)

(5.54) αἰεὶ γάρ μοι ἔωθεν ἐνικλᾶν, [$_{FR}$ ὅττι
always for me.DAT.SG be-accustomed.PERF.3SG frustrate.INF REL.ACC.SG.N
νοήσω]
intend.FUT.1SG
'For she is always accustomed to frustrate **what(ever)** I have in mind' (Hom. *Il.* 8.408)[169]

Such free relatives must be carefully distinguished from headless RCs occurring in an RC-MC pair (as in exx. 5.28 and 35), which do not occupy argument positions, and are CPs. Accordingly, as with other types of 'plain' relative, embedded free relatives may be distinguished on the surface from headless correlative RCs only by the absence of a correlative demonstrative in the MC. Nevertheless, a DP-analysis of some free relatives appears to be evidenced directly by their co-ordination with other DPs:

(5.55) *avardhan* [$_{DP}$ *vā́jāḥ*] ***utá*** [$_{DP}$ *yé* *cit* *átra*]
increase.IMPF.3PL Vaja.NOM.PL.M and REL.NOM.PL.M INDEF there
'The [Vajas] **and** [whoever (else was) there] strengthened (you)' (RV 10.73.3b)

A DP-analysis for such clause types can hardly be denied for Latin and Ancient Greek, given the certified availability of embedded relatives (cf. exx. 5.51–2); to my knowledge, there is no synchronic

[169] Example from Probert (2015: 129).

analysis of free relatives in either of these languages that treats them as CPs. For Sanskrit, however, Hock (1989, 2015) holds that these are not genuine (embedded) free relatives, but instead something he calls a "replacive" relative. By this he means these clauses are (headless) correlatives in which the correlative demonstrative has been deleted. The strongest evidence in favour of of the "replacive" theory is that there is no independent evidence for embedded RCs in Sanskrit; we do not see centre-embedding, and correlative clauses with overt resumption remain an unmarked relativisation strategy (Hock, 1989). Yet Hock does not seem to give an account as to why this deletion, which is apparently possible in all branches of IE, does not displace correlatives in Sanskrit, yet leads to the development of embedded RCs elsewhere. An alternative analysis in one in which RCs such as ex. (5.55) are genuine, DP-level free relatives in Sanskrit, just as they are in, e.g., Latin and Ancient Greek. Taken together with the possible existence of DP-level free relatives in Hittite from Probert (2006) discussed below, this would suggest that we should not exclude categorically the possibility of DP-level free relatives in PIE.

A final pattern of 'plain' RC formation should be considered. Whereas in exx. (5.47–9) the head noun and the relative pronoun are adjacent, leading to the ambiguity between a correlative vs. and embedded structure, in ex. (5.50) the head noun DUMU.MEŠ-*uš* ('boys') is separated from the relative pronoun *kuiēš* by a different constituent from within the MC: in this case, the matrix verb *uwadanzi* ('they bring'). Such orderings, in which the head noun is separated from the RC, are quite common:

(5.56) αὐτὰρ ἐγὼ θεός εἰμι, [$_{CP}$ διαμπερὲς ἥ =σε$_{WL1}$
but I god.NOM.SG.F be.PRES.1SG to-the-end REL.NOM.SG.F =you.ACC.SG
φυλάσσω / ἐν πάντεσσι πόνοις]
guard.PRES.1SG in all.DAT.PL.M toil.DAT.PL.M
'But I am the goddess, who will guard you to the end in all troubles' (Hom. *Od.* 20.47–8)

(5.57) *ego servi sumpsi* ***Sosiae*** *mi imaginem*
I slave.GEN.SG.M assume.PERF.1SG Sosia.GEN.SG me.DAT.SG appearance.ACC.SG.F
qui *cum Amphitruone abiit hinc in exercitum*
REL.NOM.SG.M with Amphitruo.ABL.SG go-away.PERF.3SG hence into army.ACC.SG.M
'I have assumed the appearance of the slave Sosia, who's gone away to the army with Amphitruo'
(Plaut. *Amph.* 124–5)

(5.58) ***vagnúm*** *iyarti* ***yám*** *vidé*
call.ACC.SG.M raise.PRES.3SG REL.ACC.SG.M possess.PRES.3SG
'He raises the call which is his' (RV 9.14.6c)

As with peripheral 'plain' relatives, such RCs permit (at least) two analyses. If we take them to be embedded, we must consider them **extraposed**. This is a common strategy to avoid excessive centre-embedding,[170] even in languages where it is possible, such as English:

(5.59) A student$_i$ [$_{RC}$ who$_i$ I taught last year] emailed me yesterday (centre-embedded)

(5.60) A student$_i$ emailed me yesterday [$_{RC}$ who$_i$ I taught last year] (extraposed)

[170] Possibly associated with processing constraints, cf. Probert (2015: 399).

There are many and various accounts for the apparent rightward extraposition of embedded RCs (see de Vries, 2002: ch. 7). I do not commit to any particular analysis here. For now it is simply worth noting that surface discontinuity between the head noun and RC does not *per se* exclude an embedded analysis.

On the other hand, these structures would also follow immediately from the account I gave of **correlative** clauses with MC-RC order. Indeed, the only distinction between what would traditionally be called correlatives and right-extraposed RCs is the absence of a correlative demonstrative in MC. Consider the following processes I established for PIE correlative clauses:

1. [**EXREF**]: *REL must be co-indexed with either (a) a pronoun or (b) a full DP, in a neighbouring CP
2. **LEFTMOST HEAD NOUN RESTRICTION**: The head noun is only spelled out in the leftmost clause, whether it is the MC or the RC

These two features essentially predict that a correlative clause base-generated with MC-RC ordering will have the surface form of a 'plain' relative: the head noun will be retained in the MC only, satisfying the [EXREF] feature on *REL. A correlative analysis of this sort does not make any strong predictions as to the adjacency of HN_{MC} and the RC. For this reason, all types of peripheral 'plain' relative are accounted for, both where there is adjacency (ex. 5.47–9), and where there is not (ex. 5.56–8). As for the presence or absence of a correlative demonstrative, the model does not make any strong predictions. The [EXREF] feature on *REL is satisfied by HN_{MC}—the presence of an overt *DEM is not required. It is optional, in the usual way demonstratives are optional according to the pragmatic context of the utterance.

To collapse all instances of 'plain' relatives into the same syntactic mechanism as correlatives is essentially the approach of Hock (1989) and Kiparsky (1995), revisited in Hock (2015). Hock notes that the correlative demonstrative is 'optional', but does not expand on the nature of its optionality. My model for PIE is that it is effectively obligatory in RC-MC configurations where HN_{MC} is deleted (LHNR), but optional for correlatives in MC-RC configuration, where HN_{RC} is deleted by LHNR. Nevertheless, the central argument is that that PIE had 'conjoined' correlatives and any variations can be derived from that original strategy. So far, I have clarified the theoretical apparatus required to implement this model, which has much to recommend to it.

Nevertheless, the two rules of correlative formation I have postulated so far cannot account for embedded relatives. Aside from centre-embedding, the 'embeddedness' of certain postnominal relatives shows up in other ways. This was argued for in Hittite at length by Probert (2006).[171] Her core observation is that in Old Hittite RCs, if the MC does not contain an overt correlative pronoun (i.e., in 'plain' relatives), the MC regularly lacks a sentence connective (*nu* etc.). Recall that in §5.2.2 above, a crucial piece of evidence in favour of treating correlative RCs as adjoined to the MC at a CP-level was the possible presence of *nu* etc. at the beginning of both RCs and MCs. I argued (following Huggard, 2015) that

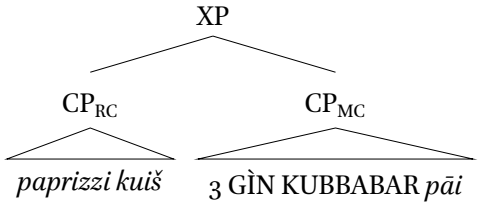
[171]The author also argues, more recently and more tentatively, that embedded relatives may be attested in Lydian (Probert, 2018).

nu occupied [Spec, ForceP], meaning both the RC and the MC contained a fully-fledged left periphery. Probert's (2006) argument is that this evidence is precisely lacking for MCs with 'plain' relatives in Old Hittite. Without a clear clause-boundary between RC and MC in 'plain' relatives, it is possible to analyse the 'plain' RC as embedded. Consider the following:

(5.61) *paprizzi* *kuiš* 3 GÌN KUBBABAR *pāi*
sully.PRES.3SG REL.NOM.SG.C 3 half-shekels-of-silver give.PRES.3SG
'The one who sullies gives three shekels of silver' (KBo 6.2 I 57)

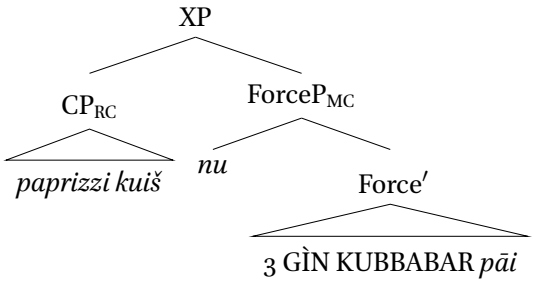
The correlative analysis for such a clause would be:

(5.62)



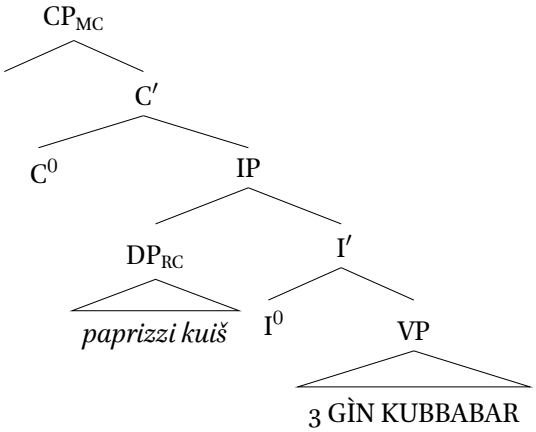
Within this grammatical context, it should be possible to insert a *nu* (vel sim.), generating the structure:

(5.63)



On the other hand, if the RC is embedded (i.e., a DP-level free relative), the sentence structure would be something along the lines of the following:

(5.64)



This would preclude the possibility of inserting *nu* between the RC and the MC, as MC left periphery occurs **before** the RC. In this instance, according to Probert's analysis, the free relative occurs first simply because the entire free relative acts as a subject DP for the matrix verb: I have modelled this position as [Spec, IP]. Once again, we find ourselves with a structure that is ambiguous at a surface level, which requires some further analysis.

As far as my model goes, in order to treat this as an example of a correlative, we would have to amend the [EXREF] feature slightly, and say that **[EXREF] must be fulfilled by either (a) a DP in the form of a correlative pronoun and/or HN_{MC}; or (b) a null *pro***. I suggested fleetingly in §5.2.3 that the latter might be possibility for correlatives in which the correlative pronoun was the subject of the MC and no clitic was available. Since Hittite, like all ancient IE languages, is pervasively *pro*-drop, it is not within the bounds of reason to posit a null *pro* in the subject position of the MC, which is able to satisfy [EXREF]. This would account for 7/8 of the relevant examples listed in Probert (2006: 31–3).[172] This hypothesis may find support in Garrett's (1990) observation that, although Hittite does have subject clitics, they are never attested with transitive verbs—a condition that holds for 6/7 of the sentences my model works for. However, the comparative data from Sanskrit, which lacks subject clitics altogether, make no such prediction. Rather, the rates of overt resumption in correlatives do not appear to correspond at all to the case function of the antecedent in the MC (Hettrich, 1988: 573–8). It seems, therefore, that if any null element can satisfy [EXREF] in Sanskrit, subject *pro* is not necessarily a preeminent candidate. Returning to Old Hittite, even if resumption by *pro* is a possibility where subject clitics are unavailable, Probert (2006) notes that overt resumption should also be possible with the accented demonstrative *apāš*; that this possibility is not exploited further supports her argument that these are genuinely embedded RCs. Finally, the lack of CP-elements (e.g., *nu*) occurring between RC and MC would still remain to be accounted for under a correlative analysis. As such, I accept the possibility that these are genuine instances of embedded relatives, though they remain ambiguous.

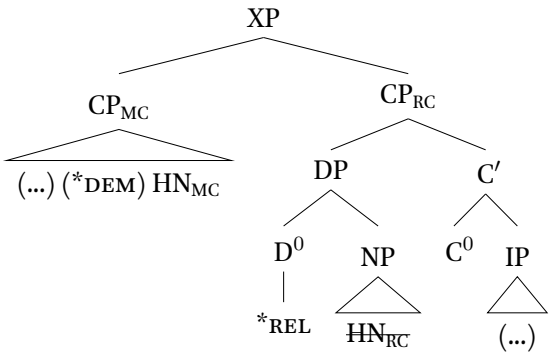
To summarise so far: many clauses that have the surface form of 'plain' relatives are ambiguous between a correlative analysis with MC-RC word order vs. genuine embedded relatives. Those which are not ambiguous are: (a) centre-embedded relatives, (b) free relatives in co-ordination structures, which are unambiguously DPs. A little bit more subjectively, we may also consider the evidence from Old Hittite postposed RCs as support for an embedded structure. What does this mean for PIE?

A recurrent theme in previous work on this topic, including Haudry (1973), Hock (1989, 2015), Kiparsky (1995), Bianchi (2000) and Probert (2006, 2015) is the possibility that ambiguity between different forms of RC could result in the reanalysis of one type in favour of the other. With the exception of Probert (2006, 2015), the argument is generally made that embedded RCs are derived from correlative ones. Historically, this falls into the broader notion that 'paratactic' structures develop into 'hypotactic' ones. Although my argument may fit into this broader theoretical picture, I eschew the categories 'paratactic' and 'hypotactic', and instead focus on the specific reanalyses within a Minimalist framework that may account for the different observed types of RC.

[172] With the alternative reading of KBo 6.2 i 7–8 (Probert, 2006: 30, n.13).

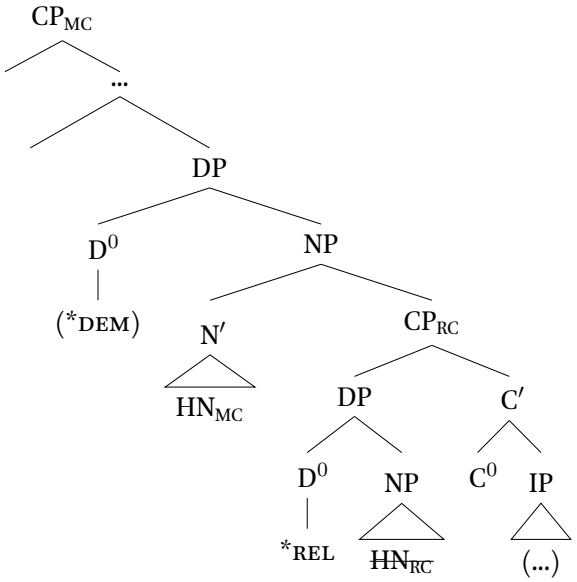
I start with the fairly uncontroversial hypothesis that PIE had a correlative clause of the type discussed in §5.3. As I have noted several times, correlative clauses of this type could generate surface 'plain' relatives, as long as the RC was peripheral to the MC. I repeat the structure from ex. (5.50):

(5.65) **Matching analysis for correlatives:**



I believe structures of this sort provide precisely the right type of ambiguity to allow for a reanalysis in which the RC is no longer adjoined to the matrix CP, but the DP containing HN$_{MC}$. With this single reanalysis, and no further changes, a correlative clause has become an embedded one under the MA:

(5.66) **Matching analysis for 'plain' relatives:**



A surface ambiguity of this sort has already been observed for Sanskrit (Hock, 1989, 2015; Davison, 2009) and Ancient Greek (Probert, 2015), though their accounts of the possible reanalyses differ. Probert (2015) does not commit to a theoretical account involving deletion, and treats the reanalysis as from a free relative in apposition to HN$_{MC}$ to a headed postnominal (embedded) RC.[173] For a free relative to be in apposition with a DP, however, the free relative itself must be a DP. Unless embedded postnominal RCs already existed, therefore, this would mean there was a stage where free relatives were DPs but (headed) correlative RCs were still CPs—a situation which is not impossible, but would require robust

[173]This hypothesis expands on earlier suggestion of this development within Greek by Ruijgh (1971).

motivation. In fact, Probert does not support this hypothesis; rather, she tentatively suggests that (restrictive) postnominal RCs (i.e., DP-level RCs) **did** exist in PIE (2015: 444), in which case the reanalysis she suggests does not present a primal origin for embedded relatives, but a sort of 'sideways' reanalysis from one type of DP-level RC to another. Hock (2015), on the other hand, suggests that it is precisely the optional deletion of correlative pronouns that leads to 'conjoined' correlatives being reanalysed as embedded; he suggests that, even synchronically, at an early stage certain ancient IE 'plain' relatives could instead be treated as underlyingly correlative, just with the necessary (and largely unexplained) deletions. My argument is that synchronically, what look like embedded RCs are probably embedded RCs; diachronically, however, it is at least possible to derive them from an original correlative strategy in PIE, with the features I have established in §5.3.

It is worth noting briefly that the reanalysis I have suggested is supported by two general psycholinguistic principles: **minimal attachment** and **late closure** (see, e.g., Frazier, 2013). Under the former, while parsing a sentence the language-comprehender will posit as few tree-structure nodes as possible; under the latter, they will attach an ambiguous phrase (e.g., the relative CP here) to the most recent phrase (e.g., the DP containing HN_{MC}), all else being equal.[174] In other words, the embedded structure (as long as it is clause-final), is easier to process than the CP-adjoined correlative analysis; this lends further credence to this particular ambiguity as a locus of reanalysis.

A complementary, albeit slightly less convincing pathway of change could be posited if we also admit the HRA into our reconstruction. In MC-RC correlatives, the occurence of HN_{MC} in clause-final position, allowing for the reanalysis, was incidental. A similarly incidental ordering could be derived in correlatives of RC-MC orderings where HN_{RC} was fronted, within the RC, to a position preceding the relative pronoun, i.e.:

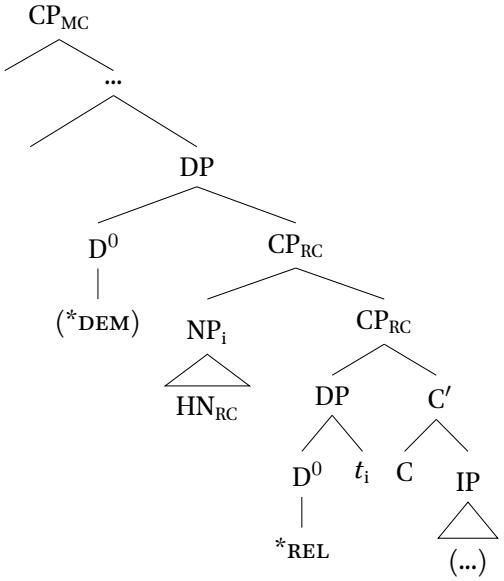
(5.67) $[_{RC}\ HN_{RC}$ *REL...$]\ [_{MC}$ *DEM... $]$

In the first place, this order would be generated by the movement of HN_{RC} to the **higher** [Spec, TopP], with movement of *REL to its customary location in the **lower** [Spec, TopP]. In this instance, the movement of HN_{RC} was not initially grammaticalised, since it was also possible for it to stay in its base-generated position within RC. Yet it is possible that the fronting of HN_{RC} in this way was reanalysed as a grammatical requirement, rather than simply a pragmatic one. In other words, rather than *REL moving upward to check the [*u*REL] feature on Top° alone, and HN_{RC} moving to a different specifier position to check a different functional feature, **both movements are driven by [*u*REL]**. This is essentially how the HRA works as envisionsed by Bianchi (1999, 2000): the entire relativised DP is fronted, with the head noun subsequently fronting to a higher specifier position (for details see §2.3). The same ambiguous structure discussed in ex. (5.65–6) could then be reanalysed into an HRA structure, where HN_{MC} is reanalysed as HN_{RC}, fronted to the very left edge of the RC:[175]

[174] Minimal attachment would further suggest that an interpretation in which the RC is a complement, rather than an adjunct, would be preferred; I do not pursue this question here.

[175] For simplicity, I have not labelled the CP specifiers involved in this process.

(5.68) **Head-raising analysis for 'plain' relatives:**



A reanalysis of the syntactic relationship between HN$_{RC}$ and *REL in linear orders of this sort is reminiscent of Haudry (1973: 155–7).[176] I say this reanalysis is less convincing because a central feature of MC-RC correlatives is that HN$_{RC}$ is not spelled out (LHNR). As far as correlative clauses are concerned, the only grammatical scenario in which the order [HN$_{RC}$ *REL] occurred was in *preposed* correlative RCs (i.e., RC-MC configurations), which cannot be straightforwardly reanalysed as postnominal, embedded RCs—on the contrary, they appear to provide a robust enough triggering environment for correlative clauses to be maintained as a relativisation strategy even when competing with a 'simpler' embedded analysis. They could, however, be readily reanalysed as extraposed DPs, in which case they would be accounted for under an analysis of the sort proposed by Cinque (2020): the correlative pronoun therefore no longer exists to satisfy [EXREF] on *REL, but simply as a form of resumption resulting from dislocation, which is optional. On the other hand, if the MA provides the pivot for change, it is then possible in a secondary process for an MA relative to be reanalysed as an HRA relative; once embedding is a possibility, the question becomes the mechanism by which the head noun reaches its position directly preceding the relative pronoun, which could be either by base-generation (MA) or movement (HRA). As I noted in §2.2, without the kind of locality diagnostics commonly used to disambiguate between the two structures, both options are possible synchronically. It may well be a conspiracy of the two that leads to the attested embedded forms, with both preposed and postposed RCs being reanalysed as embedded (i.e., DPs).

It is possible that this reanalysis was nascent in PIE at the later stages of unity. I would argue that while there can be no doubt that PIE had correlative relatives, given the volume of evidence, there is more doubt as to whether it had embedded RCs. The fact of the matter is that 'plain' relatives are so often peripheral as to offer largely ambiguous evidence in favour of an embedded analysis. Even when there is apparent external evidence, such as the absence of sentence connectors for Old Hittite 'plain' relatives,

[176]The similarities between our analyses are quite limited, however; Haudry (1973) implies that the innovation of appositive RCs leads to a scenario in which the relationship between the relative pronoun (in his model, originally indefinite in function) and the correlative pronoun 'becomes secondary and can even cease to exist' (1973: 155, trans. KRP). This clearly differs quite substantially from my hypothesis.

this does not **preclude** a correlative analysis: as Hock (2015) notes, the Hittite sentence connectors are simply less prevalent in Old Hittite than they are in later stages of the language, and their absence in 'plain' relatives could simply be a trivial archaism. I will pick up this question in the following section, where I summarise my arguments for the different syntactic types of RCs in PIE.

5.5 "Anaphoric" relatives: The crossroads

I have left open the possibility that PIE had embedded (i.e., DP-level) RCs; I do not believe there is sufficient evidence to exclude them from the latest stage of linguistic unity, before the Anatolian split. Nevertheless, across the ancient IE languages, embedded relatives appear to be marginal, and plausibly developed from the non-embedded strategy. In fact, with the few exceptions mentioned in the preceding section, we can collapse almost every type of RC in PIE into one, non-embedded category: I will call this **anaphoric relative clause**. I have established three processes that drive the derivation of the anaphoric RC:

1. Top° bearing a(n optional) strong [*u*REL] feature
2. *REL bearing an [EXREF] feature
3. LEFTMOST HEAD NOUN RESTRICTION: HN_{RC} and HN_{MC} match, and only the leftmost instance is spelled out

These conspire to yield both correlative clauses and 'plain' relatives: the former is the case when the [EXREF] feature on *REL was satisfied by a demonstrative, while the latter is when it is satisfied by a full DP. The various possibilities are listed below:

(5.69) Correlative, RC-MC order: *REL HN_{RC}... *DEM...

(5.70) Correlative, MC-RC order: *DEM HN_{MC}... *REL

(5.71) 'Plain' relative: *HN_{MC}(...) *REL...

My use of the term 'anaphoric' does not imply that PIE somehow lacked 'true' RCs, and that relativisation is somehow a secondary development of simple anaphora. On the contrary, I have argued that PIE had a specialised pronoun *REL, which had a specialised feature [EXREF], that sets it apart from simple 'anaphoric' pronouns (**só-/tó-* etc.). The RC is anaphoric in the sense that there is always a relationship of co-indexation between *REL and a constituent in the MC, whether that be a correlative pronoun or HN_{MC}. This behaviour of *REL therefore unites many 'plain' relatives together with correlatives, to the exclusion of DP-level RCs (e.g., free relatives, embedded RCs).

My definition of an anaphoric RC bears much similarity with what Lehmann (1984) calls the *umstellbar* ('adaptable') RC. In the author's typology, this unifies both preposed correlatives and postposed relatives (either 'plain' or correlative). In both cases, the RC sits aside the MC but is not embedded within it. Lehmann (1984: 129–32) focuses primarily on Vedic as an archaic IE languages that shows the 'adaptable' RC; he treats Hittite as instantiating strictly the preposed type of RC, to the exclusion of its

postposed RCs (Hoffner & Melchert, 2008: 425–6); if we include postnominal RCs in Hittite, it may well be contained within this category too, together with the other ancient IE languages which show correlative clauses alongside embedding strategies.

The generativist literature on RCs tends not to recognise Lehmann's terminology in this connection. This is partly because he takes only a vague stance on theoretical issues such as constituency, such that there is no clear Minimalist formalism of the 'adaptable' clause as he presents it. Lehmann's *umstellbar* RC is mentioned by de Vries (2002: 235), who promptly dismisses it. The main problem for such a category, de Vries argues, is that preposed correlative RCs and right-extraposed RCs (i.e., relatives with MC-RC word order) are categorically different strategies.[177] His primary criticism is derived from Srivastav (1991), listing two concerns:

1. Preposed correlative RCs contain the head noun, while postposed relatives do not
2. Correlative clauses are maximalising, while postposed relatives are usually restrictive

On this basis, he concludes that (2002: 235, emphasis mine):

> ...extraposed relatives cannot simply be analysed as correlatives that are right-adjoined or moved to the right. Rather, they behave on a par with postnominal relatives. And in fact, Hindi and related languages have a secondary postnominal strategy. **This reasoning is valid for Sanskrit**, Avestic, Hindi, Marathi, and probably for related languages like Bengali and Gujarathi.

I am not sure what evidence de Vries has that this reasoning applies to Sanskrit: Srivastav (1991: 654, n.18, emphasis mine) herself explicitly notes that 'Hock['s (1989) claim of symmetrical adjunction for Sanskrit RCs] cannot account for the...asymmetry [of correlative adjunction], **at least for Hindi**.' Furthermore, in Sanskrit preposed correlative RCs do not have such rigid semantic constraints: they can be restrictive, non-restrictive, or maximalising. I have also noted, as argued by Davison (2009), that Hindi IP-adjoined correlatives are categorically different from CP-adjoined relatives as attested in Sanskrit and Hittite, which I have reconstructed for PIE. Moreover, de Vries (2002: 235) himself notes that under his analysis, 'the issue is why these languages [viz. PIE in my reconstruction] have a hidden postnominal relative strategy whereby extraposition is obligatory.' My argument is that we do not need to posit a 'hidden postnominal' strategy in PIE—what appear to be extraposed embedded relatives can perfectly well be analysed simply as anaphoric relatives appearing in MC-RC order. The fact that HN_{RC} is deleted has no bearing on the question of embeddedness, since the mechanism of deletion (matching) occurs in both CP-level and DP-level RCs (cf. exx. 5.65–6).

Anaphoric RCs represent the 'crossroads' between embedded, postnominal relatives and non-embedded, correlative clauses, because they are capable of producing both types of structure on the surface. Reanalyses can occur in various directions: on the one hand, if RC-MC ordering is fixed, the result is a change toward a strict correlative strategy, with residual postnominal forms of the sort attested in Hittite. On the other hand, MC-RC orderings lend themselves quite clearly to a reanalysis in favour of

[177] Slightly confusingly from the persecptive of the terminology I have employed so far, de Vries (2002) considers only correlatives of the order RC-MC (i.e., where the RC is preposed) to be *correlatives*. The order MC-RC, whether or not a correlative demonstrative is present in MC, is considered postnominal.

embedded RCs, as attested in Latin and Ancient Greek. Moreover, the actualisation of a certain reanalysis does not necessarily lead to the demise of the anaphoric relativisation strategy. Perhaps the clearest example in this regard is Latin, which makes extensive use of embedded RCs as an unmarked strategy but elaborates upon the anaphoric quality of relative pronouns in the form of 'connecting' relatives.[178] In short, **every form of RC attested in the ancient IE languages *could* be derived, either synchronically, or diachronically (via reanalysis) from the anaphoric RC**. I will re-iterate that this does not constitute proof that the anaphoric RC was the only RC type in PIE—embedded relatives, as attested by free relatives and centre-embedded relatives, may well have been possible. Indeed, since postposed anaphoric RCs necessarily led to ambiguous structures, we could even argue that embedded, DP-level, postnominal RCs existed in PIE. However, I think it is reasonable to suggest that the anaphoric RC was the dominant syntactic type of RC in PIE, and that embedded RCs of any type were marginal and probably innovative.

5.6 Questions of semantics

With the syntactic formation of anaphoric RCs established, I turn finally to the interaction between the syntactic structures I have posited so far and the semantic types of RC in PIE.

The most straightforward conclusions can be made concerning restrictive (and maximalising) RCs. In short, they can be of any of the syntactic types I have established so far: anaphoric (preposed), anaphoric (postposed) or 'embedded'. The first of these is most common in those languages which favour correlative structures (e.g., Hittite and Sanskrit), but even in these languages there is no strict constraint that restrictive or maximalising relatives must be preposed. I reconstruct this level of flexibility for PIE, with the caveat that the 'embedded' RC being a late-stage innovation may not have been available for restrictive relatives at an earlier stage in PIE.

Of more interest, then, is the status of appositive RCs (ARCs) in PIE. On the basis of their attestation across the board in IE languages (albeit rarely in the earliest stages of certain branches—see §3.3.2), it seems reasonable to suggest that PIE had some form of ARC, as has been suggested by several authors (Sturtevant, 1930; Lehmann, 1984; Hettrich, 1988; Hajnal, 1997). I will argue here that ARCs instantiated a different set of syntactic behaviours in PIE than restrictive and maximalising RCs. These behaviours ultimately arise from some more general principles of apposition and ARCs cross-linguistically.

I will begin by noting that it is very likely that in PIE, ARCs were obligatorily postposed (i.e., in MC-RC configuration). This is also not a novel claim: however, my reconstruction differs from the theories such as those of Sturtevant (1930) and Hajnal (1997) in that I believe restrictive and maximalising RCs could **also** be postposed, i.e. postponement is a necessary but not a sufficient condition for an RC to be interpreted as appositive.[179] This hypothesis holds for Hittite, Ancient Greek and Latin, where preposed correlatives are consistently either restrictive or maximalising. Support for the claim may also be drawn

[178] See Pompei (2011a: 505–8) and references therein.

[179] I re-iterate my rejection of the notion that this distinction was reflected morphologically: all semantic types of RC in PIE were introduced by the same *REL (§3.2.1).

from Sanskrit, which shows a strong tendency for appositives to be postposed (Hock, 1989). What remains to be accounted for is the apparent propensity of Sanskrit at all stages, in spite of this tendency, to have preposed ARCs too (Hettrich, 1988: 696–794). I will return to this question at the end of the section.

The evidence from Sanskrit notwithstanding, it seems to be true that no language permits a preposed ARC introduced by a relative pronoun (Lehmann, 1984; de Vries, 2002, 2006; del Gobbo, 2007, 2017). While in my methodology I aim to avoid relying too heavily on typological generalisations where the comparative data appear to flout them (see §2.2), in this case I do not believe the evidence from Sanskrit is sufficient to dismiss such an overwhelmingly strong pattern out-of-hand. Furthermore, the absence of preposed ARCs is not a typological accident: rather, it follows from the underlying structures of apposition. A full discussion appositive and parenthetical structures in PIE is beyond the scope of this thesis, but it is worth elucidating the specific ways in which they may inform our reconstruction of PIE RCs, especially concerning the features of *REL.

It has been argued by several authors (Demirdache, 1991; de Vries, 2002, 2006; del Gobbo, 2007, 2017) that relative pronouns in appositive RCs are **E-type** anaphoric pronouns.[180] Chierchia (1995: 15) describes such pronouns as a 'proxy for a description whose content can be systematically reconstructed from the context'. Essentially, E-type pronouns 'stand in' for a full NP/DP which has already been introduced to the discourse, as in the following example (Chierchia, 1995: 16):

(5.72) A man$_i$ walked in. He$_i$ was wearing a hat.

On the whole, E-type pronouns must be linearly preceded by their antecedents to be felicitous:

(5.73) *He$_i$ walked in. A man$_i$ was wearing a hat.

There are exceptions to this however, such as conditional clauses:

(5.74) If he$_i$ wears a hat, a man$_i$ will not feel the cold.

I will return to these exceptions below, as they may account for the apparently deviant position of ARCs in Vedic. Proceeding with the analysis of authors such as de Vries (2002) and del Gobbo (2007), treating the relative pronoun in ARCs as E-type pronouns accounts straightforwardly for the observation that ARCs must follow their antecedents. In fact, del Gobbo (2007: 183) also notes that this constraint is in a sense 'pragmatic', noting that 'in order for the relative pronoun [in an ARC] to be properly interpreted, this has to be processed **after** its antecedent, i.e., **after** the 'head' of the RC' (emphasis mine).

If we accept that the position of ARCs with regard to the MC is universally motivated by the syntactic features intrinsic to E-type pronouns, it follows that PIE ARCs were consistently postposed. This means that for an anaphoric RC to be non-restrictive, it must occur in MC-RC configuration. This ordering alone will yield the following two features:

1. ARCs do not require a correlative demonstrative in the main clause, if the [EXREF] feature of *REL is already satisfied by a DP

[180] The terminology is apparently traced back to Evans (1980, 1982).

2. The internal head noun (HN_{RC}) of an ARC will not be spelled out, as it matches with HN_{MC} and is deleted by LHNR

These constraints would yield a superficial identity with restrictive RCs that only happen to be placed in MC-RC word order. Yet for ARCs, we may expect the second feature (lack of HN_{RC}) to follow not only from 'matching' as attested in restrictive RCs, but from the more general principles of E-type anaphora. At LF the E-type pronoun may be either an identical copy of its antecedent or a definite description of that antecedent (del Gobbo, 2007: 183–4); at PF the NP (if present) is deleted, leaving only the determiner (replaced with an anaphoric pronoun) to be spelled out. Further investigations may seek to either disambiguate these two types of deletion (matching deletion vs. E-type deletion), but for our purposes they yield the same results. It is also worth noting that there are varying theories for the underlying syntactic structure of ARCs (see, e.g., de Vries, 2006; del Gobbo, 2017), which may interact with the theory I have proposed in different ways; at this stage, however, I do not believe a different model would make significantly different predictions for the behaviour of ARCs in PIE.

I return finally to the question of preposed ARCs in Vedic Sanskrit. Hettrich (1988: 696, n.282) claims there are 244 such examples in the Rig Veda. Moreover, in direct contravention to the LHNR proposed for PIE above, there are numerous examples of preposed ARCs where the head noun is not found in the RC (HN_{RC}), but rather in the MC (HN_{MC}), such as the following example from Hettrich (1988: 697).

(5.75) *yásmin* *víśvāḥ* *ádhi* *śríyaḥ* *ráṇanti* *saptá*
REL.LOC.SG.M all.NOM.PL.F over.PW glory.NOM.PL.F delight.PRES.3PL seven
saṃsádaḥ *índram* *suté* *havāmahe*
gathering.NOM.PL.F Indra.ACC.SG.M pressing.LOC.SG.M invoke.PRES.1PL
'[The one] over whom all glories delight, and the seven gatherings—Indra, we summon at the pressing [of the Soma]' (RV 8.92.20)

For examples such as these, we may follow Probert (2015: 35–6, 133) and tentatively suggest a reading in which these are in fact free relatives (which are inherently maximalising), which stand in apposition to HN_{MC}. By way of comparison, consider the following from Homeric Greek (Probert, 2015: 133).

(5.76) Ζεῦ ἄνα, δὸς τείσασθαι, [$_{FR}$ ὅ με
Zeus.VOC.SG king.VOC.SG give.IMPV.2SG revenge.AOR.INF REL.NOM.SG.M me.ACC.SG
πρότερος κάκ' ἔοργεν], / δῖον Ἀλέξανδρον, καὶ ἐμῇς
first badly do.PERF.3SG god-like.ACC.SG.M Alexander.ACC.SG.M and my.DAT.PL.F
ὑπὸ χερσὶ δάμασσον...
under hand.DAT.PL.F subdue.AOR.IMPV.2SG
'Lord Zeus, grant me to revenge myself on him who has done me harm first—god-like Alexander—and subdue him at my hands...' (Hom. *Il.* 3.351–2)

While I concur with Probert's analysis of such clauses in Greek, I am reticent to apply this analysis across the board in Vedic. My primary concern is that treating these as free relatives would require them to be analysed as DPs, rather than CPs. This is not impossible; I have argued that Sanskrit does have

DP-style relatives, as attested in the co-ordination structures discussed in §5.4 above. However, I have argued that this analysis is fairly marginal in Sanskrit, as evidenced by the language's more general lack of 'embedded' structures (Hock, 1989, 2015). If we analyse preposed ARCs as free relatives, then we have to hold that they are consistently more felicitous when in apposition with another DP than they are on their own. On the other hand, if we take preposed ARCs not as DP-style free relatives, but as the (headless) RC part of a correlative pair, the requirement for "apposition" follows from the [EXREF] feature on the relative pronoun. The structural ambiguity here may lead to the reanalysis of a preposed ARCs as a free relative in apposition with HN_{MC}, but we would expect the result of this reanalysis to be a preponderance of free relatives without correlative demonstrative, which does not appear to materialise in Sanskrit.

Even if we were to accept a free relative account for examples such as (5.75), we would still have to account for preposed ARCs which do contain HN_{RC} (Hettrich, 1988: 696):

(5.77) *yáṃ* *vái* *sū́ryam* *svárbhānuḥ* *támasā*
REL.ACC.SG.M EMPH SUNACC.SG.M Svarbhānu.NOM.SG.M darkness.INS.SG.N
ávidhyat *āsuráḥ* *átrayas* *tám* *ánu* *avindan*
strike.IMPF.3SG āsura.NOM.SG.M Atri.NOM.PL.M DEM.ACC.SG.M again.PW find.IMPF.3PL
'The sun which Svarbhānu Āsura struck with darkness, that one the Atris found' (RV 5.40.9)

From a syntactic perspective, these tokens are simply identical to restrictive/maximalising correlative RCs (Probert, 2015: 37). To me, this suggests that they have a shared origin in the (anaphoric) correlative clause I have posited for PIE. Given what I have argued so far, we must therefore consider the possibility that the availability of preposed ARCs did not result from a syntactic reanalysis, but rather a semantic one. In short: we may argue that preposed RCs were originally restrictive/maximalising. Consider ex. (5.77): in this poem, as is common elsewhere, the sun is personified as *Sūrya*. Adopting this translational equivalent, one could force a restrictive reading of a definite entity in the English translation:

(5.78) The Sūrya that Svarbhānu Āsura struck with darkness, that one the Atris found.

Here, I have deliberately translated the RC as restrictive; in English, we see the peculiar insertion of the definite article with a proper noun in such instances, which would be ungrammatical with an ARC:[181]

(5.79) This is the Paris that I love.

(5.80) *This is the Paris, which I love.

Clearly, in the absence of a definite article (or indeed, intonational evidence, or the option to use an uninflected complementiser), the two interpretations have identical surface forms in Sanskrit. This could mean that a clause of this type is genuinely restrictive, used for rhetorical effect as part of a more general pattern in Vedic whereby 'a unique entity...although unique, is nevertheless identified with the one to which certain properties belong' (Probert, 2015: 37). This is not to say that we should translate all

[181] Example from Alexopoulou (2007).

preposed ARCs as semantically restrictive, but rather that there is an extra-syntactic pathway from one analysis to the other.

I offer a final speculation on the availability of preposed ARCs in Vedic. If we treat the relative pronoun in ARCs as an E-type pronoun, del Gobbo's (2007) prediction is that it must follow its antecedent in linear order. This follows from the more general rule that E-type pronouns must follow their antecedents when they occur in separate conjoined clauses (cf. exx. 5.72–3 above). However, as noted above, this constraint does not seem to apply to some complex sentences: specifically, an E-type pronoun in the protasis of a conditional clause can precede its antecedent in the apodosis (cf. ex. 5.74 above). Given the relative-conditional connection, it could be that the apparent lack of E-type ordering constraints in Vedic preposed ARCs offers another insight into the two clause-types; paradoxically, however, it is precisely MaxRCs that pattern with *wh*-conditional correlatives, and not ARCs. I leave a detailed investigation of the possible reanalyses in this connection, both semantic and syntactic, to further studies.

5.7 Summary

To summarise, I reconstruct the RC for PIE as follows:

1. **PIE had an inflected relative pronoun, *REL:**
   (a) *REL was a full lexical word
   (b) It bore the feature [*i*REL]
   (c) It moved to the **lower** [Spec, TopP] if the [*u*REL] feature on Top° was strong (which it probably was in pragmatically unmarked situations)
   (d) It bore the feature [EXREF], which required it to be co-indexed with either (a) a correlative demonstrative, or (b) a full DP in the MC
2. **PIE certainly had an anaphoric (correlative) relative clause:**
   (a) the RC was an independent CP and not a DP
   (b) Because the RC was a CP, it could not occupy an argument position with the MC, but was rather adjoined to the matrix CP, probably under an external projection that dominated both the RC and the MC
   (c) The orderings RC-MC and MC-RC were both possible
   (d) There were two possible instances of the head noun: one in the RC (HN_{RC}) and one in the MC (HN_{MC})
   (e) HN_{RC} and HN_{MC} were **matched**; the rightmost instance was deleted under unmarked pragmatic conditions, meaning the leftmost one was pronounced (LEFTMOST HEAD NOUN RESTRICTION, LHNR)

(f) LHNR operated indiscriminately of the ordering of RC-MC vs. MC-RC, meaning that under normal pragmatic conditions, postposed RCs did not contain an overt HN_{RC}

(g) In RC-MC configurations, although HN_{MC} was deleted by LHNR, there was still a requirement for correlative pronoun (or possibly *pro*) owing to the [EXREF] feature on *REL

(h) The correlative demonstrative could be either marked for a specific discourse function, or unmarked. If the former, it would take the form of an accented demonstrative possibly fronted to the left periphery; if the latter, it could be an enclitic pronoun targeting the lowest active left peripheral head above Fin° and subject to prosodic inversion

(i) Conversely, LHNR resulted in the relative infrequency of correlative demonstratives in MC-RC configurations: the [EXREF] feature on *REL was satisfied by HN_{MC}

3. **PIE may have had an embedded relative clause:**

(a) This was a CP either adjoined to a DP, or a complement of DP

(b) This could have developed from a reanalysis of postposed anaphoric RCs as adjoining to, or being a complement of, a clause final HN_{MC} DP, rather than adjoining to the matrix CP

(c) This reanalysis was only nascent in PIE, reflected in the unstable status of this construction in the earliest stages of Hittite, Greek and Sanskrit

(d) The DP containing the RC **could** have been embedded into the MC, occupying an argument position

4. **PIE had restrictive, maximalising and appositive RCs:**

(a) Restrictive and maximalising RCs could be of any syntactic type

(b) Appositive RCs were obligatorily **postposed**; *REL in ARCs acted as an E-type pronoun which had to be preceded by its antecedent

I locate these hypotheses within the broader picture painted by previous authors on this work in §6, where I also discuss their implications for a more general theory of PIE syntax.

Chapter 6

Conclusions and Outlook

In this final section, I summarise what I have argued in each of the previous chapters, and consider some possible implications of this work for PIE, syntactic reconstruction, and Comparative Philology more generally.

6.1 Conclusions

In the opening chapters of this thesis (§1–2), I set the scene for the reconstructive arguments put forward in the subsequent chapters (§3–5). My approach was led by a belief of the sort expressed in Clackson (2017: 204) that '[p]erhaps it is time to leave the arguments about methodology to one side, and concentrate on reconstructing syntax.' Even guided by this sentiment, the methodology I have adopted (Minimalist reconstruction) continues to require discussion and refinement, some of which I offered in §2. More specifically within Minimalism, relative clauses are themselves a rich source of discussion and debate. For this reason, I dedicated a substantial part of §2 to laying out some of the complex theoretical questions that persist within synchronic analyses of relative clauses. I demonstrated the applicability of the various hypotheses to some of the patterns attested in ancient IE languages, even though—with some notable exceptions—they are rarely the basis of such theories. This chapter allowed me to establish both the general applicability of Minimalism to the question of relative clause structure in ancient IE languages, and the possibility of using these languages as an informative source for the development of such theories.

I dedicated §3 to a detailed survey of the descendants of two PIE vocabulary items— **yó-* and **kʷí-/*kʷó-*—which seem to be the ancestors to the forms of the relative pronoun attested in the ancient IE languages. I systematically eliminated the possibility that **yó-* bore any function other than that of a relative pronoun in PIE, at least by the latest stage of genetic unity. This represents quite a different picture from the one that emerges of PIE **kʷí-/*kʷó-*, which clearly did exhibit the functions of both an interrogative and an indefinite pronoun in PIE. However, I was unable to exclude the possibility that **kʷí-/*kʷó-* was also a relative pronoun in PIE, which may either complement, supplant, or be supplanted by, **yó-*. I argued further that there was no reconstructable difference in the syntactic behaviour of **yó-* vs. **kʷí-/*kʷó-* when used as a relative pronoun. For this reason, I concluded that PIE had a relative pronoun, which I label as *REL, whose phonetic content may have surfaced as either **yó-* or **kʷí-/kʷó-*. In other

words, I argued that PIE had a unitary **functional category** in Minimalist terms (*REL), whose syntax could be reconstructed by a comparison of the behaviour of descendants of both **yó-* and **kʷí-/*kʷó-*. I held that whatever may have led to the formal discrepancy in the daughter languages, this was not reflected in the function of the relative pronoun. In the course of making this argument, I also argued against the hypothesis that **yó-* and **kʷí-/*kʷó-* may have co-existed as relative pronouns that introduced distinct semantic and/or syntactic types of relative clause.

§4 is responsible for the 'related phenomena' disclaimer in the title of this thesis. I addressed two topics—the left periphery and "Wackernagel's law"—either of which could be (and indeed, have been) the topic of their own PhD theses. I argued that the two were implicitly linked: that a theory of the PIE left periphery had to account for "second position" clitics, and *vice versa*. Their inclusion in this thesis was due to their central importance in the formation of relative clauses. In particular, the 'landing site' of a relative pronoun fronted to the left periphery (*wh*-movement) is a question that explicitly draws together these two research areas within the broader question of relative clause structure. I took Vedic as a case study, to which I compared Ancient Greek, Latin and Hittite, as representative of the early left-peripheral structures in their respective language families. On this basis, I argued that PIE had two distinct classes of "second position" clitics: pronominal (WL1) and clausal (WL2). While the position of the latter can be explained almost entirely as a PF phenomenon, I argued that the former, while also subject to phonological constraints, were raised in the narrow syntax. The rule posited for their distribution (target the lowest active left-peripheral head above Fin°) facilitated the disentanglement of various left-peripheral projections, and accounted for the differing patterns of the descendant languages. I finished the section with a reconstruction of the left periphery and "second position" clitics in PIE, with an account of the reanalyses that led to the distinct outcomes in different branches.

The PIE situation arising from §3 and §4, informed by the theory of relative clause syntax established in §2, bring us to the heart of the thesis: the reconstruction of PIE relative clauses in §5. The reconstruction consists in 23 individual conclusions of various levels of specificity; these are listed at the end of §5. My overarching argument is that PIE had what I have dubbed an **"anaphoric" relative clause**, which can account for various structures including correlative clauses and post-nominal 'plain' relatives, as well as providing a *possible* genesis for embedded relative clauses. Central this reconstruction is *REL, which I argue to have borne a feature unique to the relative pronoun: [EXTERNAL REFERENCE]. This feature is responsible for the obligatory co-reference of a relative pronoun with a either a nominal antecedent or a resumptive pronoun within a neighbouring CP. In the course of this reconstruction, I have also addressed the varying semantic types of relative clauses in PIE. In particular, I argued that in addition to restrictive and maximalising relative clauses, PIE had an appositive (non-restrictive) relative clause. I argued that while restrictive/maximalising relative clauses were syntactically 'flexible', appositive relative clauses were obligatorily postposed.

6.2 Outlook

I locate the implications of my conclusions for future research within three general categories: those that apply to PIE; those that apply to syntactic reconstruction; and those that apply at an interdisciplinary level.

First, the implications for PIE. While some areas of my reconstruction are quite specific, perhaps even to a fault, several questions remain. I start, as it were, at the beginning: what are we to make of the status of **yó-* and **k^{w}í-/*k^{w}ó-* in PIE? One may accept my argument that *REL was a unitary functional category with varying phonetic content—I have argued this is an apt approach to relative clause structure—but the fact of the matter is that we clearly do have two distinct lexical items here, whose 'original' values I have not treated. Although I chose the term 'anaphoric' to describe the major syntactic type of relative clause I have reconstructed, I explicitly noted that this should not be taken as treating *REL as simply an anaphoric pronoun. My central point was that the relationship between *REL and the head noun and/or correlative demonstrative in the main clause could not be encoded by co-constituency; the relation is rather one of obligatory co-reference. While preposed relative clauses may lend themselves to an interpretation in which *REL was originally an indefinite or interrogative (i.e., **k^{w}í-/*k^{w}ó-*), postposed relatives might suggest a genuinely anaphoric origin (i.e., **yó-*, probably from **(e)y-*). However, I have argued precisely that pre- and postposed relative clauses are instantiations of a single underlying structure (anaphoric relative clauses); there is nothing in my reconstruction to suggest at a **syntactic** level that this emerges from two originally distinct structures. As such, aetiological questions on the nature of **yó-* vs. **k^{w}í-/*k^{w}ó-* are left open.

Another question, and one that affects PIE more broadly, is the possibility of 'embedding'. The arguments I have put forward in this thesis are not conclusive on the matter; while the anaphoric relative clauses I have reconstructed were adjoined at the CP-level, and could have generated structures that were reanalysed as embedded (i.e., DP-level) relative clauses, I cannot prove conclusively that PIE originally lacked embedded relative clauses. The attested languages appear to show contradictory trends: in some languages embedded relatives may appear archaic and recessive; in others, they gain ground and displace correlative clauses as an unmarked relativisation strategy. To me, this does not show a clear 'drift' in any particular direction, but suggests that some embedded structures may have existed marginally in PIE; this could suggest that embedding was a late-stage innovation, but it does not confirm it. Clearly I have not addressed forms of embedding other than relative clauses (e.g., indirect questions), which may also be treated under the methodology of Minimalist reconstruction, combining with the conclusions of this thesis to form a more comprehensive picture of embedding as a structural possibility in PIE.

Much more could be said about the left periphery and clisis in PIE. Activity at the 'front' of the clause certainly seems to be responsible for a whole host of different interpretational functions and surface word-order variation. With a finer understanding of the role "second position" clitics play in this left periphery, we may further refine the types of left peripheral position available in PIE. For example, I have addressed interrogative clauses only insofar as they differ from relative clauses; as another type of clause that implicates the left periphery (via *wh*-movement), there is certainly scope for further research

here. Further studies may also seek to establish more detailed arguments on the possibility of a distinct category of non-clitic 'weak' pronouns, looking beyond the correlative environments I have treated in this thesis, as well as a more detailed appraisal of non-Rizzian projections (e.g., ContrFocP) which merit further comparative analysis.

Clearly there is also scope for rigorous re-visiting of these hypotheses in light of the evidence from languages I have not had the space to treat within the bounds of a doctoral thesis. This applies both across language families (e.g., introducing more evidence from Tocharian languages) and within them (e.g., more evidence from Sabellic languages alongside Latin). According to the methodology of Minimalist reconstruction, the reanalyses that led from PIE to the attested languages must be established at each stage. With the introduction of more linguistic data comes yet a larger set of reanalyses to be posited; this in turn will shape our hypotheses on the proto-grammar. Minimalist analyses of other languages can therefore be used to refine the hypotheses I have proposed, or reject them.

This brings me to the second set of implications: those that concern syntactic reconstruction more broadly. I would argue that this thesis demonstrates the kind of concrete conclusions we may reach about a proto-language via Minimalist reconstruction. Perhaps the most unique contribution this thesis has made is in using Minimalist reconstruction to diagnose syntactic change between PIE and the attested languages. While previous authors have reconstructed syntactic change in other language families with a slightly less imposing time-depth, and others have reconstructed aspects of PIE syntax on the basis of identity, I have shown that Minimalist reconstruction offers a viable method for reconstructing PIE even when (or rather, especially when) the attested languages instantiate differing syntactic patterns. This can only be done by paying close attention to the process of syntactic change via reanalysis (and actualisation), coupled with the judicious employment of universal directionality in the form of grammaticalisation theory. While there are analogues here with the Comparative Method as used for phonological reconstruction, there are fundamental differences too. Building reanalyses into the methodology clearly introduces a level of subjectivity that is not present to the same degree in phonology: different researchers may have quite different ideas about what makes a given reanalysis 'plausible and motivated' (§2.1). On the other hand, we must not pretend that hypotheses on the phonology of PIE, even very basic questions about the phoneme inventory, are somehow free of such intellectual variety. If anything, that Minimalist reconstruction opens up the possibility of such debate, not in a vacuum but informed by the vast theoretical and typological literature in syntax, demonstrates its programmatic capacity.

Finally, I dedicate my last words in this thesis to the broader disciplinary question of what Syntactic Theory and Comparative Philology may have to say to each other in the years to come. In this course of this thesis, I have cited a number of scholars who have used contemporary syntactic theory as an analytical tool for the purpose of philological research; there are also linguists whose synchronic theories are informed by ancient languages in addition to contemporary varieties. Clearly, however, there remain several disciplinary barriers one must overcome in order to make these connections. Evaluating contemporary developments in syntactic theory while also providing adequately detailed analyses of philological data requires a commitment to the significance of both in addressing one's research question. In this thesis, I aim to have demonstrated that both are integral to syntactic reconstruction. Yet

clearly reconstruction is dependent on synchronic analyses of the attested IE languages: even outside a reconstructional context, therefore, the study of the syntax of ancient languages merits this two-fold commitment. To this end, I intend this thesis to be part of a growing literature that bridges such methodological divisions, and as a(nother) step towards consilience between theoretical and philological research.

***

References

Abe, Jun. 2001. Relativized X'-theory with symmetrical and asymmetrical structure. In *Minimalization of Each Module in Generative Grammar* (J. Abe, ed.), Nagoya, Japan: Nagoya University, Faculty of Language and Culture, pp. 1–38.

Abney, Steven P. 1987. *The English Noun Phrase in its Sentential Aspect*. Ph.D. thesis, MIT.

Adams, Douglas Q. 2015. *Tocharian B: A grammar of syntax and word formation*. Innsbruck: Beiträge zur Sprachwissenschaft.

Adams, James N. 1994. Wackernagel's law and the position of unstressed personal pronouns in Latin. *Transactions of the Philological Society*, 92(2): 103–78.

Adams, James N. 2016. *An Anthology of Informal Latin, 200 BC–AD 900: Fifty texts with translations and linguistic commentary*. Cambridge: CUP.

Aissen, Judith L. 1992. Topic and Focus in Mayan. *Language*, 68(1): 43–80.

Alexiadou, Artemis, Paul Law, André Meinunger & Chris Wilder. 2000. Introduction. In *The Syntax of Relative Clauses* (A. Alexiadou, P. Law, A. Meinunger & C. Wilder, eds.), Amsterdam: John Benjamins, pp. 1–52.

Alexopoulou, Theadora. 2007. Free and restrictive relative clauses in Greek. In *Proceedings of the 17th International Conference of Theoretical and Applied Linguistics* (E. Agathopoulou, M. Dimitrakopoulou & D. Papadopoulou, eds.), Thessaloniki: Monochromia, pp. 425–37.

Allerton, David J. 1975. Deletion and proform reduction. *Journal of Linguistics*, 11(2): 213–37.

Andersen, Henning. 1973. Abductive and deductive change. *Language*, 49: 765–793.

Andrews, Avery D. 1975. *Studies in the Syntax of Relative and Comparative Clauses*. Ph.D. thesis, MIT, Cambridge, MA.

Andrews, Avery D. 2007. Relative clauses. In *Language Typology and Syntactic Description. Vol. 2.* (T.Shopen, ed.), Cambridge: CUP.

Aoun, Joseph E. & Yen-Hui A. Li. 2003. *Essays on the Representational and Derivational Nature of Grammar: The diversity of wh-constructions*. Cambridge, MA: MIT Press.

Arsenijević, Boban. 2009. {Relative {conditional {correlative clauses}}}. In *Correlatives Cross-linguistically* (A. Lipták, ed.), Amsterdam: Benjamins, pp. 131–56.

Assmann, Anke. 2013. Three stages in the derivation of free relatives. In *Rule Interaction in Grammar* (F. Heck & A. Assmann, eds.), Leipzig: University of Leipzig Press, pp. 203–45.

Bader, Françoise. 1975. Mycénien *to-so-de*, *to-so-jo*. *Minos*, 14: 85–109.

Bains, Gurpreet. 1989. *Complex Structures in Hindi-Urdu: Explorations in government and binding theory*. Ph.D. thesis, New York University, NY.

Baker, Mark C. 2008. The macroparameter in a microparametric world. In *The Limits of Syntactic Variation* (T. Biberauer, ed.), Amsterdam: John Benjamins, pp. 351–373.

Balles, Irene. 2008. Syntactic reconstruction and "morphology as paleosyntax". In *Principles of Syntactic Reconstruction* (G. Ferraresi & M. Goldbach, eds.), Amsterdam: John Benjamins, pp. 161–86.

Ballester, Xaverio. 2001. A propósito del adjetivo determinado en lituano. *Res Balticae*, 7: 67–80.

Barðdal, Jóhanna, Valgerður Bjarnadóttir, Serena Danesi, Tonya K. Dewey, Thórhallur Eythórtsson, Chiara Fedriani & Thomas Smitherman. 2013. The story of 'woe'. *Journal of Indo-European Studies*, 41(3–4): 321–77.

Beekes, Robert. 2010. *Etymological Dictionary of Greek*. Leiden: Brill.

Beekes, Robert. 2011. *Comparative Indo-European Linguistics: An introduction*. Amsterdam: John Benjamins.

Belletti, Adriana. 2001. 'Inversion' as focalization. In *Subject Inversion in Romance and the Theory of Universal Grammar* (A. Hulk & J.-Y. Pollock, eds.), Oxford: OUP, pp. 60–90.

Belletti, Adriana. 2004. Aspects of the low IP area. In *The Structure of CP and IP* (L. Rizzi, ed.), Oxford: OUP, pp. 16–51.

Beltrán, Antonino & Antonio Tovar. 1982. *Contrebia Belaisca I. El bronce con alfabeto 'ibérico' de Botorrita*. Zaragoza: Universidad de Zaragoza.

Beltrán, Francisco & Carlos Jordán. 2019. Writing and language in Celtiberia. In *Palaeohispanic Languages and Epigraphies* (A. G. Sinner & J. Velaza, eds.), Oxford: OUP, pp. 240–303.

Benincà, Paola & Cecilia Poletto. 2004. Topic, focus and V2: Defining the CP sublayers. In *The Structure of CP and IP* (L. Rizzi, ed.), Oxford: OUP, pp. 52–75.

Benveniste, Émile. 1957/1958. La phrase relative, problème de syntaxe générale. *Bulletin de la sociétée linguistique de Paris*, 53(1): 39–54.

Benveniste, Émile. 1966. *Problèmes de linguistique générale*. Paris: Gallimard.

Bertocchi, Alessandra. 1989. The role of antecedents in Latin anaphors. In *Subordination and Other Topics in Latin* (G. Calboli, ed.), Amsterdam: John Benjamins, pp. 441–62.

Bertrand, Nicolas. 2014. On tmesis, word order and noun incorporation in Homeric Greek. In *The Greek Verb: Morphology, syntax and semantics* (A. Bartolotta, ed.), Louvain-La-Neuve/Walpole, MA: Peeters, pp. 11–30.

Bhatt, Rajesh. 2002. The raising analysis of relative clauses: Evidence from adjectival modification. *Natural Language Semantics*, 10: 43–90.

Bhatt, Rajesh. 2003. Locality in correlatives. *Natural Language & Linguistic Theory*, 210: 485–541.

Bhatt, Rajesh. 2015. Relative clauses and correlatives. In *Syntax: Theory and Analysis. Vol. 1.* (T. Kiss & A. Alexiadou, eds.), Berlin: De Gruyter, pp. 708–49.

Bianchi, Valentina. 1999. *Consequences of Antisymmetry: Headed relative clauses*. Berlin/New York, NY: Mouton de Gruyter.

Bianchi, Valentina. 2000. Some issues in the syntax of relative determiners. In *The Syntax of Relative Clauses* (A. Alexiadou, P. Law, A. Meinunger & C. Wilder, eds.), Amsterdam: John Benjamins, pp. 53–82.

Biberauer, Theresa. 2011. In defence of lexico-centric parametric variation: Two 3rd factor-constrained case studies. *Workshop on Formal Grammar and Syntactic Variation: Rethinking Parameters*, Madrid.

Biberauer, Theresa, Anders Holmberg, Ian Roberts & Michelle Sheehan. 2014. Complexity in comparative syntax: the view from modern parametric theory. In *Measuring Linguistic Complexity* (F. Newmeyer & L. Preston, eds.), Oxford: OUP, pp. 102–127.

Biberauer, Theresa & Ian Roberts. 2012a. Towards a parameter hierarchy for auxiliaries: Diachronic considerations. *Cambridge Occasional Papers in Linguistics*, 6(9): 267–294.

Biberauer, Theresa & Ian Roberts. 2012b. On the significance of what doesn't happen. *14th Diachonric Generative Syntax Conference* (*DiGS 14*), Lisbon.

Biberauer, Theresa & Ian Roberts. 2017. Parameter setting. In *The Cambridge Handbook of Historical Syntax* (A. Ledgeway & I. Roberts, eds.), Cambridge: CUP, pp. 113–133.

Boeckx, Cedric. 2006. *Linguistic Minimalism: Origins, concepts, methods, and aims*. Oxford: OUP.

Boeckx, Cedric. 2007. *Understanding Minimalist Syntax: Lessons from locality in long-distance dependencies*. Oxford: Blackwell.

Borer, H. 1984. *Parametric Syntax*. Dordrecht: Foris.

Bortolussi, Bernard. 2017. Topicalizations, left dislocations and the left periphery. *Catalan Journal of Linguistics*, 16: 101–23.

Brame, Michael K. 1968. A New Analysis of the Relative Clause: Evidence for an interpretive theory. Unpublished MS, MIT.

Browning, M. 1991. *Null Operator Constructions*. New York, NY: Garland.

Brugmann, Karl. 1886–1893. *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen* (*Vols. 1–2*). Strasbourg: Karl J. Trübner.

Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In *Eurotyp: Typology of languages in Europe. Vol. 5.* (H. van Riemsdijk, ed.), Berlin/New York, NY: De Gruyter, pp. 145–234.

Carlson, Greg N. 1977. Amount relatives. *Language*, pp. 520–542.

Chantraine, Pierre. 1968. *Dictionairre Étymologique de la Langue Grecque: Histoire des mots*. Paris: Éditions Klincksieck.

Chatzopoulou, Katerina. 2018. *Negation and Nonveridicality in the History of Greek*. Oxford: OUP.

Chierchia, Gennaro. 1995. *The Dynamics of Meaning: Anaphora, presupposition, and the theory of grammar*. Chicago, IL: University of Chicago Press.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. *Lectures in Government and Binding: The Pisa lectures*. Dordrecht: Foris.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT press.

Chomsky, Noam & Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry*, 8: 425–504.

Chomsky, Noam & Howard Lasnik. 1993. The theory of principles and parameters. In *Syntax: An international handbook of contemporary research* (J. Jacobs, A. von Stechow, W. Sternefeld & T. Vennemann, eds.), Berlin: de Gruyter, pp. 506–69.

Cinque, Guglielmo. 1994. On the evidence for partial N-movement in the Romance DP. In *Paths Towards Universal Grammar: Studies in honor of Richard S Kayne* (G. Cinque, G. Koster, J.-Y. Pollock, L. Rizzi & R. Zanuttini, eds.), Washington, D.C.: Georgetown University Press, pp. 85–110.

Cinque, Guglielmo. 1996. The 'antisymmetric' programme: Theoretical and typological implications. *Journal of Linguistics*, 32(2): 447–464.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A cross-linguistic perspective*. Oxford: OUP.

Cinque, Guglielmo. 2004. "Restructuring" and functional structure. In *Structures and Beyond: The cartography of syntactic structures. Vol. 3.* (Adriana Belletti, ed.), Oxford: OUP, pp. 132–92.

Cinque, Guglielmo. 2010. *The Syntax of Adjectives: A comparative study*. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 2013. *Typological Studies: Word order and relative clauses*. New York, NY/London: Routledge.

Cinque, Guglielmo. 2020. *The Syntax of Relative Clauses: A unified analysis*. Cambridge: CUP.

Citko, Barbara. 2001. Deletion under identity in relative clauses. *North-Eastern Linguistic Society*, 31: 131–45.

Citko, Barbara. 2009. What don't *wh*-questions, free relatives, and correlatives have in common? In *Correlatives Cross-linguistically* (A Lipták, ed.), Amsterdam: Benjamins, pp. 49–80.

Clackson, James. 2007. *Indo-European Linguistics: An introduction*. Cambridge: CUP.

Clackson, James. 2017. The Comparative Method and comparative reconstruction. In *The Cambridge Handbook of Historical Syntax* (A. Ledgeway & I. Roberts, eds.), Cambridge: CUP, pp. 189–206.

Cole, Peter. 1987. The structure of internally headed relative clauses. *Natural Language and Linguistic Theory*, 5(2): 277–302.

Danckaert, Lieven. 2012. *Latin Embedded Clauses: The left periphery*. Amsterdam: John Benjamins.

Davison, Alice. 2009. Adjunction, features and locality in Sanskrit and Hindi/Urdu correlatives. In *Correlatives Cross-linguistically* (A. Lipták, ed.), Amsterdam: John Benjamins, pp. 223–262.

Dayal, Veneeta. 1996. *Locality in Wh-Quantification*. Dordrecht: Kluwer.

De Vaan, Michiel. 2015. Reduplicated demonstratives in ancient Indo-European languages. *Transactions of the Philological Society*, 113(1): 38–51.

de Vries, Mark. 2002. *The Syntax of Relativization*. Ph.D. thesis, University of Amsterdam.

de Vries, Mark. 2006. The syntax of appositive relativization: On specifying coordination, false free relatives, and promotion. *Linguistic Inquiry*, 37(2): 229–70.

del Gobbo, Francesca. 2007. On the syntax and semantics of appositive relative clauses. In *Parentheticals* (N. Dehé & Y. Kavalova, eds.), Amsterdam: John Benjamins, pp. 173–202.

del Gobbo, Francesca. 2017. More appositives in heaven and earth than are dreamt of in your Linguistics. *Glossa*, 2(1): 1–38.

Delbrück, Berthold. 1878. *Syntaktische Forschungen, Vol. 3: Die altindische Wortfolge aus dem Çatapathabrāhmaṇa*. Halle: Verlag der Buchhandlung des Waisenhauses.

Delbrück, Berthold. 1893–1900. *Vergleichende Syntax der indogermanischen Sprachen*. Strasbourg: Karl J.Trübner. 3 vols.

Demirdache, H. 1991. *Resumptive Chains in Restrictive Relatives, Appositives and Dislocation Structures*. Ph.D. thesis, MIT.

den Dikken, Marcel. 2009. Comparative correlatives and successive cyclicity. In *Correlatives Cross-linguistically* (A Lipták, ed.), Amsterdam: Benjamins, pp. 263–308.

Derksen, Rick. 2007. *Etymological Dictionary of the Slavic Inherited Lexicon*. Leiden: Brill.

Derksen, Rick. 2014. *Etymological Dictionary of the Baltic Inherited Lexicon*. Leiden: Brill.

Devine, Andrew M. & Laurence D. Stephens. 2006. *Latin Word Order: Structured meaning and information*. Oxford: OUP.

Dik, Helma. 1995. *Word Order in Ancient Greek: A pragmatic account of word order variation in Herodotus*. Amsterdam: Gieben.

Douglas, Jamie A. 2016. *The Syntactic Structures of Relativisation*. Ph.D. thesis, University of Cambridge.

Dunkel, George E. 1979. Preverb repetition. *Münchener Studien zur Sprachwissenschaft*, 38: 41–82.

Dunkel, George E. 1982. IE conjunctions: Pleonasm, ablaut, suppletion. *Zeitschrift für vergleichende Sprachforschung*, pp. 178–99.

Enrique-Arias, Andrés. 2005. Using universal principles of phonetic qualitative reduction in grammaticalization to explain the Old Spanish shift from *ge* to *se*. In *Historical Linguistics 2003: Selected papers from the 16th international conference on Historical Linguistics, Copenhagen, 11–15 August 2003* (M. Fortescue, E. S. Jensen, J. E. Mogensen & L. Schøsler, eds.), Amsterdam: John Benjamins, pp. 103–14.

Ernst, Thomas. 2004. *The Syntax of Adjuncts*. Cambridge: CUP.

Eska, Joseph F. 1995. Rethinking the evolution of Celtic constituent configuration. *Münchener Studien zur Sprachwissenschaft*, 55: 7–39.

Evans, Gareth. 1980. Pronouns. *Linguistic Inquiry*, 11: 227–62.

Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: OUP.

Flier, Michael S. 1974. *Aspects of Nominal Determination in Old Church Slavic*. The Hague: Mouton.

Forbes, Kathleen. 1958. The Relations of the particle ἄν with κε(ν) κα καν. *Glotta*, 37(3/4): 179–182.

Fraenkel, Eduard. 1964. *Kleine Beiträge zur klassischen Philologie I*. Rome: Edizioni di storia e letteratura.

Fraenkel, Ernst. 1960. *Litauisches Etymologysches Wörterbuch: Band I.* Heidelberg: Carl Winter Verlag.

Fraenkel, Ersnt. 1947. *Sprachliche, besonders syntaktische Untersuchung des kalvinistischen litauischen Katechismus des M. Pietkiewicz von 1598.* Göttingen: Vandenhœck & Ruprecht.

Frascarelli, Mara. 2000. *The Syntax-Phonology Interface in Focus and Topic Constructions in Italian*. Dordrecht: Springer Science+Business Media.

Frazier, Lyn. 2013. Syntax in sentence processing. In *Sentence Processing* (Roger P. G. van Gompel, ed.), London/New York, NY: Pyschology Press (Taylor & Francis), pp. 21–50.

Friedrich, Johannes. 1960. *Hethitisches Elementarbuch*. Heidelberg: Carl Winter Verlag.

Friedrich, Paul. 1975. *Proto-Indo-European Syntax: The order of the meaningful elements*. Butte, MT: Journal of Indo-European Studies.

García-Castillero, Carlos. 2020. *Clause Typing in the Old Irish Verbal Complex*. Berlin/Boston, MA: De Gruyter Mouton.

Garrett, Andrew. 1990. Hittite enclitic subjects and transitive verbs. *Journal of Cuneiform Studies*, 42: 227–42.

Garrett, Andrew. 1994. Relative clause syntax in Lycian and Hittite. *Die Sprache*, 36: 29–69.

Garrett, Andrew. 1996. Wackernagel's law and unaccusativity in Hittite. In *Approaching Second: Second position clitics and related phenomena* (A. Halpern & A. M. Zwicky, eds.), Stanford, CA: CLSI Publications, pp. 85–134.

Gianollo, Chiara. 2021. Indefinites and negation in Ancient Greek. *Journal of Historical Syntax*, 5: 1–38.

Gildea, Spike, Eugenio R. Luján & Jóhanna Barðdal. 2020. The curious case of reconstruction in syntax. In *Reconstructing Syntax* (J. Barðdal, S. Gildea & E. R. Luján, eds.), Leiden: Brill, pp. 1–44.

Goedegebuure, Petra. 2009. Focus structure and Q-word questions in Hittite. *Linguistics*, 47(4): 945–69.

Goedegebuure, Petra. 2014. *The Use of Demonstratives in Hittite*. Wiesbaden: Harrassowitz Verlag.

Goldstein, David. 2010. *Wackernagel's Law in Fifth-Century Greek*. Ph.D. thesis, University of California, Berkley.

Goldstein, David. 2016a. *Classical Greek Syntax: Wackernagel's law in Herodotus*. Leiden: Brill.

Goldstein, David. 2016b. Variation versus change: Clausal clitics between Homer and Herodotus. *Indo-European Linguistics*, 4: 53–97.

Gonda, Jan. 1954. The original character of the Indo-European relative pronoun *i̯o*. *Lingua*, 4: 1–41.

Gonda, Jan. 1955. Notes on the Indo-European $k^{u}i$- and $k^{u}o$- pronouns. *Lingua*, 4: 241–285.

Grassmann, Hermann. 1873. *Wörterbuch zum Rig-Veda*. Leipzig: F A Brockhaus.

Grice, Paul H. 1975. Logic and conversation. In *Syntax and Semantics, 3: Speech acts* (P. Cole & J. Morgan, eds.), New York, NY: Academic Press, pp. 41–58.

Grosu, Alexander. 2002. Strange relatives at the interface of two millennia. *GLOT International*, 6: 145–67.

Grosu, Alexander & Fred Landman. 1998. Strange relatives of the third kind. *Natural Language Semantics*, 6: 125–170.

Hackstein, Olav. 2013. The evolution of finite complementation in Tocharian. *Tocharian and Indo-European Studies*, 13: 117–48.

Haegeman, Liliane. 2012. *Adverbial Clauses, Main Clause Phenomena, and the Composition of the Left Periphery*. Oxford: OUP.

Haegeman, Liliane. 2014. West Flemish verb-based discourse markers and the articulation of the speech act layer. *Studia Linguistica*, 68(1): 116–39.

Hahn, E Adelaide. 1933. Light from Hittite on Latin indefinites. *Transactions and Proceedings of the American Philological Association*, 64: 28–40.

Hahn, E. Adelaide. 1937. Hittite *kwis kwis*. *Transactions and Proceedings of the American Philological Association*, 68: 388–402.

Hahn, E Adelaide. 1946. The origin of the relative $k^{w}i$- $k^{w}o$-. *Language*, 22: 68–85.

Hahn, E Adelaide. 1964. Relative and antecedent. *Transactions of the American Philological Association*, 95: 111–141.

Hahn, E Adlaide. 1949. The non-restrictive relative in Hittite. *Language*, 25: 346–374.

Haider, Hubert. 2000. OV is more basic than VO. In *The Derivation of VO and OV* (P. Svevonius, ed.), Amsterdam: John Benjamins, pp. 45–67.

Hajnal, Ivo. 1997. Definite nominale Dermination im Indogermanischen. *Indogermanische Forschungen*, 102: 38–73.

Hale, Kenneth. 1975. Gaps in grammar and culture. In *Linguistics and Anthropology: In honor of C. F Voegelin* (M. D. Kinkade, K. Hale & O. Werner, eds.), Lisse: de Ridder.

Hale, Mark. 1987. *Studies in the Comparative Syntax of the Oldest Indo-Iranian Languages*. Ph.D. thesis, Harvard University, Cambridge, MA.

Hale, Mark. 1996. Deriving Wackernagel's Law: Prosodic and syntactic factors determining clitic placement in the language of the Rigveda. In *Approaching Second: Second position clitics and related phenomena* (A. L. Halpern & A. M. Zwicky, eds.), Stanford, CA: Center for the Study of Language and Information, pp. 165–97.

Hale, Mark. 2007. *Historical Linguistics: Theory and method*. Oxford: Blackwell.

Hale, Mark. 2017. Preliminaries to the investigation of clitic sequencing in Greek and Indo-Iranian. In *On Looking Into Words (and Beyond): Structures, relations, analyses* (C. Bowern, L. Horn & R. Zanuttini, eds.), Berlin: Language Science Press, pp. 289–310.

Halla-aho, Hilla. 2018. *Left-Dislocation in Latin: Topics in syntax in Republican texts*. Leiden/Boston, MA: Brill.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The View from Building 20* (K. Hale & S. J. Keyser, eds.), Cambridge, MA: MIT Press, pp. 111–76.

Halle, Morris & Alex Marantz. 1994. Some key features of distributed morphology. In *MIT working papers in Linguistics 21: Papers on phonology and morphology* (A. Carnie, H. Harley & T. Bures, eds.), Cambridge, MA: MIT Press, pp. 275–88.

Halpern, Aaron L. 1992. *Topics in the Placement and Morphology of Clitics*. Ph.D. thesis, Stanford University, Stanford, CA.

Halpern, Aaron L. 1995. *On the Placement and Morphology of Clitics*. Stanford, CA: CLSI Publications.

Harris, Alice C. 2008. Reconstruction in syntax: Reconstruction of patterns. In *Principles of Syntactic Reconstruction* (G. Ferraresi & M. Goldbach, eds.), Amsterdam: John Benjamins, pp. 73–95.

Harris, Alice C. & Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: CUP.

Haspelmath, Martin. 1997. *Indefinite Pronouns*. Oxford: OUP.

Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In *Up and down the cline: The nature of grammaticalization* (O.Fischer, M. Norder & H. Perridon, eds.), Amsterdam: John Benjamins, pp. 17–44.

Haudry, Jean. 1973. Parataxe, hypotaxe et corrélation dans la phrase latine. *Bulletin de la Société de Linguistique de Paris*, 68: 147–186.

Haudry, Jean. 1979. L'antétposition de la relative en indo-européen. *La linguistique*, 15: 101–110.

Heim, Irene. 1987. Where does the definiteness restriction apply? Evidence from the definiteness of variables. In *The representation of (in) definiteness* (E. Reuland & A. Meulen, eds.), Cambridge, MA: MIT Press, volume 14, pp. 21–42.

Held, Warren H. 1957. The Hittite relative sentence. *Language*, 33(4): 7–52.

Hendery, Rachel. 2012. *Relative Clauses in Time and Space: A case study in the methods of diachronic typology*. Philadelphia, PA: John Benjamins.

Hermann, Eduard. 1912. *Über die Entwicklung der litauischen Konjuktionalsätze*. Jena: Frommansche Buchdruckerei.

Hettrich, Heinrich. 1988. *Untersuchungen zur Hypotaxe im Vedischen*. Berlin/New York, NY: Walter de Gruyter.

Hickey, Raymond. 2002. Internal and external forces again: Changes in word order in Old English and Old Irish. *Language Sciences*, 24(3): 261–83.

Hill, Virginia. 2007. Vocatives and the pragmatics-syntax interface. *Lingua*, 117: 2077–105.

Hirt, Hermann. 1937. *Indogermanische Grammatik. Teil VII: Syntax II Die Lehre vom einfachen und zusammengesetzen Satz*. Heidelberg: Carl Winter Verlag.

Hock, Hans H. 1982. Clitic verbs in PIE or discourse-based verb fronting? Sanskrit *sá hovāca gā́rgyaḥ* and congeners in Avestan and Homeric Greek. *Studies in the Linguistic Sciences*, 12(2): 1–38.

Hock, Hans H. 1989. Conjoined we stand: Theoretical implications of Sanskrit relative structures. *Studies in the Linguistic Sciences*, 19(1): 93–126.

Hock, Hans H. 1993. Some peculiarities of Vedic-prose relative clauses. In *Proceedings of the VIIIth World Sanskrit Conference, Vienna 1990* (G. Oberhammer & R. Mesquita, eds.), Vienna: Verlag der Österreichischen Akademie der Wissenschaft, pp. 19–29. (Wiener Zeitschrift für die Kunde Südasiens, Supplement to vol. 36).

Hock, Hans H. 1996. Who's on first? Toward a prosodic account of P2 clitics. In *Approaching Second: Second position clitics and related phenomena* (A. L. Halpern & A. M. Zwicky, eds.), Stanford, CA: Center for the Study of Language and Information, pp. 199–270.

Hock, Hans H. 1997. Chronology or genre? Problems in Vedic syntax. In *Inside the Texts, Beyond the Texts: New approaches to the study of the Vedas. Proceedings of the International Vedic Workshop, Harvard University, June 1989* (M. Witzel, ed.), Cambridge, MA/Columbia, MO: Department of Sanskrit and Indian Studies, Harvard University (distributed by South Asia Books).

Hock, Hans H. 2000. Genre, discourse, and syntax in early Indo-European, with emphasis on Sanskrit. In *Textual parameters in older languages* (S. C. Herring, P. van Reenen & L. Schøsler, eds.), Amsterdam: Benjamins, pp. 163–95.

Hock, Hans H. 2015. Proto-Indo-European verb-finality: Reconstruction, typology, validation. In *Proto-Indo-European Syntax and Its Development* (L. Kulikov & N. Lavidas, eds.), Amsterdam: John Benjamins, pp. 51–78.

Hoffman, Johann B. 1950. *Etymologisches Wörterbuch des Griechischen*. Munich: R. Oldenbourg.

Hoffner, Harry A., Jr. 1995. About Questions. In *Studio Historiae Ardens: Ancient Near Eastern Studies Presented to Philo H. J. Houwink ten Cate on the Occasion of His 65th Birthday* (T. P. J. van de Hout & J. de Roos, eds.), Leiden: Nederlands Historisch-Archeologisch Institut te Istanbul, pp. 87–104.

Hoffner, Harry A., Jr. & H. Craig Melchert. 2008. *A Grammar of the Hittite Language. Part I: Reference Grammar*. Winona Lake, IN: Eisenbrauns.

Hofling, C. Andrew. 2000. *Itzaj Maya Grammar*. Salt Lake City, UT: University of Utah Press.

Hopper, Paul J. & Elizabeth C. Traugott. 1993. *Grammaticalization*. Cambridge: CUP.

Horrocks, Geoffrey. 2011. Latin syntax. In *A Companion to the Latin language* (J. Clackson, ed.), Chichester: Wiley-Blackwell, pp. 118–143.

Huggard, Mattyas. 2011. On Wh-(non)-movement and internal structures of the Hittite preposed relative clause. In *Proceedings of the 22nd Annual UCLA Indo-European Conference* (S. W. Jamison, H. C. Melchert & B. Vine, eds.). Bremen: Hempen Verlag, pp. 83–104.

Huggard, Mattyas. 2015. *Wh-words in Hittite: A study in syntax-semantics and syntax-phonology interfaces*. Ph.D. thesis, UCLA.

Hulsey, Sarah & Uli Sauerland. 2006. Sorting out relative clauses. *Natural Language Semantics*, 14: 111–37.

Izvorski, Roumyana. 1996. The syntax and semantics of correlative proforms. In *Proceedings of NELS 26* (K. Kusumoto, ed.), Amherst, MA: GLSA, pp. 133–47.

Jacobson, Pauline. 1995. On the quanitificational force of English free relatives. In *Quantification in Natural Languages* (E. Bach, E. Jelinek, A. Kratzer & B. H. Partee, eds.), Dordrecht: Kluwer Academic Publishers, pp. 451–86.

Jagić, Vatroslav. 1879. *Quattuor evangeliorum Codex glagoliticus olim Zographensis nunc Petropolitanus*. Berlin: Weidmann.

Jasanoff, Jay. 1999. Some relative forms of the verb in Old Irish. In *Compositiones indogermanicae: in memoriam Jochem Schindler* (H. Eichner, H. C. Luschützky & E. Sadovski, eds.), Prague: Enigma, pp. 205–21.

Joseph, Brian D. & Richard Janda. 1988. The how and why of diachronic morphologization and demorphologization. In *Theoretical Morphology: Approaches in modern linguistics* (M. Hammon & M. Noonan, eds.), New York, NY: Academic Press, pp. 193–210.

Junghare, Indira. 1973. Restrictive relative clauses in Marathi. *Indian Linguistics*, 34: 251–262.

Kachru, Yamuna. 1973. Some aspects of pronominalization and relative clause construction in Hindi-Urdu. In *Papers on South Asian Linguistics* (B. B. Kachru, ed.), Urbana, IL: University of Illinois, pp. 87–103.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Kayne, Richard S. 2014. Why isn't *this* a complementizer? In *Functional Structure from Top to Toe* (P. Svevonius, ed.), Oxford: OUP, pp. 188–231.

Kennedy, Benjamin H. 1962. *Revised Latin Primer*. London: Longmans.

Kerr, Judith. 1968. *The Tiger Who Came to Tea*. New York, NY: HarperCollins.

Keydana, Götz. 2011. Wackernagel in the language of the Rigveda: A reassessment. *Historische Sprachforschung*, pp. 80–107.

Keydana, Götz. 2018. The syntax of Proto-Indo-European. In *Handbook of Comparative and Historical Indo-European Linguistics. Vol. 3* (J. Klein, B. Joseph & M. Fritz, eds.), Berlin/Boston, MA: Mouton de Gruyter, pp. 2195–228.

Kiparsky, Paul. 1995. Indo-European origins of Germanic syntax. In *Clause structure and Language Change* (A. Battye & I. Roberts, eds.), Oxford: OUP, pp. 140–69.

Klein, Jared. 2003. Āmreḍitas and related constellations in the Rigveda. *Journal of the American Oriental Society*, 123(4): 773–802.

Kloekhorst, Alwin. 2007. *Etymological Dictionary of the Hittite Inherited Lexicon*. Leiden.

Kloekhorst, Alwin. 2014. *Accent in Hittite: A study in plene spelling, consonant gradation, clitics and metrics*. Wiesbaden: Harrassowitz Verlag.

Koch, Christoph. 1992. Zur Vorgeschichte des relativen Attributivkonnexes im Baltischen und Slavischen. In *Indogermanisch, Slawisch und Baltisch: Materialen des vom 21.–22. September 1989 in Jena in Zusammenarbeit mit der Indogermanischen Gesellschaft durchgeführten Kolloquiums* (B. Barschel, ed.), Munich: Verlag Otto Sagner.

Kratzer, Angelika. 1999. Beyond 'oops' and 'ouch'. How descriptive and expressive meaning interact. *Cornell Conference on Theories of Context Dependency*, Ithaca, NY, 26th March 1999.

Krisch, Thomas. 1990. Das Wackernagelsche Gesetz aus heutiger Sicht. In *Sprachwissenschaft und Philologie. Jacob Wackernagel und die Indogermanistik heute. Kolloquium der Indogermanischen Gesellschaft, Basel, 1988* (H. Eichner & H. Rix, eds.), Wiesbaden: Reichert, pp. 64–81.

Krisch, Thomas. 2017. Proto-Indo-European syntax. In *The Indo-European Languages* (M. Kapović, A. G. Ramat & P. Ramat, eds.), London/New York, NY: Routledge, pp. 111–52. 2nd edition.

Kroll, Wilhelm. 1910. Der lateinische Relativsatz. *Glotta*, 3(1): 1–18.

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World Lexicon of Grammaticalization*. Cambridge: CUP, 2nd edition.

Ladusaw, William A. 1994. Thetic and categorical, stage and individual, weak and strong. In *Proceedings from Semantics and Linguistic Theory IV* (M. Harvey & L. Santelmann, eds.), Ithaca, NY: Cornell University, Department of Modern Languages and Linguistics, pp. 220–29.

Lambert, Pierre-Yves. 1994. *La langue gauloise*. Paris: editions errance.

Langacker, Ronald W. 1977. Syntactic reanalysis. In *Mechanisms of Syntactic Change* (C. N. Li, ed.), Austin, TX: University of Texas Press, pp. 57–139.

Ledgeway, Adam. 2018. On the decline of edge-fronting from Latin to Romance. In *Word Order Change* (A. M. Martins & A. Cardoso, eds.), Oxford: OUP, pp. 264–278.

Lees, Robert B. 1960. *The Grammar of English Nominalizations*. The Hague: Mouton.

Lees, Robert B. 1961. The constituent structure of noun phrases. *American Speech*, 36: 159–68.

Lehmann, Christian. 1980. Der indogermanische **kwi-/kwo-* Relativsatz im typologischen Vergleich. In *Linguistic Reconstruction and Indo-European Syntax*, Amsterdam: John Benjamins, pp. 155–182.

Lehmann, Christian. 1984. *Der Relativsatz: Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik*. Tübingen: Gunter Narr Verlag.

Lehmann, Winfred P. 1974. *Proto-Indo-European Syntax*. Houston, TX: University of Texas Press.

Lewis, Henry & Holger Pedersen. 1937. *A Concise Comparative Celtic Grammar*. Göttingen: Vandenhoeck & Ruprecht.

Lightfoot, David W. 1979. *Principles of Diachronic Syntax*. Cambridge: CUP.

Lightfoot, David W. 2002. Myths and the Prehistory of Grammars. *Journal of Linguistics*, 38(1): 113–136.

Lodge, Gonzalez. 1962. *Lexicon Plautinum. Vol. 2*. Hildesheim: Georg Olms Verlagsbuchhandlung, 2nd edition.

Lowe, John J. 2011. Ṛgvedic clitics and 'prosodic movement'. In *Proceedings of the LFG11 Conference* (M. Butt & T. Holloway King, eds.), Stanford, CA: CSLI Publications, pp. 360–80.

Lowe, John J. 2014. Accented clitics in the Ṛgveda. *Transactions of the Philological Society*, 112(1): 5–43.

Lubotsky, Alexander. 1988. The Old Phrygian Areyastis-inscription. *Kadmos*, 27(1): 9–26.

Luján, Eugenio R. 2009. On the grammaticalization of **k^{w}i-/* k^{w}o*-relative clauses in Proto-Indo-European. In *Grammatical change in Indo-European Languages* (V. Bubenik, J. Hewson & S. Rose, eds.), Amsterdam: John Benjamins, pp. 221–234.

Lunt, Horace G. 2001. *Old Church Slavonic Grammar*. Berlin/New York, NY: Mouton de Gruyter, 7th edition.

Luraghi, Silvia. 1990. *Old Hittite Sentence Structure*. London/New York, NY: Routledge.

Luraghi, Silvia. 1998. The grammaticalization of the left sentence boundary in Hittite. In *The Limits of Grammaticalization* (A. G. Ramat & P. Hopper, eds.), Amsterdam: John Benjamins, pp. 189–210.

Luraghi, Silvia. 2001. The development of local particles and adverbs in Anatolian as a grammaticalization process. *Diachronica*, 18(1): 31–58.

Lutz, Uli, Gereon Müller & Arnim von Stechow. 2000. Introduction. In *Wh-Scope Marking* (U. Lutz, G. Müller & A. von Stechow, eds.), Amsterdam: John Benjamins, pp. 1–16.

Macdonell, Arthur A. 1916. *A Vedic grammar for students*. Oxford: Clarendon Press.

Mahajan, Anoop. 2000. Relative asymmetries in Hindi correlatives. In *The Syntax of Relative Clauses* (A. Alexiadou, P. Law, A. Meinunger & C. Wilder, eds.), Amsterdam: John Benjamins, pp. 201–229.

Mańczak, Witold. 1980. Irregular Sound Change due to Frequency in German. In *Historical Morphology* (J. Fisiak, ed.), The Hague: Mouton, pp. 309–19.

Manolessou, Io & Geoffrey Horrocks. 2007. The development of the definite article in Greek. *Studies in Greek Lingistics*, 27: 224–36.

Martirosyan, Hrach. 2009. *Etymological Dictionary of the Armenian Inherited Lexicon*. Leiden/Boston, MA/Cologne: Brill.

Matić, Dejan. 2003. Topic, Focus, and discourse structure. Ancient Greek word order. *Studies in Language*, 27: 573–633.

McCone, Kim. 1997. *The Early Irish Verb*. Maynooth: Department of Old Irish, National University of Ireland.

McFadden, Thomas. 2004. *The position of Morphological Case in the derivation: A study on the syntax-morphology interface*. Ph.D. thesis, University of Pennsylvania.

Meillet, Antoine. 1912. L'évolution des formes grammaticales. Reprinted in *Linguistique Historique et Linguistique Générale* (A. Meillet, ed.). 1948. Paris: Champion, pp. 130–158.

Melchert, H. Craig. 1985. Hittite *imma* and Latin *immō*. *Historische Sprachforschung*, 98(2): 184–205.

Melchert, H. Craig. 2017. Anatolian. In *The Indo-European Languages* (M. Kapović, ed.), London/New York, NY: Routledge, pp. 171–204. 2nd edition.

Meyer, Robin. 2013. *Classical Armenian Relative Clause Syntax*. Master's thesis, University of Oxford.

Miklosich, Franz. 1868–1879. *Vergleichende Grammatik der slavischen Sprachen* (*4 Vols.*). Vienna: Wilhelm Braumüller.

Miller, Gary. 1975. IE: VSO, SOV, SVO, or all three? *Lingua*, 37: 31–52.

Milsark, Gary. 1974. *Existential Sentences in English*. Ph.D. thesis, MIT.

Moorhouse, Alfred C. 1959. *Studies in the Greek Negatives*. Cardiff: University of Wales Press.

Munaro, Nicola. 2006. Looking for the atoms of the left periphery: On the interface legibility of interjections and particles. *Interface Legibility at the Edge*, Bucharest, 25th–27th June 2006.

Neeleman, Ad & Reiko Vermeulen. 2012. The syntactic expression of information structure. In *The Syntax of Topic, Focus and Contrast: An interface-based approach* (A. Neeleman & R. Vermeulen, eds.), Berlin/Boston, MA: De Gruyter Mouton, pp. 1–38.

Nespor, Marina & Irene Vogel. 2007. *Prosodic Phonology. With a new forward*. Berlin: Walter de Gruyter.

Newmeyer, Frederick. 1998. *Language Form and Language Function*. Cambridge, MA: MIT Press.

Nida, Eugene A. 1946. *Morphology: The descriptive analysis of words*. Ann Arbor, MI: University of Michigan Press.

Norde, Muriel. 2009. *Degrammaticalization*. Oxford: OUP.

Orel, Vladimir. 1998. *Albanian Etymological Dictionary*. Leiden/Boston, MA/Cologne: Brill.

Otrębski, Jan. 1956–1965. *Gramatyka języka litewskiego* (*3 vols.*). Warsaw: Państwowe Wydawynictwo Naukowe.

Partee, Barbara. 1973. Some transformational extensions of Montague grammar. *Journal of Philosophical Logic*, 2(4): 509–534.

Petit, Daniel. 2009. La préhistoire des adjectifs déterminés du baltique et du slave. *Bulletin de la sociétée linguistique de Paris*, 104: 311–60.

Picard, Marc. 1997. On the Effects of Frequency-induced Phonological Change. *Diachronica*, 14: 109––118.

Pinkster, Harm. 2021. *The Oxford Latin Syntax. Vol. 2: The complex sentence and discourse*. Oxford: OUP.

Pires, Acrisio & Sarah G. Thomason. 2008. How much syntactic reconstruction is possible? In *Principles of Syntactic Reconstruction* (G. Ferraresi & M. Goldbach, eds.), Amsterdam: John Benjamins, pp. 27–72.

Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern/Munich: Francke Verlag.

Pompei, Anna. 2011a. Relative Clauses. In *New Perspectives on Latin Historical Syntax. Vol. 4: Complex sentences, grammaticalization, typology* (P. Baldi & P. Cuzzolin, eds.), Berlin: Mouton de Gruyter, pp. 427–547.

Pompei, Anna. 2011b. Relative clauses of the "third type" in Latin? In *Formal Linguistics and the Teaching of Latin: Theoretical and applied perspectives in comparative grammar* (R. Oniga, R. Iovino & G. Giusti, eds.), Newcastle: Cambridge Scholars Publishing, pp. 117–32.

Pompei, Anna. 2014. Phrasal verbs and preverbs in Homeric Greek between telicization, incorporation and valency change. In *The Greek Verb: Morphology, syntax and semantics* (A. Bartolotta, ed.), Louvain-La-Neuve/Walpole, MA: Peeters, pp. 253–276.

Probert, Philomen. 2003. *A New Short Guide to the Accentuation of Ancient Greek*. London: Bristol Classical Press.

Probert, Philomen. 2006. Clause Boundaries in Old Hittite Relative Sentences. *Transactions of the Philological Society*, 104(1): 17–83.

Probert, Philomen. 2008. Mycenaean o- is accusative; jo- is nominative. *Glotta*, 84: 126–68.

Probert, Philomen. 2015. *Early Greek Relative Clauses*. Oxford: OUP.

Probert, Philomen. 2018. Are correlative pronouns always overt in Lydian? In *Vina Diem Celebrent: Studies in Linguistics and Philology in honor of Brent Vine* (D. Gunkel, S. W. Jamison, A. O. Mercado & K. Yoshida, eds.), Ann Arbor, MI/New York, NY: Beech Stave Press, pp. 363–79.

Puhvel, Jaan. 1984. *Hittite Etymological Dictionary. Vol. 1.* The Hague: Mouton.

Puri, Vandana. 2011. The influence of English on the history of Hindi relative clauses. *Journal of Language Contact*, 4: 250–68.

Radanović-Kocić, Vesna. 1988. *The Grammar of Serbo-Croatian Clitics: A synchronic and diachronic perspective*. Ph.D. thesis, University of Illinois, Urbana, IL.

Radanović-Kocić, Vesna. 1996. The Placement of Serbo-Croatian Clitics: A prosodic approach. In *Approaching Second: Second position clitics and related phenomena* (A. L. Halpern & A. M. Zwicky, eds.), Stanford, CA: CSLI Publications, pp. 429–46.

Radford, Andrew. 1997. *Syntax: A Minimalist introduction*. Cambridge: CUP.

Radford, Andrew. 2004. *Minimalist Syntax: Exploring the structures of English*. Cambridge: CUP.

Ram-Prasad, Krishnan J. 2018. *The Syntax of "Local Particles" in Sanskrit*. Master's thesis, University of Cambridge.

Ram-Prasad, Krishnan J. 2020. Disentangling the Vedic Left Periphery. *Cambridge Indo-European* (*E Caucus*) *Seminar*, 28th Oct 2020.

Ram-Prasad, Krishnan J. & Ian Roberts. 2022. PCM and Syntactic Reconstruction: An Indo-European case study. *The Parametric Comparison Method Workshop*, Cambridge, 18th February 2022.

Ramat, Paolo. 1992. Thoughts on degrammaticalization. *Linguistics*, 30: 549–60.

Ramchand, Gillian & Peter Svenonius. 2014. Deriving the functional hierarchy. *Language Sciences*, 46: 152–74.

Rawlins, Kyle. 2008. (*Un*)*conditionals: An investigation in the syntax and semantics of conditional structures*. Ph.D. thesis, University of California at Santa Cruz.

Reece, Steve. 2009. *Homer's Winged Words*. Leiden: Brill.

Reinöhl, Uta. 2016. *Grammaticalization and the Rise of Configurationality in Indo-Aryan*. Oxford: OUP.

Ringe, Donald. 2017. *From Proto-Indo-European to Proto-Germanic*. Oxford: OUP, 2nd edition.

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar* (L. Haegeman, ed.), Dordrecht: Springer, pp. 281–337.

Rizzi, Luigi. 2001. On the position "Int(errogative)" in the left periphery of the clause. *Current studies in Italian syntax*, 14: 223–251.

Roberts, Ian. 2004. The C-system in Brythonic Celtic Languages, V2 and the EPP. In *The structure of CP and IP* (L. Rizzi, ed.), Oxford: OUP, pp. 297–328.

Roberts, Ian. 2007. *Diachronic Syntax*. Oxford: OUP, 1st edition.

Roberts, Ian. 2021. Second Positions: A synchronic analysis and some diachronic consequences. In *Continuity and Variation in Germanic and Romance* (S. Wolfe & C. Meklenborg, eds.), Oxford: OUP, pp. 297–328.

Roberts, Ian & Anna Roussou. 2003. *Syntactic Change: A Minimalist approach to grammaticalization*. Cambridge: CUP.

Rochemont, Michael S. & Peter W. Culicover. 1997. Deriving dependent right adjuncts in English. In *Rightward movement* (D. Beerman, D. LeBlanc & H. van Riemsdijk, eds.), Amsterdam: John Benjamins, pp. 279–300.

Roma, Elisa. 2007. Relativisation strategies in insular Celtic languages: History and contacts with English. In *Europe and the Mediterranean as Linguistic Areas: Convergencies from a historical and typological perspective* (P. Ramat & E. Roma, eds.), Amsterdam/Philadelphia, PA: John Benjamins, pp. 245–88.

Rosinas, Albertas. 1975. Ar baltų **i-*, **į̄o-*, resp. **ī*, **į̄ā-* kamienai buvo reliatyvinai? *Baltistica*, 11(2): 165–70.

Roussou, Anna. 1998. Wh-interrogatives: from Classical Greek to Modern Greek. *Selected papers on theoretical and applied linguistics*, 12(1): 109–26.

Ruijgh, Cornelis J. 1971. *Autour de "te épique": Études sur la syntaxe grecque*. Amsterdam: Hakkert.

Salvi, Giampaolo. 2005. Some firm points on Latin word order: The left periphery. In *Universal Grammar in the Reconstruction of Ancient Languages* (K. É. Kiss, ed.), Berlin: De Gruyter, pp. 429–56.

Salzmann, Martin. 2006. *Resumptive Prolepsis: A study in indirect A'-dependencies*. Utrecht: LOT.

Salzmann, Martin. 2017. *Reconstruction and Resumption in Indirect A'-Dependencies: On the syntax of prolepsis and relativization in (Swiss) German and beyond*. The Hague: Mouton De Gruyter.

Samuels, Bridget. 2005. On the left periphery of Anatolian. *Linguistic Analysis*, 35(1): 275–97.

Sauerland, Uli. 1998. *The Meaning of Chains*. Ph.D. thesis, MIT.

Sauerland, Uli. 2003. Unpronounced heads in relative clauses. In *The Interfaces: Deriving and Interpreting Omitted Structures* (K. Schwabe & S. Winkler, eds.), Amsterdam: John Benjamins, pp. 205–226.

Schachter, Paul. 1973. Focus and relativization. *Language*, 49(1): 19–46.

Schmitt-Brandt, Robert. 1973. Vergleich der indogermanischen Nebensatzkonstruktionen. In *Indogermanische und Allgemeine Sprachwissenschaft* (G. Redard, ed.), Wiesbaden: Reichert, pp. 125–141.

Senn, Alfred. 1966. *Handbuch der litauischen Sprache. Band I: Grammatik*. Heidelberg: Carl Winter Verlag.

Seppänen, Aimo. 1997. The genitives of the relative pronouns in present-day English. In *Taming the Vernacular: From dialect to written standard language* (J. Cheshire & D. Stein, eds.), London: Longman, pp. 152–69.

Sevdali, Christina. 2013. Ancient Greek Infinitives and Phases. *Syntax*, 16(4): 324–61.

Shield, Kenneth. 1980. Fast Speech and the Origin of the Standard English Verbal Suffix *-s*. *Journal of English Linguistics*, 14: 24–35.

Siddiqi, Daniel. 2010. Distributed Morphology. *Language and Linguistics Compass*, 4: 524–542.

Sideltsev, Andrei V. 2017. Accented Clitics in Hittite? *Transactions of the Philological Society*, 115(2): 176–211.

Sideltsev, Andrei V. 2018. The evolution of free-choice indefinites in Hittite. *Journal of Near Eastern Studies*, 77(2): 279–300.

Sideltsev, Andrei V. & Maria Molina. 2015. Enclitic -(*m*)*a* 'but', clause architecture and the prosody of focus in Hittite. *Indogermanische Forschungen*, (120): 209–54.

Sihler, Andrew L. 1995. *New Comparative Grammar of Latin and Greek*. Oxford: OUP.

Smyth, Herbert W. 1956. *A Greek grammar for colleges*. Cambridge, MA: Harvard University Press.

Spevak, Olga. 2006. Les enclitiques en latin. *Indogermanische Forschungen*, 111: 249–74.

Spevak, Olga. 2010. *Constituent Order in Classical Latin Prose*. Amsterdam: Benjamins.

Spevak, Olga. 2012. Les enclitiques *enim* et *autem* dans la diachronie du latin. In *Les Évolutions du latin* (A. Christol & O. Spevak, eds.), Paris: Collection Kubaba, pp. 339–56.

Srivastav, Veneeta. 1991. Relative clauses in Hindi and learnability. In *Papers on South Asian Linguistics* (V. Srivastav, J. W. Gair & K. Wali, eds.), pp. 133–160. Special edition of *Cornell University Working Papers in Linguistics*, Vol. 8.

Strachan, John. 1907. An Old-Irish Homily. *Ériu*, 3: 1–10.

Sturtevant, Edgard H. 1930. Relatives in Indo-European and Hittie. *Language Monograph*, 7: 141–149.

Subbarao, Karumuri. 1984. *Complementation in Hindi Syntax*. Delhi: Academic Publications.

Szabolsci, Anna. 2001. The syntax of scope. In *The Handbook of Contemporary Syntactic Theory* (M. Baltin & C. Collins, eds.), Oxford: Blackwell, pp. 607–34.

Szemerényi, Oswald. 1980. *Einführung in die vergleichende Sprachwissenschaft*. Darmstadt: Wissenschlaftliche Buchgesellschaft.

Szemerényi, Oswald. 1996. *Introduction to Indo-European Linguistics*. Oxford: OUP, 4th edition.

Takano, Yuji. 2003. How antisymmetric is syntax? *Linguistic Inquiry*, 34(3): 516–526.

Taylor, Ann. 1990. *Clitics and Configurationality in Ancient Greek*. Ph.D. thesis, University of Pennsylvania.

Thomas, Werner. 1966. Bemerkungen zum Gebrauch von toch. A *yo* und B. *wai*. *Central Asiatic Journal*, 11(4): 264–74.

Thompson, Rupert J. E. 2002. What the Butler saw: Some thoughts on the Mycenean o- ~ jo- particle. *Minos*, 37: 317–37.

Thurneysen, Rudolf. 1946. *A Grammar of Old Irish*. Dublin: Dublin Institute for Advanced Studies. Translated by D. A. Binchy and O. Berkin.

Timberlake, Alan. 1977. Reanalysis and actualization in syntactic change. In *Mechanisms of Syntactic Change* (C. N. Li, ed.), Austin, TX: Univeristy of Texas Press, pp. 141–177.

Vailliant, André. 1950–1977. *Grammaire comparée des langues slaves (5 vols.)*. Paris: Klinksieck.

van Emde Boas, Evert, Albert Rijksbaron, Luuk Huitink & Mathieu de Bakker. 2019. *Cambridge Grammar of Classical Greek*. Camrbridge: CUP.

van Orman Quine, Willard. 1962. *Identity, Ostension, and Hypostasis*. Ph.D. thesis, MIT, Cambridge, MA.

Vázquez-González, Juan G. & Jóhanna Barðdal. 2019. Reconstructing the Ditransitive Construction for Proto-Germanic: Gothic, Old English and Old Norse-Icelandic. *Folia Linguistica Historica*, 40(2): 555–620.

Vergnaud, Jean-Roger. 1974. *French Relative Clauses*. Ph.D. thesis, MIT.

Verma, Manindra K. 1966. *The Noun Phrase in Hindi and English*. Delhi: Motilal Banarsidass.

Wackernagel, Jacob. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen*, 1: 333–436.

Wali, Kashi. 1982. Marathi correlatives: A conspectus. In *South Asian Review: Studies in South Asian languages and linguistics* (P. J. Mistry, ed.), Jacksonville, FL: South Asian Literary Association, pp. 78–88.

Walkden, George. 2013. The correspondence problem in syntactic reconstruction. *Diachronica*, 30(1): 95–122.

Walkden, George. 2014. *Syntactic Reconstruction and Proto-Germanic*. Cambridge: CUP.

Watkins, Calvert. 1963. Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb. *Celtica*, 6: 1–49.

Watkins, Calvert. 1976. Toward Proto-European syntax: problems and pseudo-problems. In *Diachronic Syntax* (S. Steever, C. Walker & S. Mufwene, eds.), Chicago, IL: Chicago Linguistic Society, pp. 305–326.

Watkins, Calvert. 1991. Etymologies, equations and comparanda: Types and values, and criteria for judgment. In *Patterns of change, change of patterns: Linguistic change and reconstruction methodology* (P. Baldi, ed.), New York, NY: Mouton de Gruyter, pp. 167––82.

Watkins, Calvert. 1995. *How to Kill a Dragon: Aspects of Indo-European poetics*. Oxford: OUP.

Watkins, Calvert. 2010. Notes on the Hittite Funerary Ritual for a prince or princess. In *ipamati kistamati pari tumatimis: Luwian and Hittite Studies Presented to J. David Hawkings on the occasion of his 70th birthday* (I. Singer, ed.), Tel Aviv: Emery and Claire Yass Publications in Archaeology, Tel Aviv University, pp. 244–8.

Wegener, Phillip. 1874. Der lateinische Relativsatz. In *Programm des Bugenhagenschen Gymnasiums* (R. Bouterwek, ed.), Trzebiatów: Schnellpressendruck von Fr. Lehfeldt.

West, Martin L. 2011. *Old Avestan Syntax and Stylistics*. Berlin/Boston, MA: De Gruyter.

Wexler, Kenneth & M. Rita Manzini. 1987. Parameters and learnability in binding theory. In *Parameter Setting* (T. Roeper & E. Williams, eds.), Dordrecht: Reidel, pp. 41–76.

Williamson, Janice. 1987. An indefiniteness restriction for relative clauses in Lakhota. In *The Linguistic Representation of (In)definiteness* (E. Reuland & A. Meulen, eds.), Cambridg, MA: MIT Press, pp. 168–90.

Willis, David. 2011. Reconstructing last week's weather: Syntactic reconstruction and Brythonic free relatives. *Journal of Linguistics*, 47(2): 407–446.

Willis, David. 2013. Negation in the history of the Slavonic languages. In *The History of Negation in the Languages of Europe and the Mediterranean: Volume I Case Studies* (D. Willis, C. Lucas & A. Breitbarth, eds.), Oxford: OUP, pp. 341–98.

Willmott, Jo. 2013. Negation in the history of Greek. In *The History of Negation in the Languages of Europe and the Mediterranean: Volume I Case Studies* (D. Willis, C. Lucas & A. Breitbarth, eds.), Oxford: OUP, pp. 299–340.

Windhearn, Ryan. 2020. *Evidence from Innovation: Reconstructing disharmonic headedness for Proto-Indo-European*. Ph.D. thesis, Cornell University.

Windhearn, Ryan. 2021. Disharmonic Headedness in Homeric Greek and Tocharian and Implications for Proto-Indo-European Reconstruction. *Journal of Historical Syntax*, 4: 1–29.

Windisch, Ernst. 1869. Untersuchungen über den Ursprung des Relativpronomens in den indogermanischen Sprachen. *Curtius Studien*, 2: 201–419.

Winter, Werner. 1984. *Reconstructional comparative linguistics and the reconstruction of the syntax of undocumented stages in the development of languages and language families*. The Hague: Mouton.

Wodtko, Dagmar S. 2000. Wörterbuch der keltiberischen Inschriften. In *Monumenta Linguarum Hispanicarum* (J. Untermann, ed.), Wiesbaden: Reichert, volume V.1.

Yates, Anthony. 2014. Luwian 'indeterminate' relative clauses and their prehistory. *Harvard GSAS Workshop*, 12th November 2014.

Ziegler, Sabine. 1993. Zur Entwicklung der Relativsätze mit dem Relativpronomen *io- in den keltischen Sprachen. *Münchener Studien zur Sprachwissenschaft*, 54: 251–70.

Zinkečius, Zigmas. 1957. *Lietuvių kalbos įyardžiuotinių būdvardžių istorijos bruožai*. Vilnius: Valstybinė politinės ir mokslinės literatūros leidykla.

Zwicky, Arnold M. 1977. *On Clitics*. Bloomington, IN: Indiana University Linguistics Club.