



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

RECONSTRUCTING THE LEFT
PERIPHERIES OF
PROTO-INDO-EUROPEAN

DANNY L. BATE

THESIS SUBMITTED TO THE UNIVERSITY OF EDINBURGH IN
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
PHD IN LINGUISTICS AND ENGLISH LANGUAGE

SCHOOL OF PHILOSOPHY, PSYCHOLOGY AND LANGUAGE SCIENCES
COLLEGE OF HUMANITIES AND SOCIAL SCIENCE
UNIVERSITY OF EDINBURGH

SUPERVISORS: CAROLINE HEYCOCK, RONNIE CANN, ROBERT TRUSWELL



©DANNY L. BATE, EDINBURGH, UK, 2024

ABSTRACT

This thesis is an exercise in syntactic reconstruction, the proposal of the word-order patterns in a prehistoric language. The language in question is Proto-Indo-European, the hypothetical ancestor of the Indo-European language family. More specifically, this thesis reconstructs the abstract syntax behind the ‘left periphery’ (i.e. the beginning) of its clauses and noun phrases, pursued according to the theoretical framework of generative grammar. It is therefore a case study in the Minimalist approach to syntactic reconstruction of Walkden (2014). It begins by introducing the idea of linguistic reconstruction and the methods, chiefly the comparative method, by which it is undertaken. It then works through the debate over the feasibility of reconstructing not only sounds and vocabulary, but also word order; Walkden’s key principle is that syntactic heads and categories offer us the *comparanda* that can function as counterparts to the phonemes and lexical items used in traditional reconstruction. The thesis then introduces the representative set of seven historical Indo-European languages that are to be individually analysed and then compared, in order to propose what features of their syntax can be reconstructed back to the proto-language. These languages are: Latin, Ancient Greek, Vedic Sanskrit, Old Church Slavonic, Old English, Old Norse and Old Irish. The next chapter sets out the case for the functional category of ‘C’ in the syntax of Proto-Indo-European, and its primary function as the syntactic locus of clause type and super-/subordinate status. The following chapter expands on that underlying syntax by adding a ‘Topic’ component to the left periphery, as well as suggesting a lower position responsible for clausal focus. The thesis then turns to the noun phrase; on the basis of the language set, this is reconstructed as having a similarly structured syntax. It comprises the categories of ‘D’, the locus of (in)definiteness, and ‘Emphasis’, which is responsible for the fronting of constituent elements within the noun phrase. The final chapter summarises the reconstructions and offers remarks on their similarities, strengths and weaknesses, as well as avenues for further research.

LAY SUMMARY

Our languages have been around much longer than the historical record that we have for them; thousands of years of language are lost to us in the present day. However, linguistic prehistory is not entirely dark. Linguistic reconstruction shines a light into stages of language that predate the written record. Perhaps the most famous product of this endeavour is Proto-Indo-European, a hypothetical origin thought up to account for the innumerable similarities between a vast spread of languages across Eurasia, including English, French, German, Russian, Greek, Persian and Hindi-Urdu. With its origins in the 17th century, this idea of a lost ancestral language is widely accepted today, since it remains the best explanation for the systematic correspondences in sounds, vocabulary and grammar shared by the aforementioned Indo-European languages. The reconstruction is so fine-grained that we can further speculate about the location and society of its speakers; our reconstructed lexicon tells us a lot about their ties of kinship, their use of technology, and their interactions with the natural world.

But what about word order? We have so many of the words of Proto-Indo-European, but how did its speakers string those words together into sentences? This is the driving issue of this thesis: to discuss and pursue the reconstruction of the syntax of Proto-Indo-European. Specifically, it compares a set of documented Indo-European languages (including Latin, Sanskrit and Old English) to reconstruct the word-order patterns at the beginning of clauses and noun phrases. For example, when we form questions in English, we tend to put the question word (e.g. *who*, *what*) first in its clause (e.g. *what do you see?*). Did Proto-Indo-European do likewise? When we build long phrases around our nouns in English (e.g. *the five long chapters*), we tend to put functional words like *a* and *the* in first position. Was the prehistoric ancestor of English similar in this regard?

This doctoral thesis is a deep dive into the syntax of one particular prehistoric language. Its ultimate goal is the reconstruction of abstract syntactic structures for its clauses and noun phrases. To achieve that goal, it begins with a discussion of the feasibility of reconstructing word order, and how the concepts of current syntactic theory give us a good way to proceed. It then works through the historical evidence, analysing and comparing the set of representative Indo-European languages. It concludes with the ultimate reconstructions, as well as suggestions for how to take the analysis of the thesis further.

DECLARATION OF ORIGINALITY

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

ACKNOWLEDGEMENTS

While this thesis is a product of an extraordinary time for humanity, a period of great personal isolation, it certainly did not emerge from a vacuum of support. This document before you would not exist, were it not for the University of Edinburgh, whose wonderful linguists took a chance on a strange, vaguely syntactic research proposal four years ago. I will forever be profoundly grateful to the department and the university for making my doctoral journey possible, allowing me to live my best linguistic life for three and a half happy years.

Among those linguists, I must specifically thank my three supervisors, Caroline, Ronnie and Rob. I must, but I lack the words to do sufficiently. It has been a pleasure and a privilege to bask in their brilliance on a bi-monthly basis, and to receive so much from them in expertise, academic perspective and friendship. Likewise, I must thank Bettelou for her support and input during the project. Without these linguists' collective guidance, this thesis and my whole PhD experience would be all the poorer. While I cannot ever repay them, I can try to pass on what they have gifted me.

I also want to acknowledge the contribution of countless unknown scholars. Any thesis exists in a landscape of previous literature, and while this thesis recognises some of it (in twenty-eight pages of references), this is only the tip of the academic iceberg atop which this work stands. Without the hard work of librarians and lexicographers, translators and transcribers, archivists and archaeologists, this project would not have been possible.

More personally and concretely, I must thank my dear friends, both in Scotland and the Czech Republic, for their company and support over these past few years. In Edinburgh, I am grateful to so many people, such as to Tom, for being my host and guide to the place, and to Gilly, Sarah, Lisa and Vilde, for adopting me as one of the girls. In my Prague life, I must thank Rhianydd, Bara, Matouš, Anna, Albert, Líba and Petra for various different reasons, not least their friendship. My sincere thanks go especially to Krishnan, my companion, confidant and California collaborator. From Merton Street to Hollywood Boulevard, it has been wonderful to explore Indo-European syntax with such a brilliant scholar at the beginning of a brilliant career.

More personally still, I will always be immeasurably grateful to my parents, not only for the whole life thing, but more recently for all the support, encouragement and belief that have got me to this point. Indeed, their interests were the foundation for my own, and so underlie this whole thesis; the syntactic aspect may not have been their doing, but the historical side of it absolutely was. Sharing their passion for the past, through history books, archaeological digs and day trips to castles, cathedrals and country houses, was a great gift, and it put me on the path to this PhD.

Then there is my fiancée, Karolína. Quite what I did to deserve my relationship with such an unfailingly kind and generous person remains unclear. She has been my selfless rock of stability and strength as the world spins, and a source of joy and comfort while we have been together or apart. Along with our dear grumpy cat, she has been my team, cheering me on one word at a time, from day one to the finish line. She has encouraged me to do this work (or often to take a break from it), listened to me both wax lyrical and complain about it all, and kept me going in sickness and in health. This thesis is dedicated, with my whole heart, to her.

*Fáelid-sem cu ndéne dul
hi·nglen luch inna géchrub
hi·tucu cheist ndoraid ndil
os mé chene am fáelid*

...

*Hé fesin as choimsid dáu
in muid du·ngní cach óenláu
du thabairt doraid du glé
for mu mud céin am messe*

CONTENTS

ABSTRACT	ii
LAY SUMMARY	iii
DECLARATION	iv
ACKNOWLEDGEMENTS	v
1 INTRODUCTION	1
1.1 General Introduction to the Thesis	1
1.2 The Reconstruction of Prehistoric Languages	4
1.2.1 An Introduction to Reconstruction	4
1.2.2 The Debate Over Syntactic Reconstruction	9
1.2.3 Previous Approaches to Syntactic Reconstruction	15
1.2.4 Generative Approaches to Syntactic Reconstruction	25
1.3 Theoretical Assumptions and Method	31
1.3.1 Theoretical Assumptions: a Primer for Current Generative Syntax .	31
1.3.2 Method for the Thesis	38
1.4 The Indo-European Language Set	42
1.4.1 Latin	43
1.4.2 Ancient Greek	44
1.4.3 Vedic Sanskrit	45
1.4.4 Old Church Slavonic	46
1.4.5 Old English and Old Norse	48
1.4.6 Old Irish	50
1.4.7 The Exclusion of Hittite and Its Consequences	51

2	C IN PROTO-INDO-EUROPEAN	53
2.1	Introductions and Chapter Aims	53
2.2	Lexical Elements of the Syntactic Category C	56
2.2.1	Subordinating Elements in Early Indo-European	56
2.2.2	Interrogative Elements in Early Indo-European	104
2.2.3	Declarative Elements in Early Indo-European	116
2.2.4	Summary: the Lexical Evidence for C	121
2.3	Syntactic Movement to CP	122
2.3.1	<i>Wh</i> -movement in Early Indo-European	123
2.3.2	Movement to CP in Imperative Clauses	147
2.3.3	Movement to CP in Declarative Main Clauses	158
2.3.4	Movement to CP in Subordinate Clauses	159
2.3.5	Pronominal Movement to C: Wackernagel's law Reconsidered	160
2.4	Conclusions: the Case for PIE C	186
3	BEYOND C: TOPIC, FOCUS AND DISCOURSE-CONFIGURATIONALITY IN PROTO-INDO-EUROPEAN	190
3.1	Introductions and Chapter Aims	190
3.1.1	Chapter Aims	190
3.1.2	Introducing Indo-European Discourse Configurationality	192
3.2	Topic in Early Indo-European and PIE	198
3.2.1	Topics in Latin, Ancient Greek and Vedic Sanskrit	198
3.2.2	Topics and Early Germanic Word Order	228
3.2.3	Topics in Old Church Slavonic and Old Irish	249
3.2.4	Reconstructing Topic for Proto-Indo-European	255
3.3	Focus in Early Indo-European and PIE	261
3.3.1	Focus and its Functions	261
3.3.2	Low Focus in Proto-Indo-European	264
3.3.3	Against a High Focus for Proto-Indo-European	272
3.4	Conclusions: Topics, Foci and Early Indo-European	290
4	THE LEFT PERIPHERY OF THE NOUN PHRASE IN PROTO-INDO-EUROPEAN	294
4.1	Introductions and Chapter Aims	294
4.1.1	Chapter Aims	294
4.1.2	Theoretical Background: The Syntax of Noun Phrases	296
4.2	Ancient Noun Phrases: Latin, Sanskrit and Ancient Greek	301

4.2.1	The Latin Noun Phrase	301
4.2.2	The Vedic Sanskrit Noun Phrase	324
4.2.3	The Ancient Greek Noun Phrase	343
4.2.4	The Ancient Noun Phrase: Intermediate Summary	366
4.3	The Early Germanic Noun Phrase	368
4.3.1	Evidence for the Early Germanic DP: Word Order	368
4.3.2	Evidence for the Early Germanic DP: Morphology	383
4.3.3	Evidence for an Early Germanic EmphasisP: Noun Fronting and Possessor-Article Orders	392
4.3.4	The Early Germanic Noun Phrase: Intermediate Summary	399
4.4	Slavic and Celtic Noun Phrases	402
4.4.1	The Old Church Slavonic Noun Phrase	402
4.4.2	The Old Irish Noun Phrase	414
4.5	Conclusions: Reconstructing a Proto-Noun Phrase	424
5	THESIS CONCLUSIONS	435
5.1	The Reconstructed Left Peripheries	435
5.2	Discussion of the Reconstructions and Wider Context	439
5.3	Further Work	444
	APPENDICES	448
A	APPENDICES	449
A.1	Abbreviations	450
A.2	Sources Information	451
A.2.1	Latin	451
A.2.2	Ancient Greek	451
A.2.3	Sanskrit	451
A.2.4	Old Church Slavonic	452
A.2.5	Old English	452
A.2.6	Old Norse	452
A.2.7	Old Irish	452
	REFERENCES	483

INTRODUCTION

1.1 GENERAL INTRODUCTION TO THE THESIS

What can we believe about the word order of a prehistoric language? This is the driving question of this thesis, which is an extensive exercise in syntactic reconstruction. That is to say, it aims to make convincing claims about the word-order patterns of one particular hypothetical ancient language: Proto-Indo-European. This language, arguably the most famous ‘proto’-language¹, is hypothetical, since it was spoken in prehistory and left no written record behind. Our evidence for it comes instead from the countless similarities among the large group of its documented linguistic descendants, known collectively as the

¹From Ancient Greek *prôtos* ‘first’.

1.1. GENERAL INTRODUCTION TO THE THESIS

Indo-European language family. On this basis, it is today widely accepted that Proto-Indo-European existed, or rather that there was once a language that closely resembled what we have reconstructed.

The language and its reconstruction are by no means new ideas; the genealogy of the concept can be traced as far back as the 17th century and to linguists like Marcus Zuerius van Boxhorn (1612-1653), who proposed an ancestral language as an explanation for the similarities shared by languages like Latin, Greek and German. The study of this common origin accelerated with the work of Sir William Jones (1746-1794) and his famous pronouncement on the comparison of Latin, Ancient Greek and Sanskrit, namely how they share

“... a stronger affinity, both in the roots of verbs and the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists.”

(Jones, 1798)

In these early insights are two axioms of linguistic reconstruction: that human language predates the historical record, and that we can learn about prehistoric languages by comparing their descendants for which we do have historical evidence. There are no direct witnesses to Proto-Indo-European, but because Latin, Ancient Greek, Sanskrit and others are documented, their common parent is not entirely lost to us today.

1.1. GENERAL INTRODUCTION TO THE THESIS

Since its beginnings, the reconstruction of Proto-Indo-European has concentrated heavily on the sounds and vocabulary of the language; as §1.2 recounts, this focus is entirely reasonable. Yet there is more to any language than these two components, and this thesis is one among recent attempts² at reconstructing aspects of how its vocabulary was assembled into longer sequences – that is, its syntax. This new fervour for syntactic reconstruction has been given inspiration and direction by the methodology of Walkden (2013, 2014), who proposes that integrating current concepts of generative grammar places our syntactic reconstructions on good foundations. This thesis follows that method, and has as its goal the reconstruction of the ‘left peripheries’ of two syntactic domains of Proto-Indo-European word order.³ These domains are the clause and the noun phrase. Integral to this task is the syntactic analysis of a set of representative historical Indo-European languages (Latin, Ancient Greek, Vedic Sanskrit, Old Church Slavonic, Old English, Old Norse and Old Irish), which were selected at the outset of this doctoral project on the basis of their antiquity, genealogical distribution and accessibility. These seven therefore provide the necessary substance for reconstruction.⁴

The thesis structure is as follows. §1.2 first discusses linguistic reconstruction in detail. §1.3 then sets out the theoretical background and method for the thesis, while §1.4 introduces the representative languages with philological biographies. Following this substantial

²Namely, Barðdal et al. (2013, 2020), Windhearn (2020) and Ram-Prasad (2022).

³‘Left’ is of course metaphorical language, based on the left-to-right direction of European writing systems; what it really means is the beginning of the two domains.

⁴The vast majority of the examples given to substantiate the claims of thesis are real, taken from historical documents written in the seven languages of the set. Each example is transliterated into Latin letters, and comes with a gloss, a translation (the author’s own), a language label and a source reference.

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

introductory chapter, the substance of the thesis begins by reviewing and reconstructing two syntactic ingredients of the clausal left periphery; each gets its own chapter, with Chapter 2 for C and Chapter 3 for Topic. The case for the noun phrase cannot be so neatly divided up, so it is discussed in its entirety in Chapter 4. Chapter 5 reviews and compares the two reconstructions, and concludes with general points and directions for further work. An appendix to the thesis provides information about the linguistic abbreviations and resources used for the historical examples.

1.2 THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

1.2.1 AN INTRODUCTION TO RECONSTRUCTION

Linguistic reconstruction shines a light into the darkness of prehistoric language. It takes as axiomatic the assumption that all the language that we have documentation for, both modern and ancient, is by no means the full extent of humanity's linguistic output, but rather follows on from millennia of preceding language that is undocumented and therefore lost to us.

Nearly lost, that is. Through similarities across languages, we may theorise something of the ancestral stages of which the languages are descendants. When two or more languages show similarities so frequently and systematically that they cannot have arisen through contact or mere coincidence, the only reasonable conclusion becomes that the similarities and the languages themselves share a common source. Sometimes that common source is

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

known to us; the Romance languages of today developed out of Latin, for which we have an abundant record. In other cases, the source was long since defunct before the historical era. In these cases, the ancestor must remain hypothetical, and our reconstructions of it cannot be externally verified, except in rare instances of newly discovered documents. They are instead evaluated according to how well they account for all the descendant data that they are intended to explain.

Much of the enterprise of linguistic reconstruction has been driven by the study of one particular language family: Indo-European. Its ultimate target has been the prehistoric language of *Proto-Indo-European*. Within that endeavour, phonological and lexical reconstruction have led the way. This is to say, the **comparative method**, by which reconstruction is usually undertaken, has been fueled by the regular correspondences in sounds and meanings between documented words across the Indo-European family. For instance, the clear phonological and semantic similarities between Latin *pater*, Ancient Greek *patér*, Sanskrit *pitá*, Old English *fæder* and Old Irish *athair* (all meaning ‘father’) lead us first to posit an unattested ancestor from which all five derive, and second to reconstruct what sounds and meaning that ancestor must have had to produce the five reflexes in this ‘correspondence set’.

Individual lexical reconstructions, here Proto-Indo-European **ph₂tér*, rely on the concurrent reconstruction of proto-phonemes from which the vocabulary was built. Phonemic reconstruction in turn relies on the Neogrammarian hypothesis (Osthoff and Brugmann, 1878) that sound change is *regular*; a sound will change in every instance of a particular phonological environment. For instance, the Proto-Indo-European voiced aspirate sound

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

*ǵ^h became /h/ in Latin whenever it preceded a vowel in word-initial position.⁵ Any exceptions to this regularity⁶ are treated as cases of an alternative environment, analogy with another word, or a borrowing from another language. Armed with regular sound changes, we can apply them to all words that contain the particular sound and undo the change back to the form of the proto-word, even in the absence of a full correspondence set to confirm that reconstruction.

Yet recognising sound changes is only possible through first identifying and comparing words in which the changed sounds appear, since the word provides the necessary corresponding phonological environments. Both the higher-level unit (words and bound morphemes) and the lower-level unit (their constituent sounds) can share ‘diachronic identity’ with cognate reflexes across the different languages compared, and the cognacy of one level both corroborates and is corroborated by the cognacy of the other. This interdependence Walkden (2013, 2014) names the Double Cognacy Condition:

“In order to form a correspondence set, the contexts in which postulated cognate sounds occur must themselves be cognate.”

(Walkden, 2013, 101)

The condition is a key factor in the application of the comparative method to syntactic reconstruction, discussed in §1.2.4.

In reconstructing sounds and the phonological form of words, the comparative method

⁵E.g. *ǵ^héiōm ‘winter’ > *hiems*.

⁶E.g. *ǵ^hh₂éns ‘goose’ > *ānser*, not **hānser*.

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

involves further assumptions about possible, reasonable and likely instances of change, based both on the features of the descendant words, and on typological observations of languages more widely. The ancestor of *pater*, *patér*, *pitá*, *fæder* and *athair* is believed to have had an initial *p in part simply because of its frequency in the correspondence set. This principle is the ‘majority rule’ (Roberts, 2007, 500-1). However, this rule is not enough. Among the Indo-European ‘brother’-cognates (Latin *frāter*, Ancient Greek *phrátēr*, Sanskrit *bhrātā*, Old English *brōþor*, Old Irish *bráthair*, etc.), only Sanskrit *bhrātā* maintains the voiced aspirate *b^h with which their common etymon (**b^hréh₂tēr*) is widely thought to have begun. This sound is therefore reconstructed as a phonological common denominator of the features of its reflexes. This goes against the majority rule, but it does satisfy a ‘directionality rule’, since the shifts of a voiced labial aspirated stop *b^h into the labial fricative /f/, the labial aspirated stop /p^h/ and the voiced labial stop /b/ are phonologically natural and conceivable changes. Likewise, in the case of the ‘father’-set, initial *p is also reconstructed because the lenition of an original *p to an innovative /f/ in Old English and Ø in Old Irish is a more plausible path of progression than the fortition of an original /f/ into /p/ in each of the three older languages. This sense of natural-ness in sound change is not a given though; it is drawn instead from the observation in documented languages of the greater occurrence of lenition over its inverse.

Age offers another principle; the two aforementioned lexical reconstructions are further supported instead by the age of the documented reflexes (the ‘elder rule’). In terms of antiquity, Vedic Sanskrit trumps the other languages in the two correspondence sets, which have had more time to undergo changes away from their common source. The fact that the

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

Vedic vocabulary displays /p/ and /b^h/ therefore additional lends weight to their putative presence among the phonemes of Proto-Indo-European. All in all, linguistic reconstruction is not a straightforward process, but rather a cumulative effort that draws from myriad sources and works with certain guiding principles, to propose a proto-form that explains all the data most cogently.⁷

This traditional lexico-phonological reconstruction has furthermore worked with functional elements that come under the traditional umbrella of ‘grammar’. This for example includes case endings, which, like lexical roots, have phonological form and meaning. Within Indo-European, the appearance and distribution of Latin *-s*, Ancient Greek *-s*, Sanskrit *-h*, Gothic *-s* and Hittite *-s* allow us to group them together as another correspondence set, and from this to reconstruct an ending **-s* that expressed nominative case and singular number for animate-referent nouns. Likewise, functional lexemes like interrogative pronouns, personal pronouns and adpositional particles can be securely reconstructed.

However, what is the feasibility of reconstructing other aspects of grammar? What about the word order of Proto-Indo-European? We know much about the sounds of this prehistoric language, and how those sounds were assembled into words, but can we know how its speakers then assembled those words into sentences?

⁷Some of the conclusions that result from the comparative method are subsequently considered typologically strange, such as the near-absence of the non-aspirated **b* in Proto-Indo-European (Olander, 2022b).

1.2.2 THE DEBATE OVER SYNTACTIC RECONSTRUCTION

In brief, there has been an attitude of pessimism around the idea of reconstructing syntax; Gildea et al. (2020) consider the idea to have been “regarded as bootless, frowned upon and even lambasted”. A key difference between syntactic and classical lexico-phonological reconstruction is that while the patterns and rules of syntax manipulate lexical items, they do not share their ‘visibility’ in sound (and therefore in historical texts). Consequently, since reconstruction works primarily through comparison and cognacy between languages, it is unclear what in syntax we can first compare to pursue reconstruction.

Moreover, phonemes are identifiable, concrete components of a spoken language, forming a limited set and passed on from one generation of speakers to the next – or more accurately, acquired and formulated anew by each generation to match their phonetic input. Sounds thus give the linguist a finite set of discrete elements to compare and trace back in a chain of acquisition to prehistoric origins. Out of these phonological atoms, a finite lexicon of elements with phonological form and meaning can then be reconstructed. With sentences and word order, however, it is not so clear what is being acquired through the generations between Proto-Indo-European and its descendant languages.

“Sentences are formed, not learned; morphemes and simple lexemes are learned, not formed ... Syntax deals almost exclusively with entities not learned, but constructed.”

(Winter, 1984, 622-3)

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

This is to say, while we may learn phonemes, morphemes and words from previous generations, sequences made from those words are for the most part not simply learned and parroted, but rather are generated afresh.⁸ We do not inherit sentences, so sentences cannot be cognate. Yet something syntactic must be passed on; a child's arrangements of given elements will typically match the arrangement in their parents' language. Word-order preferences and rules must be abstracted by the child from their linguistic input, just as phonemes are. However, the rules remain abstract and are explicitly formulated only by the teacher or the syntactician.

Identifying and comparing these rules are serious hurdles for syntactic reconstruction. Within the long arrangements of lexical elements in our early sources for the Indo-European languages, we lack the obvious cognate *comparanda* that we can say share diachronic identity and a prehistoric origin. While Old English /f/ and Latin /p/ can be said to be corresponding cognate sounds, acquired down the generations, the typical Subject-Verb order of Old English and Latin sentences cannot be so confidently claimed to have the same diachronic identity through time; it could be coincidence. Without clear cognates, we therefore lack the corresponding elements in that chain of acquisition that can lead us back in time from Latin, Ancient Greek, Sanskrit and other languages, to the syntax of Proto-Indo-European. This, named the "correspondence problem" (Lightfoot, 2002), poses a grave challenge to syntactic reconstruction, since "the first law of comparative grammar is that you've got to know what to compare" (Watkins, 1976, 312).

⁸With the exception, of course, of idioms and formulaic expressions.

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

“One may naturally wonder if Latin, Greek, Sanskrit, and English ‘*passive*’ and ‘*predication*’ do instantiate the ‘same’ entity in the sense and to the extent that we think Lat. *iugum*, ζυγόν, Skt. *yugám*, ModE *yoke* do.”

(Longobardi, 2008, xii)

There are different factors too behind phonological change and syntactic change. Phonology mostly changes for intra-phonetic reasons, such as assimilation between adjacent sounds. Regular phonological change is “blind to all considerations of morphological, syntactic, semantic and discourse categories” (Pires and Thomason, 2008, 68), and this is what makes it so useful for reconstruction, since it means that the various changes away from a phonological proto-form tend to be predictable and applicable across the board. Syntactic change is not so confined in its causes; non-syntactic cognitive factors (such as morphological and analogy) must be considered too. We seem to lack the same regularity and predictability in syntactic change as that which we have for the phonological.

Lightfoot (1979, 2002) represents a pessimistic extreme in the matter of syntactic reconstruction. For him, the damning problems are threefold. The first is the “correspondence problem” with identifying *comparanda* in syntax and then using them to reconstruct a proto-form. His second objection concerns directionality in syntactic change. Unlike in phonology, we lack the means to say whether a syntactic change is more likely to occur in one direction than in another, and therefore to say which of the *comparanda* in question are conservative or innovative with respect to the proto-language. For example, if three out of four related languages have a default Adjective-Noun order, but the fourth is Noun-

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

Adjective, how are we to tell which maintains the syntax of their common ancestor? Unlike with Old English /f/ and Latin /p/, we cannot appeal to some general syntactic features or precedent changes to determine which order is more likely to be the original.

Lightfoot's related third problem is that of the matter of chaotic change in the transmission of syntax (Lightfoot, 2002). This is to say, at the point of acquisition in infancy, there may be a radical reanalysis of the syntax that only accidentally generates output that matches the input. Being abstract schemata, the syntax of one generation does not come into contact with that of the older generation; this leaves the intergenerational transmission of syntax open to instances of seemingly chaotic change. In Lightfoot's view (2002, 133), this chance of chaos deprives us of the means to "predict backwards" and thereby reconstruct earlier links in the chain of acquisition.

Such methodological complications have contributed to the air of pessimism around syntactic reconstruction. For Pires and Thomason (2008), the debate over the feasibility of the endeavour remains "unsettled". However, Lightfoot's objections have not gone unanswered. Against the second objection, Walkden (2013) argues that both lexico-phonological and syntactic reconstruction draw on additional criteria to supplement our theories of change, such as frequency, probability, typological data and Occam's razor. Directionality in language change is not the only guide to a cogent reconstruction, and even still it is only a matter of likelihood based on observation, rather than of necessary directions of sound change. The lenition may be more common and therefore seem more likely in the history of Indo-European, but its inverse, fortition, is nonetheless attested. Just as it may prove difficult to propose which out of a Noun-Adjective and an Adjective-Noun order is

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

the original, there may be similar cases in lexico-phonological reconstruction in which no directionality across a correspondence set of cognates is clear.

Moreover, words develop along pathways of grammaticalisation and certain word-order arrangements do show ‘syntacticisation’ over time. Walkden (2014, 91) notes that Germanic languages underwent a shift from an “information-structurally conditioned pattern” to one determined primarily by syntactic requirements. Our earliest sources for North and West Germanic attest to a stage between the two extremes. Old Norse, for example, displays an alternation between verb-second and verb-first clauses conditioned by topicality (see §3.2.2), but North Germanic since then has given way to a generalised verb-second order. Here therefore we have an example of a direction in word-order change. On this basis we may posit that prehistoric Germanic word order, prior to those first sources, was closer to the former, discourse-configurational pattern.

Against the third problem, Willis (2011) and Ram-Prasad (2022) have refuted in detail Lightfoot’s objection on the grounds of chaotic change. In short, while the potential for drastic reanalysis remains, in reality the internal grammars of infant and parent cannot be so different if they are to generate matching output, as they do so consistently. Syntax is remarkably stable between the generations, because it is constrained by the linguistic data first received in acquisition. Syntactic variation can only flourish at the micro-level; any more radical divergences will produce noticeably different output, which will in turn be corrected. Intergenerational change not only in syntax but in phonology too arises from inconspicuous discrepancies in the abstract schemata, which otherwise successfully hit on the same features as the parent schemata, and so generate the same sounds or word order.

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

Really radical reanalysis between the generations is rare.

These responses to the second two of Lightfoot's objections not only rescue syntactic reconstruction, but even demonstrate the commonalities and difficulties that it shares with the traditional reconstruction of lexemes and phonemes.⁹

“In considering syntax from the historical point of view, it would be absurd to think that the physical body of a sentence or similar utterance can be historically transmitted. The same is valid for phonology or morphology; it is the phonological and morphological structure of the language which constitutes a set of linguistic systems which undergo historical transmission, not any given physical realization. The underlying syntactic structure of a sentence, for example the rules of arrangement of its constituent elements, can be presumed in the same fashion to form a linguistic system or set of systems functioning in time, and historically transmittable. As such, it is susceptible to analysis by the comparative method, as well as by other techniques of historical linguistics.”

(Watkins, 1963, 2)

However, Lightfoot's correspondence problem remains a serious challenge. Opinions have varied as to the efficacy of the comparative method with regards to syntax. Some approaches try to use it; others avoid it and look for other means of making cogent proposals of prehistoric syntax. Those that use the method have to answer the questions of what

⁹A fair charge laid against syntactic reconstructions is that the syntactic 'units' to be compared are unclear and abstract; yet the same charge of abstraction surely applies to phonemes too (cf. Jeffers, 1976, 4)

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

exactly we can compare and of what in a language's syntax can be said to be inherited through the generations. The matter of reconstructing syntax therefore depends greatly on what we think syntax actually is.

1.2.3 PREVIOUS APPROACHES TO SYNTACTIC RECONSTRUCTION

To see to what extent previous attempts have succeeded in navigating these issues of theory and methodology, we may start with the Neogrammarians of the 19th century, such as Delbrück (1878, 1893) and Wackernagel (1892). For Delbrück, the Proto-Indo-European declarative clause must have had a basic word order of Subject-Object-Verb, on the basis of the frequency and apparent basic-ness of the SOV order in early Indo-European languages, chiefly in Sanskrit. The Sanskrit clause displays an SOV order with such consistency that SOV can be considered a default order. Reconstructing SOV for Proto-Indo-European is therefore supported by the majority and elder rules, and also by the comparative evidence of Latin, Ancient Greek and Avestan. For Delbrück, it is this clausal arrangement, maintained across the emerging family, that has been passed on since the *ur*-grammar. This view is strengthened by the later-deciphered evidence from Hittite, which is also consistently SOV (Sideltsev, 2015a).

However, SOV is by no means rigid across early Indo-European. Sanskrit, Ancient Greek and Latin display an impressive capacity for word-order variation. In principle and practice, all sequences of subject, object and verb are possible. This variety is not random, but rather appears to operate according to the requirements of discourse; that is

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

to say, word order in these languages is used to relate the propositional content of a clause to its external context. Early Indo-European languages are “discourse-configurational” (Kiss, 1995a). Even the seemingly ‘basic’ SOV order is a product of this; Lowe (2015, 38) argues that SOV in Sanskrit is the most common clausal arrangement “given the typical information-structural arrangement of a clause”, since subjects are frequently topical and objects are “likely to be more discourse-prominent than the verb”. In other words, SOV may not in fact reflect a basic Sanskrit order, but rather be equal in status with the other orders that arise from the discourse-informational arrangement of the early Indo-European clause.

Applied to early Indo-European in general, this view of word order is a blow to the possibility of simply reconstructing SOV as the fundamental word order of Proto-Indo-European. It deprives SOV of any special status, explaining its common occurrence not as the product of a syntactic configuration, but as a typical assignment of clausal constituents to the available discourse-informational positions. It is these stable abstract positions that are instead what early Indo-European languages have in common, and what speakers of Latin, Ancient Greek, Sanskrit, etc. have inherited. Instead of the basic syntax of the verb and its arguments, these are instead better targets of syntactic reconstruction, as in Chapter 3.

A separate, long-recognised word-order phenomenon is that of Wackernagel’s law (1892). It is not strictly a law, but rather an observation that certain unstressed elements in the early Indo-European clause tend to appear second in the overall order, as if hosted by whatever comes first. Typical Wackernagelian elements include coordinating conjunctions,

1.2. *THE RECONSTRUCTION OF PREHISTORIC LANGUAGES*

auxiliary verbs and object pronouns, and this pattern appears across Indo-European (in Sanskrit, Ancient Greek, Latin, Hittite, Old Irish and beyond). It is their wide distribution that leads us to propose this ‘law’ for the proto-language too; for Watkins (1964, 1036), the behaviour of enclitics is “one of the few generally accepted syntactic statements” about Proto-Indo-European. However, the type and position of Wackernagelian elements are not rigid across early Indo-European languages, so §2.3.5 of this thesis takes a closer look at the data and conditions for Wackernagel’s law, to propose what exactly about it can be reconstructed back to the proto-language.

A problem that pertains to both Delbrück and Wackernagel, raised by Jeffers (1976), is that the Neo-Grammarians struggle to deal with difference. It relies on the appearance of the same feature (be it SOV order or second-position elements) across a number of related languages; if several early Indo-European languages have it, the common ancestor must have had it too. This follows the majority rule. Yet the success of traditional reconstruction is in how it incorporates both identity and systematic difference. The goal of correspondence sets and establishing cognates is really to reconstruct on the basis of how the elements differ. What would Delbrück and Wackernagel do with elements of word order in which there is no clear cross-Indo-European commonality? The relative ordering of nouns and their adjective modifiers shows great variation in early Indo-European; if asked to reconstruct one order as the original, Delbrück and Wackernagel “either flip a coin or throw [their] hands up in despair” (Jeffers, 1976, 5); they lack a framework of syntax and syntactic change to model what structural properties the two might have in common, and how one order might have emerged out of the other. In sum, these early proposals of

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

prehistoric syntax by Delbrück, Wackernagel and their generation were a good start, but there was still work to do.

In a “veritable renaissance” (Watkins, 1976, 305) for syntactic reconstruction, a new approach occurred in the 1970s, one which can deal with difference. It took as its fuel the typological universals of Greenberg (1963). From these language universals arose the notion of ‘consistency’: that there is a natural drive in language towards general typological types, in which specific features cluster together (such as a clausal VSO order and prepositions, according to Universal 3). Languages that go against these types are said to be inconsistent and ‘disharmonic’, the explanation for which is that they are languages undergoing change from one consistent type to another. Such universal shifts can be applied not only for explaining attested changes, but also for going beyond those changes into prehistory; Lehmann (1974) and Friedrich (1975) use the typological universal to claim that the variety of word orders across Indo-European are the results of the drive away from an originally harmonic grammar. For Lehmann, this included an SOV clausal order; for Friedrich, SVO.

There are flaws though in this typological approach to reconstruction, primarily in the universals from which its theoretical strength derives. Greenberg’s universals have lost much of their force since the 1960s; the number of exceptions identified (both in and out of the languages in Greenberg’s original set) have robbed the universals of the premise that they must indicate something deeper and truly universal in language, reducing them to typological tendencies. With this goes the implication that unattested languages must have also adhered to the same word-order patterns. For Watkins (1976), the comparative

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

evidence for Proto-Indo-European rejects any straightforward assignment to word-order types, such as whether it had a Verb-Object or Object-Verb order.

“The facts are such, the comparative method is such, that both word-order patterns, marked verb-initial and unmarked verb-final, must be reconstructed for the proto-language ... much of the typological debate about Indo-European as OV versus VO is purely and simply a pseudo-problem.”

(Watkins, 1976, 316)

Lehmann’s argument from typological inconsistency depends on the premise that Proto-Indo-European itself was perfectly consistent in its word order, in contrast with its descendants. As Jeffers (1976, 7) notes, this is an unrealistic view of any language. The typological approach cannot be considered to be a productive avenue for reconstruction today, although Walkden (2014, 61) suggests that typological generalisations remain a helpful heuristic to compare syntactic reconstructions against.

Turning away from the comparative method, another approach is to reconstruct on the basis of grammaticalisation, the unidirectional process by which lexical items change to become less referential and more functional. For example, English *will*, a modal verb that marks the future tense, developed from Old English *willan*, a lexical verb with the meaning of ‘to want’. Grammaticalisation is typically accompanied by a reduction in phonology; independent words become clitics and affixes. For example, definiteness is marked on Romanian nouns by a form of the obligatory suffix *-ul*, which developed through

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

grammaticalisation out of the non-obligatory, non-bound Latin demonstrative *ille*. Givón (1971) offers the following maxim:

“Today’s morphology is yesterday’s syntax”

(Givón, 1971, 413)

Concentrating on verbal affixes in Swahili, Givón uses this idea of unidirectional change and the origins of functional items to pursue the *internal* reconstruction of aspects of Proto-Bantu; this he achieves by undoing the effects of grammaticalisation within the Swahili verb. The strength of this method is that it circumvents the issue of the comparative method. Through Givón’s principle, reconstruction of earlier stages is internal to that particular language. While the proposal of grammaticalisation pathways is necessarily comparative, drawing on changes attested in other languages, it does at least avoid the correspondence problem and the question of diachronic identity in syntax across languages.

There is a weakness though in the main principle; not all functional morphology comes from formerly independent words. Phonology could be another possible origin. For example, Ripano, a Romance language spoken in Marche, Italy, has a highly productive system of gender-agreement morphemes, which emerged not from stand-alone words, but from the ‘morphologisation’ of word-final vowels (Ferrari-Bridgers, 2010). Initial-consonant mutations are a key component of the nominal morphology of Insular Celtic languages, yet they arose simply from the inter-word effects of sandhi (Hickey, 1996). We cannot consider all functional words and morphology as results of grammaticalisation. Furthermore, reconstruction on the basis of grammaticalisation pathways can tell us little about all the

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

word-order possibilities in prehistoric stages of language. Since the process typically involves the reduction of elements into clitics and affixes, we may assume lexical elements that have undergone it once had greater word-order freedom, and that an order X-Y (in which Y is a suffix) arose from an older, common order of X Y, in which the two were unbound elements. However, we cannot say what other orders were possible; it does not follow that the prehistoric syntax could not also produce a Y X order. Givón's principle of grammaticalisation remains extremely useful, but it should complement other methods of reconstruction. Where there is some evidence for the process having taken place, even if only its final stages, we have grounds to follow the path of grammaticalisation back into prehistory.

Any approach that sets out to work with the comparative method must suggest an answer to Lightfoot's correspondence problem. One answer is the approach of Harris and Campbell (1995) and their "cognate sentence patterns". Just as phonological reconstruction makes use of correspondences between languages (as in the example of 'father' and 'brother' examples above), Harris and Campbell seek to identify syntactic correspondences through short sentences that "have the same propositional meaning, and the grammatical morphemes of which correspond to each other systematically" (Ferraresi and Goldbach, 2008, 9). With examples from two Kartvelian languages, they compare the various features of two sentences equivalent in meaning, identifying not only the morphology but also the syntactic 'patterns', such as their systems of case alignment, verbal agreement and Subject-Adverb-Verb word order. It is these patterns therefore that two languages may share and have as their common inheritance.

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

This approach has met with criticism; the problem is not the empirical data used in Harris and Campbell’s cognate sentences, but rather the extent to which the empirical data corresponds to the real syntactic rules that govern them, as well as the unclear meaning of ‘pattern’. Their cognate sentences are only the surface appearance of syntax; the rules or structures that produce the surface may have crucial differences.

“The sentences found as part of linguistic data cannot be used directly to establish correspondences among different grammars. Only the internal properties of these grammars, which are identified by analyzing the linguistic data, can be used ... identifying such grammatical properties must be both a preliminary step (regarding the daughter languages) and an ultimate goal (regarding the ancestral language)”

(Pires and Thomason, 2008, 44)

Imagine two sentences in two languages, A and B, which fulfil Harris and Campbell’s criteria for being cognate sentences, sharing both meaning and functional elements. In the two sentences, the verb is the initial constituent. It is a far leap from this observation to proposing that the languages correspond in this regard, and an even further leap to say that they descend from an ancestor with the same feature. In language A, the clausal-initial verb could be in its default position, while in B its position may be unusual, marked and produced for a particular purpose.

- (1) a. *canunt* *trēs* *virī* *novī*
sing.PRS.IND.ACT.3PL three.M.NOM.PL man.M.NOM.PL new.M.NOM.PL

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

‘Three new men are singing’

(Latin. Artificial example)

- b. *canait* *trí* *fir* *nuai*
sing.PRS.IND.ACT.3PL three.M.NOM.PL man.M.NOM.PL new.M.NOM.PL
‘Three new men are singing’

(Old Irish. Artificial example)

(1a) and (1b) are plausible Latin and Old Irish examples, in which the order of the elements, including the verb, mirrors their equivalents. With the match in meaning, morphology (such as the nominative-case agreement) and even etymology, these are cognate sentence patterns for Harris and Campbell. However, in the Latin, the clause-initial position of *canunt* would be unusual and therefore marked, perhaps for an emphatic or existential reading, while *canait* occupies the normal position for finite verbs in Old Irish. It would not be reasonable therefore on the basis of (1a) and (1b) to say that they descend from a language with default verb-initial clauses. Likewise, the order of elements in the noun phrase is considerably less rigid in Latin than in Old Irish, and their mirroring in examples like (1a) and (1b) would be more simply ascribed to coincidence than an inherited pattern. The observable data of equivalent sentences may not fully represent the syntactic rules and structure as they really exist, displaying instead an accidental convergence of syntactic output. As Fortson (2006) puts it, “word order is not syntax, but a by-product of it”.

This issue of deceptive data is especially relevant to historical languages, for which the limited data may result in a misleading picture of the grammar. If what we have for an ancient language is fragmentary or composed in non-prose language, we cannot take the

1.2. *THE RECONSTRUCTION OF PREHISTORIC LANGUAGES*

word order of specific sentences as fully representative of the vernacular, especially in the absence of living native speakers who could give us a sense of normality and markedness in word order. Certain writers may mimic the style and therefore the syntax of texts in other languages, or break syntactic rules to conform to metrical templates of accent or number of syllables. We must first interrogate our sources as historical artifacts, to build a proper picture of the context of the language that they witness, and only then to apply a theoretical framework to ascertain the abstract syntax behind that language.

All of this goes to say that if we compare specific sentences, the frequency of word orders like SOV, and superficial patterns like Wackernagel's law, this does not amount to comparing the actual syntactic rules of early Indo-European languages. These phenomena are only symptomatic of underlying rules; to hit upon that bedrock of syntax, we must dive deeper.

This is where the tradition of generative grammar can be of use; it takes as foundational the view that syntax is something interior that exists to generate something exterior. It prioritises the internal grammar of an individual, which is formulated during childhood acquisition on the basis of linguistic data from the child's surroundings. This in turns generates the child's own output. This narrower view of what syntax is satisfies Fortson's problem (2006) of a "lack of precision" regarding what we aim to reconstruct. Generative grammar also has an answer for the question of what in syntax is inherited from one generation to another; each child abstracts from the word-order data and formulates anew the syntax that will generate the right output (much as they abstract phonemes from phonetic input). The internal grammars of parent and child "do not come into any direct

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

relation” during acquisition (Pires and Thomason, 2008, 35), only via the intermediary of the output.

Pires and Thomason (2008, 45) state that there are three tasks for syntactic reconstruction. It requires the proposal of hypotheses about:

- The properties of the mental grammars that could generate the outputs possible in the daughter languages
- The properties of the mental grammars that could have been internalised by speakers of the ancestral language
- 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

Generative grammar comes ready with a wealth of theoretical technology to attempt these tasks. However, it is by no means a monolithic school of thought.

1.2.4 GENERATIVE APPROACHES TO SYNTACTIC RECONSTRUCTION

Two camps within generative grammar present themselves as avenues for syntactic reconstruction: the parametric camp (Roberts, 1998, 2007; Longobardi, 2001b) and the Minimalist camp (Willis, 2011; Walkden, 2013, 2014; Windhearn, 2020; Ram-Prasad, 2022). The two differ as to how they they conceptualise the syntax to be reconstructed, and how they go about that reconstruction.

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

For Roberts and Longobardi, the concept of a syntactic parameter is of enormous utility for reconstruction. Parameters are supposed to constitute a universal switchboard-like system of binary settings, given one of two values during acquisition so as to generate the right word-order features of a particular language. Different values for these parameters are responsible for the syntactic variation that we observe cross-linguistically, and it predicts that genealogically related languages will be similar in their values. Parameters offer good *comparanda* for reconstruction, being simple (in principle each with only two possible values) and having “deductive depth” (Longobardi, 2001b), in that few parameters can explain a great number of surface features. For Longobardi and Roberts, Universal Grammar and the syntax behind all languages comprise “a finite set of binary parameters which defines the possible variation in grammatical systems” (Roberts, 2007, 505), so it is not parameters *per se* that are the target of reconstruction, but rather their set values. If we can first identify what parameters and values determine the observable word-order features of a language, we have something conceptually simple and concrete to compare across languages.

Roberts (2007, 505-7) demonstrates this approach with null subjects and the relative order of verbs and objects in early Indo-European languages. The alternations between null and obligatory subjects, and OV and VO can be framed as two binary parameters. In the second case, an OV order has the value 0, while VO is 1. Of the four languages in Roberts’s practice example, three have the value 0, including Sanskrit. Reconstruction of an ancestral 0 value and OV order is therefore supported both by the majority and the elder rule across this set.

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

The alternative camp is likewise grounded in the Minimalist program (Chomsky, 1995), in particular in the “Borer-Chomsky conjecture” (Baker, 2008).

“All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon”
(Borer, 1984)

In other words, the underlying syntax is constructed from discrete building blocks (‘heads’), which are stored in a mental repository, and which each have their own syntactic category and requirements to be satisfied through the operations of syntax. Syntactic reconstruction that builds on this conjecture has been championed by Walkden (2013, 2014) and since pursued by Windhearn (2020) and Ram-Prasad (2022). A key insight is that this ‘item-based’ view offers syntactic counterparts to the phonemes uncontroversially integral to lexico-phonological reconstruction. Just as we can reconstruct the phonemes *b^h, *p and *s for Proto-Indo-European, so too could we propose heads of the categories Verb, Noun and Adjective.

“This approach enables units of syntactic variation, lexical items, to be seen as analogous to the units of phonological variation ... in syntax we might reconstruct the lower level unit, lexical items, through their context of appearance in sentences attested in the daughter languages”
(Walkden, 2014, 55)

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

In contrast with a great switchboard of parameters hardwired in the brain, this view of syntax is minimal, involving a handful of syntactic operations that manipulate acquired lexical items. Each of the latter exists as a syntactic head, stored in a pre-syntactic repository, and each has specific features (akin to the phonological features of a phoneme), such as its category, what it takes as its complement, and what movement it triggers (see §1.3.1). Combined by syntax, these heads interact to satisfy their features. It is this computation that result in the observable phenomena of word order and morphology.

Under this view (referred to as the ‘Minimalist’ approach), the task of syntactic reconstruction is therefore to identify these items and their syntactic categories across a family of languages, and then on this comparative basis, to reconstruct either specific heads or at least their broader syntactic categories back into prehistory.

While identifying syntactic cognates remains challenging, the theory at least proposes that discrete elements exist out there to be compared. Their presence can be ascertained through both their form and their word-order behaviour. Walkden (2013) thus identifies for syntax the two levels of correspondence by which lexico-phonological reconstruction satisfies the Double Cognacy Condition (see §1.2.1). Just as reconstructing phonemes involves comparing cognate sounds in cognate words, cognate syntactic heads are realised by lexical items that occur in comparable “distributional patterns ... i.e. the syntactic environments in which they can be found in the daughter languages” (Walkden, 2013, 111). On this two-level basis, even though the lexical items themselves may differ, we have the grounds to identify cognate syntactic categories to which the items belong.¹⁰

¹⁰The gold standard of syntactic reconstruction is when not only the syntactic category and word-

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

To illustrate, we could compare the subordinators equivalent to English *that* in the Romance languages, such as French *que* and Italian *che*. On the basis of their form, function and word-order distribution, we have the triple correspondence then to reconstruct a specific head of the category of C(omplementizer) for the languages' common ancestor; this is of course corroborated by the attestation of Latin *quod* (or *quia*). The same process is applicable to early Indo-European languages like Latin and Ancient Greek, comparing their subordinators *quod* and *hóti* 'that, because'. As lexical items, they are not cognate, and so they cannot be reconstructed back to a single syntactic head, but their matching function (subordination) and distribution in the clause (clause-initial) still give us two levels of correspondence to propose a diachronically identical syntactic category of C in the two languages. To this category *quod*, *hóti* and other lexical items belong, the category having been acquired along the subordinators' phonological form and function by generations of speakers. We can posit more specific syntactic features, such as the type of clause that they head; Latin *quod* and Ancient Greek *hóti* are C_{Finite}. Once the presence of cognate heads and categories has been satisfactorily demonstrated in the syntax of the descendant languages, we can then propose a proto-head or a general proto-category as the source of their cognacy.

It should be noted that these two approaches are not at odds theoretically; both are compatible with current Minimalism (Chomsky, 1995), although the concept of parameters order distribution, but also the lexical items themselves are cognate, as established by lexico-phonological reconstruction. This 'triple cognacy' could be achieved in the case of verbs like Ancient Greek *eidénai*, Old English *witan* and Old Church Slavonic *věděti*, all meaning 'to know'. On this basis, we can reconstruct not only the category of V, but a specific proto-V-head, with the phonological form **wóyde*.

1.2. THE RECONSTRUCTION OF PREHISTORIC LANGUAGES

predates it, being a product of the earlier Government and Binding framework¹¹ (Chomsky, 1981, 1982; Chomsky and Lasnik, 1993). Instead, they differ chiefly in methodology, that is, in how they describe syntax, analyse it and go about reconstructing it. The goal of parametric reconstruction is to propose a set of cross-linguistic parameters for quantitative comparison, while the goal of the ‘Minimalist’ approach is to reconstruct qualitatively from a set of syntactic heads, but these are really two sides of the same coin; parameters capture the word order produced by the heads of the Chomsky-Borer view.

But which of the two camps should this thesis follow? It is not necessary to dismiss one outright, but it is to prioritise one for the moment. A general principle is to try the more restrictive framework first.¹² Out of these two views of syntax, the Minimalist approach is the more restrictive. Parameters may be broad, ambiguous and hard to define.¹³ The Borer-Chomsky view meanwhile reduces all syntactic features and variation to discrete, uniform syntactic items. As Walkden notes, their comparable status with the phonological units traditionally used in the comparative method means that they seem like more appropriate entities to identify and compare. The Minimalist approach therefore seems more interested in working with the comparative method to produce specific proto-forms of syntax; parametric comparison is “concerned mainly with language phylogeny” (Windhearn, 2020, 7).

¹¹‘Principles and Parameters’ is for many a synonymous alternative name for the framework.

¹²If it is too restrictive to work with the data, this suggests a problem with the framework and the less restrictive approach can then be attempted; the same cannot be said for the inverse.

¹³The relative ordering of verbs and objects across the Indo-European family cannot be captured in a single parameter. It also seems unfounded to assume that all syntactic phenomena must necessarily exist through binary settings.

1.3. THEORETICAL ASSUMPTIONS AND METHOD

This thesis joins Willis, Walkden, Windhearn and Ram-Prasad in adopting the Minimalist view of syntax for the purpose of syntactic reconstruction. Its role in this thesis is not intended to promote its singular superiority, nor deny the potential successes of other approaches. Rather, it is founded on the view that this methodology can and does produce theories of reconstruction and change as equally cogent as those within the lexico-phonological tradition. All things considered, as Clackson (2017, 204) concludes in his review of syntactic reconstruction:

“Perhaps it is time to leave the arguments about methodology to one side, and concentrate on reconstructing syntax.”

1.3 THEORETICAL ASSUMPTIONS AND METHOD

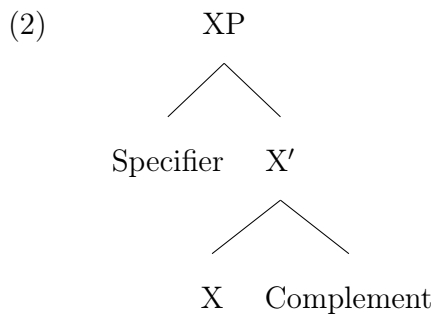
1.3.1 THEORETICAL ASSUMPTIONS: A PRIMER FOR CURRENT GENERATIVE SYNTAX

While it would be difficult to summarise all of generative thinking, an introduction to the concepts behind the analysis of this thesis is necessary, especially for non-syntactician readers, such as interested philologists. This section therefore expands on the concepts and approach briefly sketched in §1.2.3 and §1.2.4.

As mentioned, the ‘atom’ of syntax in this school of thought is the **head**. Every head has

1.3. THEORETICAL ASSUMPTIONS AND METHOD

a syntactic category, and every item of vocabulary, both words and bound affixes¹⁴, gives lexical expression and owes its word-order position to (at least) one head in the underlying syntax. Larger structures are built when a head takes another syntactic constituent as its **complement**. It also has the option to incorporate something else as its **specifier**. The latter two components belong to the **projection** of the head, and the overall unit of head, complement and specifier is a **phrase**. This produces the classical ‘X-bar’ understanding of the abstract building blocks of syntax (Chomsky, 1970), in which ‘X’ is a head of unspecified category.

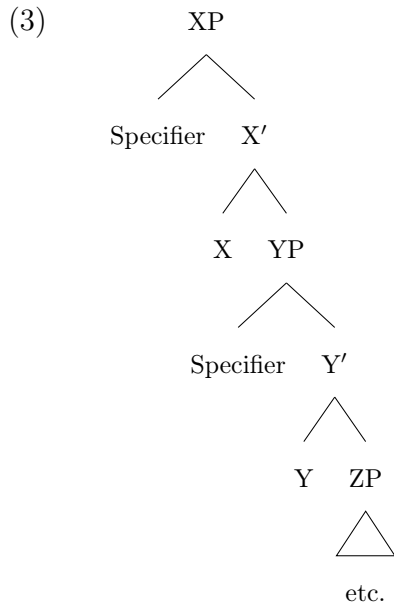


Every head has a syntactic category, which sets the category of the whole phrase; the English words *eat*, *go* or *think* arise from heads of the category of V(erb), each forming its own Verb Phrase (VP). By combining phrases, syntax builds a hierarchical structure, as in (3). This produces the grammatical word orders of a language.¹⁵

¹⁴There is famously no straightforward definition of a ‘word’; this phrasing is intended to include morphemes that are both free (e.g. English *cat*, *blue*, *walk*) and bound (e.g. the suffixes *-s*, *-ing*, *-ed*), since both are generated from the same underlying syntactic structure, an idea central to the Distributed Morphology approach (cf. Halle and Marantz, 1993).

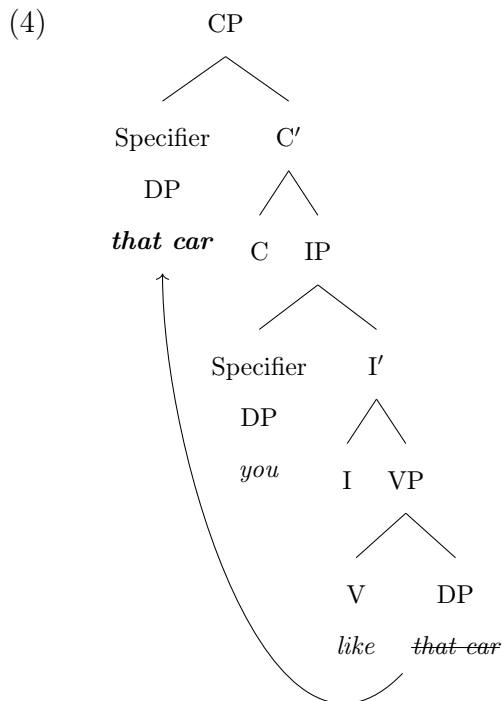
¹⁵Note: a triangle-shape at the bottom of a diagram like (3) signals that there is more to the syntactic structure, but this is not shown for the sake of space.

1.3. THEORETICAL ASSUMPTIONS AND METHOD



Syntax also involves the secondary operation of **movement**. This modifies the initial structure by moving a head or phrase from where it is ‘base-generated’ to somewhere else in the structure. Movement can be detected by an unusual location of a word or phrase in the final order. If the object of the verb in a languages typically follows the verb, but occasionally precedes for the purpose of contrast, the latter minority of orders offer evidence for the movement of the object higher than the verb. The syntax in (4) demonstrates this analysis with the English sentence *that car, you like*.

1.3. THEORETICAL ASSUMPTIONS AND METHOD

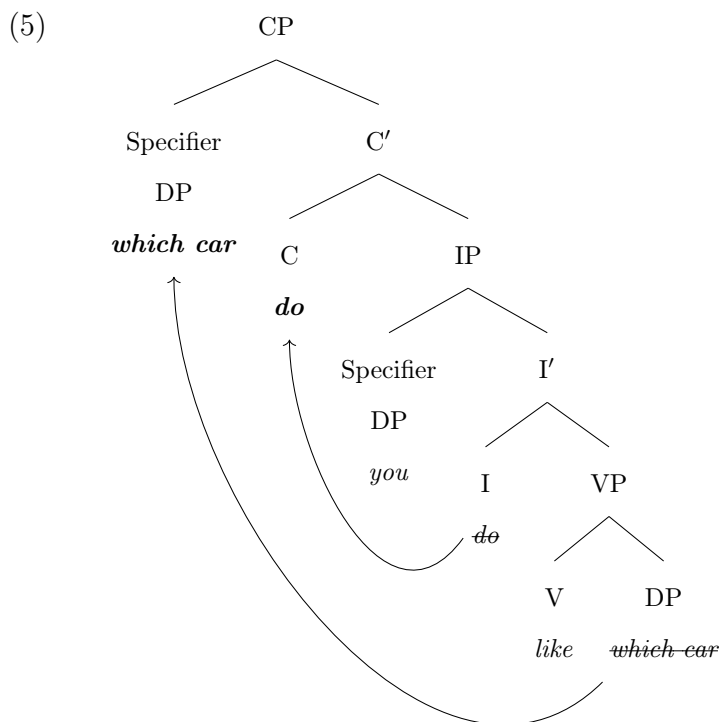


In (4), the whole Determiner Phrase *that car* is treated as moved from its basic position as the complement of the verb, to the specifier position of CP ('Spec,CP'). Being a hierarchical structure, components towards the top of a structure like (4) are said to be 'high'; those below them are 'low'. The syntactic configuration in (4) will ultimately generate the clause *that car, you like*, after the post-syntactic product is passed on for further stages of derivation. The final sequence is referred to as the linear or surface order.

One further key idea is lexically **null** elements. By this term, it is meant that there exist heads that are likewise stored in the pre-syntactic lexicon and form part of the syn-

1.3. THEORETICAL ASSUMPTIONS AND METHOD

tactic structure, but lack dedicated expression in sound or sign. Put simply, they are silent. This naturally prompts the question of where our evidence for them comes from. As syntactic heads, null elements still have their own features that they satisfy through the operations of syntax; crucially, they may trigger movement. English questions, with their distinct word-order patterns, are generally considered to involve movement upwards towards a phonologically null $C_{\text{Interrogative}}$ head. In (5), both the *wh*-phrase *which car* and the auxiliary verb *do* move up to the specifier and head positions of the CP to produce the question *which car do you like?*



1.3. THEORETICAL ASSUMPTIONS AND METHOD

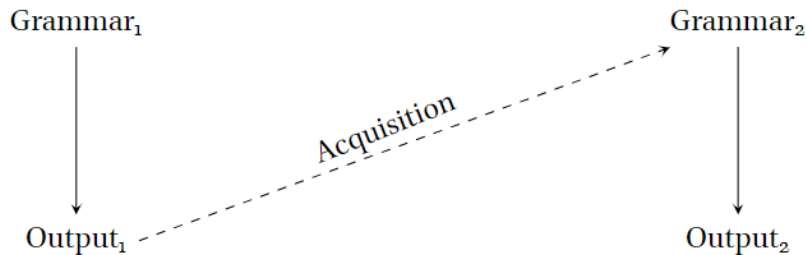
If we identify movement, we therefore have a means to identify null elements in the syntax that cause that movement; this is a key aspect of theory for this thesis.¹⁶

As a penultimate point, while a definitive view on the matter is outside the business of this thesis, one further assumption is that these syntactic categories are neither universal nor innate, but rather are formulated anew during childhood acquisition. Just as phonemes, morphemes and words are abstracted and learned from the linguistic input, so too does the acquirer conceive the broad syntactic categories like C and V, and their specific features that produce the ordered strings of elements received. Consequently, all syntactic reconstruction in this thesis is not understood to rely on the properties of so-called Universal Grammar, except perhaps the core syntactic operations of merging and moving (Chomsky, 1999). It proceeds only where the comparative evidence of the language set leads it; that destination may be the conclusion that certain syntactic items were different or even absent in the older stages of a language. This allows room for genuine change in syntax over time, which occurs at the point of acquisition. The transmission of syntax from one generation to the next was formulated by Andersen (1973) as a ‘Z’ shape. Reanalysis happens when Output_1 is received by the younger generation as their primary linguistic data, and used to abstract a Grammar_2 that is slightly different from the older

¹⁶All of this theory is consistent with the general consensus within generative grammar today, being a *koine* drawn from both Government and Binding theory (Chomsky, 1981, 1982) and the later Minimalist program (Chomsky, 1995). However, for the purposes of this thesis, the theoretical assumptions are not bound to many subsequent points of analysis popular within Minimalism, such as Kayne’s (1994) syntactic antisymmetry, which understands the relative order of specifier, head and complement within the phrase to be fixed. With this light level of commitment to the framework of generative grammar, it is hoped that much of the theory employed and analysis proposed may be easily ‘translatable’ into other syntactic schools of thought.

Grammar₁.

(6)



The ‘Z-model’ of syntactic acquisition, adapted from Andersen (1973, 767).

Yet, as noted in §1.2.2, discrepancies between internal grammars must be very minor; if they were not, noticeable discrepancies in the output would arise, and be corrected. This thesis therefore is also couched in the assumptions that such change is in general very slow and gradual, and that (pre)historical languages were not radically different from today’s, following the Uniformitarian Principle (Roberts, 2007, 274).

Lastly and relatedly, this thesis furthermore does not seek to unearth any syntax beyond the boundaries of the Indo-European family. Any resulting commonalities with the syntax of non-Indo-European languages should be taken as interesting avenues for further comparison and research, not as features of any syntactic universals. If the syntactic picture of Proto-Indo-European that this thesis paints is considered to be universal to human languages, then this view corroborates the reconstruction; if it is not considered

1.3. THEORETICAL ASSUMPTIONS AND METHOD

universal, then the reconstruction can nonetheless stand on the evidence and analysis put forward in the thesis. Universality is of no consequence; what matters is the strength of the reconstruction.

1.3.2 METHOD FOR THE THESIS

The process of researching for this thesis involved a series of reviews, working through the aforementioned seven languages one by one (though Old English and Old Norse were grouped together). Each of these reviews followed the same protocol: to survey lexical and word-order phenomena in the clauses and noun phrases of that language. To be manageable, the scope of the thesis was restricted to the clausal and nominal ‘left periphery’; it concentrated on the beginning of the two domains, researching what lexical items and syntactic patterns we observe there.

The scope was broad though enough to include a variety of general types of clause: declarative clauses, interrogative clauses, imperative clauses, and both finite and non-finite subordinate clauses. The working assumption was that there is a syntax common to all clauses in each language, and that their varying features are a product of differences in their underlying heads. To access that common syntax, multiple types of clause had to contribute their evidence. Likewise, a wide range of constituents found in noun phrases was considered besides the head noun, namely: adjectives, dependent genitive phrases, numerals, articles, demonstratives and quantifiers.

The data have mostly come second-hand. This thesis relies on the observations, gener-

1.3. THEORETICAL ASSUMPTIONS AND METHOD

alisations and analyses of previous scholars who specialise in the individual languages. The sources include large descriptive grammars¹⁷, comprehensive grammars that may or may not assume a particular theoretical framework¹⁸, shorter papers that present statistical findings from various corpora¹⁹, and papers that offer specific theoretical analysis.²⁰ The plan has been to draw from a wide pool of secondary sources that together provide a full and accurate picture of the data, to compensate for the absence of statistical investigations undertaken personally.

These reviews of the surface syntactic phenomena had a driving motive: to identify the components of the underlying syntax that could produce them, conceptualised as heads belonging to certain syntactic categories. As outlined in §1.3.1, syntax involves the arrangement of these heads into a hierarchical structure. The evidence for them comes from two main sources. For some functional categories, there will be morphophonological form to point the way; certain lexical items can be said to be generated from heads of that category on the basis of similar function and word-order behaviour. For others, lexical exponents and therefore their ‘visibility’ will be lacking, yet distinct word-order patterns can be ascribed to their features. In sum, the two sources are:

- Lexical items (i.e. words, roots, affixes, clitics)
- Instances of movement

¹⁷Such as Pinkster’s *Oxford Latin Syntax* (2015; 2021).

¹⁸Such as Faarlund’s *Syntax of Old Norse* (2004) or Thurneysen’s *A Grammar Of Old Irish* (1946).

¹⁹As in Cichosz’s study (2017) of Old English verb-initial clauses.

²⁰Such as Adger’s treatment (2006) of the Old Irish clause.

1.3. THEORETICAL ASSUMPTIONS AND METHOD

This understanding of what syntax is and involves motivated the reviews into seven early Indo-European languages. Within the left periphery of these languages, certain words can be straightforwardly understood to arise from a stable functional category and heads in the underlying syntax. Though we lack native speakers and their grammatical intuitions for these historical languages²¹, we still have a lot of extant written material for them; through quantitative and qualitative research on these sources, we can build a picture of the patterns in their word order. A foundational point is that from the previous literature, it is possible to gain an accurate idea of patterns, frequency, variation and ‘normality’ in the word order of the historical texts.²² By comparing the frequencies of the constructions that we find, we can identify productive and basic arrangements of constituents. Exceptions are always attested, but exceptional orders do not in principle negate any claims that historical languages had syntactic configurationality, and that we in the present day can recognise it.

To illustrate, secondary scholarly sources provided the information to ascertain the distribution of subordinators in Ancient Greek sources; these described not only what lexical items the texts display, but also their position relative to the rest of their clause. Evidently, Ancient Greek has a category of words that mark a clause as subordinate to another clause in its propositional content. This category includes *hóti* ‘that/because...’, *ei* ‘if...’ and *hína* ‘so that...’. From these words, we have enough evidence to propose a

²¹And therefore the negative evidence on which so much of generative thought and research are founded.

²²Syntactic analysis of historical languages always suffers from a dual problem of ‘too much’ and ‘too little’; we have too much variety in our attested word order to propose neat, exception-less analyses, yet also too little philologically safe data to build any analysis on.

1.3. THEORETICAL ASSUMPTIONS AND METHOD

syntactic category of C(omplementizer) to which they belong, and through membership of which they share the same word-order features. The picture of the data is that C-elements in Ancient Greek overwhelmingly stand first in their clause, with a minority of orders in which they come second.

This proposal of C in Ancient Greek draws additional strength from apparent cases of movement. *Wh*-words and phrases, such as pronominal *tís* ‘who’ or adverbial *poû* ‘where’, display a “strong tendency” (Goldstein, 2015, 71) to occur in the initial position of their clause, regardless of the category of the interrogative element or the length of the clause. So, just as clauses with *hóti* and *ei* have a C head that indicates their subordinate status, Ancient Greek questions may have their own interrogative element of category C, which makes its presence felt through movement. What we propose for Modern English works well for Ancient Greek too.

There are therefore two avenues by which this project has reached its conclusions about the syntax of early Indo-European languages, and identified the syntactic heads that fuel the Minimalist approach to syntactic reconstruction. Understood through Andersen’s Z-model in (6), the historical data are the (literary) output through which we can diagnose the underlying grammar. Having completed these synchronic syntactic analyses for a set of related languages, the thesis reached a position to compare them, and then to derive the syntactic heads and categories from a proto-syntax, in effect following multiple Z-lines of acquisition over millennia back to their prehistoric point of convergence.

1.4 THE INDO-EUROPEAN LANGUAGE SET

As mentioned, the first order of business is to undertake Pires and Thomason's first task: to propose the internal grammars of the daughter languages of the ancestral language to be reconstructed. Regarding Proto-Indo-European specifically, Fortson agrees; "a proper theory of the development of the syntax of Indo-European languages is ... not yet attainable" until we first have "theoretically sophisticated analyses" of the syntax of a broad set of Indo-European languages (Fortson, 2006). Since a total survey of Indo-European clauses and noun phrases would be a task for several lifetimes, a subset of representative Indo-European languages is required. These should be early in their date of attestation, on the assumption that "the older a daughter language is, the more likely it is to be closer to the ancestor ... forms in an older language may carry more weight in determining a reconstructed form" (Roberts, 2007, 359). They furthermore should capture a wide spread across Indo-European, in order to target the prehistoric point of convergence of all the family, and thereby avoid a reconstructing only a branch on the Indo-European tree. Following these principles, this thesis concentrates on the following seven languages: Latin, Ancient Greek, Vedic Sanskrit, Old Church Slavonic, Old English, Old Norse, Old Irish.

These fit the principles above, given that each language is an early or the earliest source for the Italic, Hellenic, Indo-Aryan, Slavic, Germanic and Celtic sub-families of Indo-European. With this broad grouping, we can reduce the chances of reconstructing what are actually products of inter-dialectal contact or external influences, and be more

1.4. THE INDO-EUROPEAN LANGUAGE SET

confident that the commonalities across the family that we identify are inheritances from a prehistoric common ancestor.

Each language of the set is historical; it is known from textual sources only. Granted, there is a great disparity in their dates of attestation; there is an almost three-thousand-year divide between the oldest (Vedic Sanskrit) and the most recent (Old Norse), but to meet the criterion for a wide distribution, this cannot be helped. This selection may surprise some scholars in what it includes and excludes. The following biographies therefore introduce the seven languages and justify their inclusion. Additional factors were the size of the language's corpus, the feasibility of accessing the data and previous scholarship, and avoiding the interference of foreign languages in our extant sources. These factors were behind the exclusion of some languages; corpus size and a lack of syntactic complexity meant that languages like Gaulish and Phrygian are absent from the set, while Classical Armenian was avoided on account of the considerable influence of Persian on the language.

1.4.1 LATIN

The Latin language first emerges into the historical record as the primary language of the city of Rome and the region surrounding the Tiber river. Our earliest sources date to the 7th century BC, when Latin was only one among many languages in the Italian peninsula; Latin is the chosen representative of the wider Italic language family. 'Early' Latin is the specific focus; this groups together Old Latin (up to *c.*100 BC) and Classical Latin (*c.*100 BC to *c.*200 AD), and so to draw from a large corpus of prose and plays by Roman

1.4. THE INDO-EUROPEAN LANGUAGE SET

authors like Plautus, Cicero, Caesar and Livy. It excludes Late Latin and also Medieval Latin, which thrived in the medieval world as an elite second language that stuck closely to classical standards, while the vernacular continued its development into the Romance language family.

1.4.2 ANCIENT GREEK

Ancient Greek offers another strand of Indo-European, spoken across the eastern Mediterranean region, with a heartland in the southern Balkans and Aegean islands. Even more so than 'Latin', the term 'Ancient Greek' is a simple name that masks a vast amount of linguistic variation, both chronological and geographical. Used generally, it covers almost two millennia of language, from the 16th century BC and up to even the 4th century AD. This thesis draws evidence from both Homeric (*c.*800 BC) and Classical Greek (*c.*500 BC - *c.*300 BC), the latter being language of prose writers like Plato and Herodotus. While Homeric is the older form and therefore has a privileged position in Greek's contribution to comparative Indo-European research, the focus of this thesis also includes later Classical Greek, in an effort to mitigate the constraints of metric templates on the word order of Homer and Hesiod. It excludes the even earlier Mycenaean Greek, because of the syntactic simplicity of its accessible texts.

1.4.3 VEDIC SANSKRIT

The Sanskrit language represents the Indo-Aryan and greater Indo-Iranian branches of Indo-European. With the possible exception of the scarcely attested Mitanni Indo-Aryan (Masica, 1991, 35-7), Sanskrit is the oldest documented Indo-Aryan language, with thousands of years of use and surviving attestation up to the present day. A traditional division is into the older Vedic Sanskrit (*c.*1500–*c.*600 BC) and the later Classical Sanskrit, though the latter does not derive directly from the former (Oberlies, 2017, 448); they rather represent two stages and regional dialects of Old Indo-Aryan.

Vedic Sanskrit is the language of the Vedas, religious and philosophical works that were transmitted orally for centuries. There are four canonical Vedas, the oldest being the Rigveda (RV). Our written sources are considerably later than the language that they witness, with the oldest documents for the Rigveda only dating to the 11th century AD. The Rigveda contains over a thousand hymns in ten books and is readily available in accessible sources, and so it is the primary Sanskrit source used in this thesis.²³ Books 2-8 are “generally accepted” to be “the most ancient core” (Cardona, 2017, 310) in terms of their language and date of composition, dated by many to *c.*1500 BC. This places them among our oldest sources for Indo-European.

Occasional mention is made of Avestan, an ancient Iranian language close to Vedic

²³The concept of the *sam̐skṛtá-* (‘refined’) language is a post-Vedic development, motivated by the relegation of Sanskrit to specialist uses and its linguistic divergence from the vernacular, which became the varieties of Middle Indo-Aryan. Since this thesis draws overwhelmingly from the language of the Rigveda, simply ‘Vedic’ would be an appropriate alternative label for the examples.

1.4. THE INDO-EUROPEAN LANGUAGE SET

Sanskrit in both antiquity and linguistic structure. Much of what pertains to Sanskrit word order applies also to Avestan (see Hale 1993; West 2011, 105-29; de Vaan and García 2014, 103-4 *inter alios*).

1.4.4 OLD CHURCH SLAVONIC

For this thesis, the representative of the Slavic sub-family of Indo-European is Old Church Slavonic. This liturgical language provides in manuscripts the bulk of our oldest extant Slavic sources, which date back as far as the 10th century AD, though some are copies of earlier documents. The Old Church Slavonic literary tradition begins in the 9th century with the missionary work led by the brothers Constantine (later Cyril) and Methodius. Tasked with proselytising the Slavs of Greater Moravia, they required translations of important Christian texts. These were based on the Slavic variety spoken in and around their home city of Thessaloniki; this dialect is situated genealogically within South Slavic (Langston, 2018, 1398-9).²⁴ There is also no natural break between *Old Church Slavonic* and the rest of ecclesiastical Slavic language since; the term is limited by convention.

“We reserve the name OCS for the language of a relatively small group of texts that are thought to have some direct connection to the original Cyrillo-Methodian mission or the subsequent work of their disciples in Bulgaria-Macedonia, and which preserve certain archaic features. These texts were composed and

²⁴However, Old Church Slavonic is a transnational, literary language and its production was not limited to one geographical area. The South Slavic dialect used by Constantine and Methodius provided an influential standard, but regional variation is detectable in the different recensions of the language.

1.4. THE INDO-EUROPEAN LANGUAGE SET

copied from the second half of the ninth century through the eleventh century, but the majority of the surviving manuscripts date to the eleventh century.”
(Langston, 2018, 1399)

Hence, the documents with such a connection to the original mission, the Czech-Moravian and the Macedonian-Bulgarian recensions, make up the accepted Old Church Slavonic canon.²⁵ Early manuscript sources, which avoid the interference of later copyists, include the *Codex Marianus* (*C. Mar.*), the *Codex Zographensis* (*C. Zogr.*), the *Codex Suprasliensis* (*C. Supr.*) and the Kievan Folia. In terms of genre, the corpus contains translations of the gospels and psalms, prayers, sermons, missals and two narrative accounts of the lives of Cyril and Methodius. With the likely exception of the two *vitae*, this genre means that the canon comprises translations of Greek (or Latin) originals. This has bearing on the syntax of our sources; imitation of the Greek word order seems to have driven the Slavic either in its first composition or in later copying.

“The word order in the O.C.S. Gospel texts overwhelmingly follows the Greek as faithfully as is possible, and deviations usually turn out to coincide with known Greek variants. We may wish to assume that the original O.C.S. version had independent and native word order, but the O.C.S. codices offer only hints as to what it may have been.”
(Metzger, 1977, 440)

²⁵Moreover, specific grapho-phonological features are used as criteria for inclusion or exclusion, most notably the expected correct usage of the nasal vowels ϵ and ρ .

1.4. THE INDO-EUROPEAN LANGUAGE SET

Consequently, any conclusions about its syntax must be couched in philological caution. Discrepancies between Greek and Slavic grammar are of great utility.

1.4.5 OLD ENGLISH AND OLD NORSE

The representatives chosen for the Germanic subfamily of Indo-European may be surprising for two reasons: the choice and the number. To address the first, it should be acknowledged that the usual specimen for early Germanic is Gothic, which has both an early date of composition (mostly 4th century AD) and a substantial corpus. However, while Gothic is of enormous utility for historical phonologists, there are doubts as to the Germanicity of its word order; Gothic surface syntax, especially that of the language of Wulfila's Bible translations, looks to have been greatly determined by the syntax of original texts in Greek. This very general view has naturally been challenged; Walkden (2014, 12) writes that it "seems to be based, more often than not, on simple intuition", and various scholars have striven to identify specific 'real' Germanic features within the available Gothic word order. While we can no doubt make progress through comparison, inference and a focus on discrepancies between Greek and Gothic morphology, this thesis circumvents the issue by simply avoiding Gothic.

Instead, Old English and Old Norse represent the Germanic family. Although considerably later in date than Gothic, the two have large corpora, containing poetry and prose, much of which is original composition in that language. Both have data to offer the goals of this thesis, and by including more than one early Germanic language, their commonal-

1.4. THE INDO-EUROPEAN LANGUAGE SET

ities allow us to get back to their immediate proto-language, Proto-Northwest-Germanic (Ringe, 2006, 213-4). This was a language as old as, if not older, than attested Gothic.

The term ‘Old English’ covers the varieties of West Germanic that arrived and were spoken in Britain following the departure of the Roman administration, likely starting from the 5th century AD, up until the end of the 11th century and the Norman Conquest. While we do have examples of Old English from those earliest times, the extant language is fragmentary and short until the more extensive manuscript material of the late 9th century onwards (Gneuss, 1991). As well as poetry like *Beowulf*, prose texts are numerous, such as the Old English *Martyrology* (9th century) or Aelfric’s homilies (10th century).²⁶

Old Norse is a North Germanic language, or rather part of a dialectal continuum. The term, precisely used, refers to the language of texts first written in Norway and Iceland from the twelfth to fourteenth centuries (Haugan, 2000, 5). While sources of North Germanic are available from other areas and from earlier eras, these texts remain the oldest substantial witnesses to North Germanic. Their dating is significantly later than Gothic, and even Old English, making Old Norse a contemporary instead of Middle English. In terms of prose, the corpus comprises numerous sagas (legendary and historical narratives) and factual works like the *Prose Edda*.

²⁶There is considerable geographical variation in Old English sources, but dialectal differences do not seem to have any serious bearing on the data for word order.

1.4.6 OLD IRISH

Old Irish is the name given to a historical stage of the Irish language, dating to between *c.*600 and *c.*900 AD. It is not the oldest documented stage of Irish; it postdates Primitive Irish, known from a small corpus of short inscriptions written in the Ogham script and dating to between the fourth and sixth centuries. Our Primitive Irish sources reveal that the language underwent drastic phonological and consequently grammatical changes during the early medieval period, but they are of little use for complex syntactic analysis. Studying the syntax of the extinct Continental Celtic languages, such as Gaulish, would likewise struggle with the same paucity of data. Old Irish then is the most suitable representative of the Celtic branch of Indo-European, being the oldest form of Celtic for which the attested evidence will “permit a full synchronic description” (Stifter, 2009, 59).

The following three manuscript texts are major contemporaneous witnesses to Old Irish: the Würzburg glosses on the Pauline Epistles (Wb.), the Milan glosses on a commentary on the Psalms (Ml.) and the St. Gall glosses on Priscian’s *Institutiones* (Sg.). These sources date from the second half of the 8th century to the mid-9th century, and consist of annotations to Latin manuscripts, written by Irish scholars in a prose style taken to reflect something close to the everyday language of the glossators. This thesis draws primarily from these three sources.

1.4.7 THE EXCLUSION OF HITTITE AND ITS CONSEQUENCES

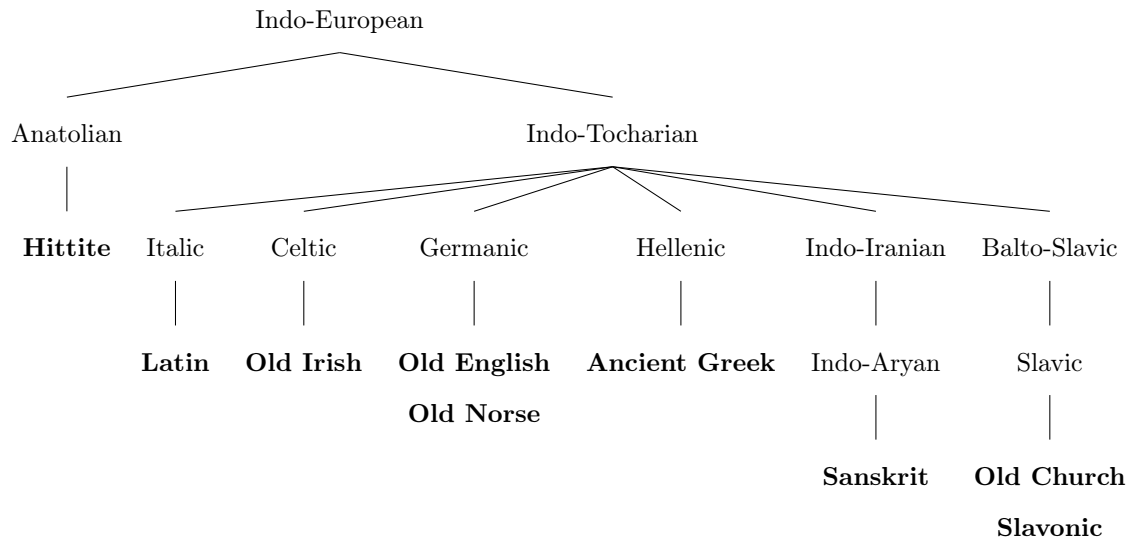
Among the early Indo-European languages absent from this set, Hittite is perhaps the most glaring omission. It certainly meets the criteria; the language is very old (second millennium BC), has a substantial corpus and could represent another sub-family of Indo-European (Anatolian). The texts are long enough to allow detailed syntactic analysis, both in and out of a specific theoretical framework (cf. Melchert and Hoffner, 2008; Sideltsev, 2015a).

At the start of this project, Hittite was excluded from the set because of the challenges of working with the language. Hittite is written in cuneiform, which requires dedicated (self-)tuition to master its syllabic and logographic systems, as well as its tradition of transliteration into Roman letters. Hittite, while long acknowledged to be Indo-European, also differs greatly from the rest of the family in phonology, morphology and vocabulary. For Hittite to make a useful and original contribution to the syntactic analysis of this thesis, these challenges demanded a level of skill that the researcher at the outset did not have. An elementary understanding has since been acquired, so occasional reference to Hittite is made. The language is a natural avenue for further work.

Until undertaking that work, the reconstructions of the thesis cannot be claimed to pertain to Proto-Indo-European in its oldest hypothetical state, but only to late Proto-Indo-European or ‘Proto-*Core*-Indo-European’. This, the “Indo-Tocharian” node in (7), developed after the split with what would become the Anatolian branch (Kloekhorst, 2022, 216), and is the immediate common ancestor of the language set.

1.4. THE INDO-EUROPEAN LANGUAGE SET

(7)



The ‘neo-traditional’ model of the Indo-European family, adapted from Olander (2022a, 5).

Further work on whether Hittite conforms to or rejects the analysis, and what we can say about the older stage of the proto-language is an exciting prospect.

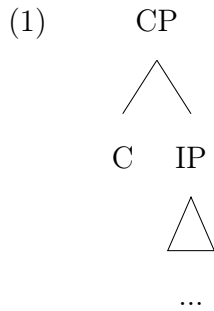
C IN PROTO-INDO-EUROPEAN

2.1 INTRODUCTIONS AND CHAPTER AIMS

The task of this chapter is to lay out the evidence for the hypothesis that the syntax of the Proto-Indo-European (finite) clause contained a head of the functional category C. The function of C was to encode information about the type and status of its clause. By ‘type’, this means its purpose and (il)locutionary act (i.e. asserting, commanding, requesting information, (see Austin, 1975)). Furthermore, this C component encoded the clause’s status as either an independent clause or one in some way subordinate to another. In this thesis, these two key functions of C and qualities of a clause are often subsumed for short under the term ‘type’.

2.1. INTRODUCTIONS AND CHAPTER AIMS

To the category C belonged syntactic heads with different clause-typing features, i.e. $C_{\text{Declarative}}$, $C_{\text{Interrogative}}$, and in the absence of reconstructed lexical exponents, we must further posit that C heads could be lexically null. The C head occupied a high position in the overall syntactic structure, above the basic positions for the lexical components of the clause, such as the finite verb and its nominal arguments. This abstract configuration looks so:



The argument is more specifically that C in Proto-Indo-European is the unitary origin of a set of five general phenomena in the left periphery of the early Indo-European clause:

- Lexical markers of clause type
- The syntactic distinction between main and subordinate clauses
- *Wh*-movement
- Certain instances of clause-early verbal position
- ‘Second-position’ pronouns

2.1. INTRODUCTIONS AND CHAPTER AIMS

Concurrently, the chapter aims to demonstrate that the adoption of this C is an analysis superior to two alternatives: that the Proto-Indo-European left periphery either had ‘less syntax’ (i.e. no C), or that it had ‘more’ (i.e. several heads responsible for these phenomena). Against the first, this chapter argues that the phenomena are too similar and consistently present across the language set to be ascribed to unrelated, parallel developments, and should be explained as inherited from a language with configurational syntax. Against the second, the view is that the word-order data mentioned above do not require the proposal of a more complex syntactic structure. Behind the rejection of both analyses is the principle of parsimony; a unitary CP in Proto-Indo-European efficiently gives us what we want.

As discussed in Chapter 1, the reconstruction of a proto-category of C relies completely on demonstrating the presence of C in finite and non-finite clauses in the descendent Indo-European languages. This presence is detected both on account of lexical items that belong to the category, and the instances of syntactic movement that C triggers. This chapter therefore builds the case in the following order: first, it reviews the lexical elements that belong to the category of C across the language set; second, it discusses a variety of cases of movement to either the C head or the specifier of CP. It then summarises and concludes with a description of the features and behaviour of C in Proto-Indo-European, according to what the comparative evidence allows us to conclude.

2.2 LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

2.2.1 SUBORDINATING ELEMENTS IN EARLY INDO-EUROPEAN

It seems appropriate to begin this line of argument with subordinators (also known as conjunctions, subordinating conjunctions, complementizers), since these are the members *par excellence* of the syntactic category C; the label ‘C’ arose as shorthand for ‘complementizer’. Subordination is a syntacto-semantic quality of a clause that is used to group together a wide range of clause types, including declarative, interrogative, temporal, causal, conditional and complement subordinate clauses. In brief, elements that mark their clause as semantically and syntactically subordinate to another are a consistent feature of early Indo-European languages. Across the seven languages considered, their word-order position is very stable, standing either initially or at least early in their clause. This is no trivial fact; equivalent subordinators in other languages may be clause-internal (e.g. Wakhi) or clause-final (e.g. Japanese) (Dryer, 2013a).

LATIN SUBORDINATORS

Latin for one has a broad range of elements that communicate various kinds of relationship between their clause and another.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

- (2) a. *patrem hūc orātō [ut*
 father.M.ACC.SG hither entreat.FUT.IMP.ACT.2SG that
veniat]
 come.PRS.SBJ.ACT.3SG
 ‘Would you entreat my father to come here’
 (Latin. Plautus, *Asinaria* 740)
- b. *dixī [quia mustēla comēdit]*
 say.PRF.IND.ACT.1SG that weasel.F.NOM.SG eat.PRF.IND.ACT.3SG
 ‘I said that the weasel had eaten it’
 (Latin. Petronius, *Satyricon* 46.4)
- c. [*dum haec in conloquiō geruntur]*
 while this.N.NOM.PL in discussion.N.ABL.SG bear.PRS.IND.PASS.3PL
Caesarī nūntiātum est
 Caesar.M.DAT.SG announce.PRFP.NOM.SG be.PRS.IND.3SG
 ‘While these things were happening in the discussion, it was announced to
 Caesar...’
 (Latin. Caesar, *Gallic War* 1.46)
- d. *quōs quidem ego [sī ullō modō*
 who.M.ACC.PL indeed I.NOM.SG if any.M.ABL.SG way.M.ABL.SG
fieri possit] nōn tam ulcisci
 happen.PRS.INF be-able.PRS.SBJ.3SG not so take-revenge.PRS.INF.PASS
studeō
 strive.PRS.IND.ACT.1SG
 ‘Whom indeed, if it could by any means be done, I strive not so much to take
 revenge on...’
 (Latin. Cicero, *Catilinarian Orations* 2.8)

The *ut*-clause in (2a) is subordinate, because it is what the father is to be asked to do, and so *ut* simply introduces the object of the main-clause verb. *Quia* (2b) and *quod* (5a)

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

may likewise have this primarily grammatical function, except with *realis* clauses. The subordinator *dum* in (2c) marks its clause as the wider temporal context of the event in the main clause. *Sī* in (2d) presents a condition on which the main clause depends.

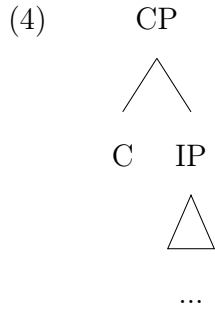
In (2), the subordinators all occupy a clause-initial place in the clause. This is the regular position for markers of subordination across Latin; (3) below contains two subordinate clauses, one embedded inside the other, in each of which again the subordinating device (the temporal *ubi* and the relative pronoun *quam*) is the first element.

- (3) [*ubi* *ea* *diēs* [*quam* *cōstituerat* *cum*
 when that.F.NOM.SG day.F.NOM.SG which.F.ACC.SG set.PPRF.IND.ACT.3SG with
lēgātīs] *vēnit*]...
 ambassador.M.ABL.PL come.PRF.IND.ACT.3SG
 ‘When that day that he had appointed with the ambassadors came...’

(Latin. Caesar, *Gallic War* 1.8)

Pinkster (2021, 955) summarises the situation: “in finite subordinate clauses the first position of the clause is restricted to subordinating devices ... unless there are pragmatically prominent competing expressions”. On this basis, we can propose that there is a functional syntactic category of C to which Latin subordinators belong, and which is a part of the underlying syntax of Latin subordinate clauses. Subordinators’ position in the linear left periphery means that the C head is present in the highest ‘zone’ of the syntax; C selects the rest of the clause as its complement, and the CP may be itself selected as an object, subject or adjunct within another clause. This syntactic behaviour of C can be modelled as follows:

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C



However, the observed position of Latin subordinators is not rigid. Constituents of the subordinate clause sometimes precede them, although this is not without limits and it does not contradict the establishment of C. The subordinator cannot appear clause-finally, and the greater the number of pre-subordinator constituents there are, the rarer the word order. Pinkster (2021, 979) explains this minority of orders as pragmatic fronting from within the clause to the left of the subordinator in the linear order. (5) illustrates such subordinator-second orders.

- (5) a. *equidem sciō iam [filius quod amet meus istanc meretricem ē proxumō Philaenium]*
indeed know.PRS.IND.ACT.1SG now son.M.NOM.SG that
love.PRS.SBJ.ACT.3SG my.M.NOM.SG that.F.ACC.SG courtesan.F.ACC.SG
from nearest.N.ABL.SG Philaenium.F.ACC.SG
‘Indeed I now know that my son may love that courtesan from nearby, Philae-
nium’

(Latin. Plautus, *Asinaria* 1.1)

- b. [*cōsul postquam dētractārī certāmen*
consul.M.NOM.SG after refuse.PRS.INF.PASS battle.N.ACC.SG

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

vīdit] *posterō* *diē* *in cōnsilium*
 see.PRF.IND.ACT.3SG following.M.ABL.SG day.M.ABL.SG in council.N.ACC.SG
advocāvit

summon.PRF.IND.ACT.3SG

‘After the consul saw that battle was refused, the next day he called a council’

(Latin. Livy, *History* 37.39, from Pinkster (2021, 979))

- c. *nunc* [*internōsse* **ut** *nōs* *possītis*
 now know-apart.PRS.INF.ACT that we.ACC.PL be-able.PRS.SBJ.2PL
facilius] *ego* *hās* *habēbō*...

more-easily I.NOM.SG this.F.ACC.PL have.FUT.IND.ACT.1SG

‘Now so that you can tell us apart more easily, I will have these...’

(Latin. Plautus, *Amphitruo* 1. prologue, from Pinkster (2021, 979))

The elements that precede the subordinators in (5) are fronted to signal their prominence in the discourse; *filius* is fronted as the topic of conversation in (5a), just as the consul that precedes the subordinator *postquam* in the example (5b) is the topic of Livy’s narrative. In (5c), *nunc* serves an introductory, attention-grabbing function, and can be analysed as belonging to the superordinate clause, but the infinitive *internōsse* belongs to the subordinate *ut*-clause; we can analyse *internōsse* as fronted before *ut*, being the topic of the overall sentence.¹ The evidence therefore suggests a topic-associated position above the elements in C, which hosts fronted material; this is discussed in Chapter 3.

It must be noted that all the subordinators considered so far appear in *finite* subordinate clauses; Latin also has morphology at its disposal to build non-finite verbs and clauses. A given verb may form infinitives, participles, supines, gerunds and gerundives. These

¹Namely how to distinguish the speaker, the god Mercury, from a servant who he is imitating. Alternatively, this is comedy, so its position relative to *ut* may instead be set by the usual iambic metre.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

verb forms can have arguments and adjuncts of their own, and can express a range of tenses and voices. Non-finite clauses therefore almost match finite clauses in propositional content, and there is some, albeit limited, evidence that non-finite clauses also arise from a syntax with a CP structure like (4). Perhaps the two best known non-finite constructions are the Ablative Absolute (6) and the Accusative with Infinitive (7). The former makes use of the ablative case to join nouns, adjectives and participles as one unit; it is a very productive construction, often providing finite clauses with circumstantial information. The Accusative with Infinitive, comprising an infinitive verb and an accusative subject, is equally productive in Latin; “from Old Latin to Late Latin ... it represents the canonical form of a complement sentence” (Ferraresi and Goldbach, 2003, 5).

- (6) *omnibus rēbus ad profectiōnem comparātis...*
 all.F.ABL.PL thing.F.ABL.PL for departure.F.ACC.SG prepare.PRF.F.ABL.PL
 ‘With all things ready for the departure...’

(Latin. Caesar, *Gallic War* 2.12.2)

- (7) a. *difficile est [amicitiam manēre]*
 difficult.N.NOM.SG be.PRS.IND.3SG friendship.F.ACC.SG remain.PRS.INF.ACT
 ‘It is difficult for friendship to remain’

(Latin. Cicero, *On Friendship* 37)

- b. [*rem tē valdē bene gessisse*]
 matter.F.ACC.SG you.ACC.SG very well handle.PRF.INF.ACT
rūmor erat
 rumour.M.NOM.SG be.IPRF.IND.3SG
 ‘The rumour was that you had handled the matter very well’

(Latin. Cicero, *Letters to Friends* 1.8.7)

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

Unlike finite subordinate clauses, the Ablative Absolute and Accusative with Infinitive here are not marked by a lexical subordinator like *ut* or *quod*. Moreover, they can exhibit the verbal features of tense and voice, but not mood or person. For these reasons, we might claim that the underlying syntax of non-finite Latin clauses in Latin is less complex, lacking certain functional projections, such as CP. However, the issue of case contradicts this; sentences like (7a) and (7b) raise the question of where the subordinate subject receives its accusative case from. While the Accusative with Infinitive often occurs as the object of a main-clause verb, it is not limited to this context. The accusative case of the subject is therefore not dependent on the donation of case by the higher verb. The absence of a case assigner outside the subordinate clause permits us to propose a lexically null case assigner within it (Cecchetto and Oniga, 2002; Melazzo, 2005). Specifically, this is a null element of the category C; consequently, the Accusative with Infinitive clause can also be considered a CP. By adopting this analysis, the theoretical consequence is that some Latin non-finite clauses may match finite clauses in syntactic structure.² All things considered, subordinate clauses in Latin, not only finite but also non-finite, support the proposal of the functional category C, which occupies a high position in the underlying syntax.

ANCIENT GREEK SUBORDINATORS

From the Homeric period onwards, Ancient Greek displays a plethora of subordinators.

²Melazzo (2005) notes that this parity should not come as a surprise; the two types of subordinate clauses (finite and non-finite) exist alongside each other in Latin and a common structure would later facilitate the replacement of the Accusative with Infinitive and the generalisation of finite *quod*-clauses in vernacular language.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

Temporal subordinators (‘when’, ‘until’) include *hēōs* and *hóte*; the conditional (‘if’) subordinator is *ei*; purposive clauses can be formed with *hína*, and complement clauses with *hóti* and *hōs*, which also have causal uses. The first two types form subordinate clauses that generally precede the main clause (Dunn, 1988, 68), providing it with background information or a dependent condition. Some subordinators co-occur or even fuse with the modal particle *án* to express a more indefinite meaning, such as *hótan* ‘whenever’.

- (8) a. *hoútoi élegon [hóti Kúros =mèn*
 this.M.NOM.PL say.IPRF.IND.ACT.3PL that Cyrus.M.NOM.SG and
téthnēken...]
 die.PRF.IND.ACT.3SG
 ‘They said that Cyrus was dead’

(Ancient Greek. Xenophon, *Anabasis* 2)

- b. [*hína Sōkrátēs tò eiōthòs*
 so-that Socrates.M.NOM.SG the.N.ACC.SG be-used-to.PRFP.ACC.SG
diapráxētai...]
 accomplish.AOR.SBJ.MED.3SG
 ‘So that Socrates might wangle what he’s used to...’

(Ancient Greek. Plato, *Republic* 1.337)

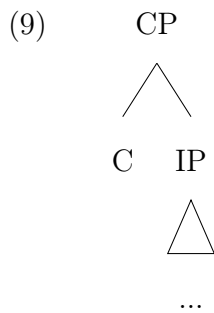
- c. [*hóte tous árkhontas hōmológeis [ouk*
 when the.M.ACC.PL leader.M.ACC.PL agree.IPRF.IND.ACT.2SG not
anamartétous eînai]]...]
 unerring.M.ACC.PL be.PRS.INF
 ‘When you agreed that leaders are not infallible...’

(Ancient Greek. Plato, *Republic* 1.340)

As (8) exemplifies, the usual position of subordinators is clause-initial, preceding all the

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

remainder of the clause, including clausal settings (Allan, 2012, 23). On the basis of this common and productive syntactic category, and its fairly strict position within the clause, it is reasonable to derive it from a C head and a CP. This CP again stands above the lexical ‘zone’ of the syntax, comprising the verb and its arguments.



Like Latin, Ancient Greek also has a wide range of verbal morphology to build non-finite clauses; there is possible evidence that these too have the CP structure in (9). One clause-like construction involves the Ancient Greek infinitive, which can be used on its own or preceded by the neuter singular definite article. The ‘articular infinitive’ can itself be the complement of various prepositions; these together create adverbial clauses like the temporal context in (10).³

- (10) *[en tô_i phroneîn =gàr mēdèn]*
in the.N.DAT.SG think.PRS.INF.ACT for nothing.N.ACC.SG
hēdistos bíos
sweetest.M.NOM.SG life.M.NOM.SG
‘For when thinking of nothing is life sweetest’

³The structure seems to be a post-Homeric innovation, since it is attested only once in the Greek of that era (*Odyssey* 20.52) (Fykias, 2013).

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

(Ancient Greek. Sophocles, *Ajax* 545)⁴

The Genitive Absolute (11) combines a noun (phrase) and a participle to supply additional information to the main clause. In Homeric language, it generally expresses the time at which the main proposition takes place, but goes on to develop causal, concessive and conditional functions (Spieker, 1885, 316). This non-finite clause requires at least a subject and a verb, as does the Accusative with Infinitive construction (12).

- (11) *Kûros* =*d' oûn anébē* *epì tà*
 Cyrus.M.NOM.SG and then go-up.AOR.IND.ACT.3SG over the.N.ACC.PL
órē [*oudenòs kōlúontos*]
 mountain.N.ACC.PL no-one.M.GEN.SG prevent.PRSP.M.GEN.SG
 ‘And then Cyrus climbed over the mountains without hindrance’

(Ancient Greek. Xenophon, *Anabasis* 1.2.22, from Smyth (1920, §2070))

- (12) *oîmai* [*theón =tina autòn ep'*
 suppose.PRS.IND.MED.1SG god.M.ACC.SG some.M.ACC.SG he.M.ACC.SG upon
autèn agageîn tèn timōrían]
 same.F.ACC.SG lead.AOR.INF.ACT the.F.ACC.SG punishment.F.ACC.SG
 ‘I imagine that some God brought him unto the same punishment’

(Ancient Greek. Lycurgus, *Against Leocrates* 1.91)

Most non-finite subordinate clauses in Ancient Greek do not have an introductory element to mark the clausal boundary, but there are exceptions; *prín* ‘before’ and *hóste* ‘so that’

⁴The equals sign (=) is used in examples throughout to indicate clisis and separate out a enclitic or proclitic element from its host. In particular, it is used with ‘second-position’ conjunctions and pronouns that follow Wackernagel’s law (see §2.3.5). It does not correspond to anything in the original Greek, Sanskrit or Old Irish sources.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

precede infinitival verbs and their arguments (Sevdali, 2007, 134-9). Like their finite counterparts, these elements stand in first position, and their behaviour can be derived from a CP structure with a non-finite C head. As for the articular infinitive though, this likewise consistently displays an Article-Infinitive order, so we might identify the definite article as another C-type element. However, the definite article (and any governing preposition) exhibit the same behaviour as they do with ordinary nouns, so the articular infinitive is most simply explained as containing a syntactically unchanged definite article, while the infinitive has a nominal feature that allows for their combination; the construction offers no evidence for a CP.

As for non-finite clauses without a clear lexical subordinator, the Accusative with Infinitive construction at least may match the CP-structure of a finite clause; this is the analysis of Sevdali (2007, 169). One argument in support of this is that finite and non-finite clauses can be coordinated, as in (13), in which the pair of coordinating particles *mèn* and *dé* joins the two complement clauses of *eidénai*.

- (13) *egò toínun ... humâs =mèn pántas eidénai*
 I.NOM.SG therefore you.ACC.PL and all.M.ACC.PL know.PRF.INF.ACT
hēgoûmai [hóti egò =mèn orthôs
 think.PRS.IND.MED.1SG that I.NOM.SG and correctly
légō] [toûton =dè hoútō skaiòn eînai]...
 speak.PRS.IND.ACT.1SG this.M.ACC.SG and so stupid.M.ACC.SG be.PRS.INF
 ‘I therefore think that you all know that I speak correctly and that this man is so
 stupid...’

(Ancient Greek. Lysias, *Against Theomnestus* 1.15, from Sevdali (2007, 134))

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

This argument from Sevdali (2007, 134) relies on the assumption that coordination only conjoins units of the same syntactic and categorial status. If we accept this, it follows that finite [*hóti egò mèn orthôs légō*] and non-finite [*toúton dè houtō skaiòn ênai*] are equivalent in syntactic structure. A second argument is from the case of the subordinate subject. As with Latin, subjects of the Ancient Greek Accusative with Infinitive bear accusative case, even when there is no obvious case assigner in the higher clause; these clauses can be the complement of impersonal and passive verbs (Fykias, 2013, §6). If there is a case assigner, it must be something within the non-finite clause, so the null C analysis of the Latin Accusative with Infinitive may apply to Ancient Greek also. Thirdly, there is evidence for the some movement of enclitic pronouns to a C-like position in non-finite clauses; this is discussed in §2.3.5. For these reasons, we have grounds to consider some types of non-finite clause in Ancient Greek as arising from an underlying CP-syntax, like their finite counterparts.

VEDIC SANSKRIT SUBORDINATORS

Vedic Sanskrit noticeably differs from Latin and Ancient Greek. Subordination in a syntactic sense seems lacking in the language; its subordinate finite clauses do not display the syntactic integration available in the previous two languages. They appear at either edge of their main clause, but not embedded within it. The consensus is that this indicates an absence of clausal embedding and of a distinct syntactic status for subordinate clauses.⁵

⁵A philological point to consider is that what we see in Vedic language may not be the full grammatical picture. The Vedas are composed in ritual language that was transmitted for centuries through an oral tradition; such compositions have been observed more widely to prefer ‘additive’ language over subordi-

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

Viti (2007, 50-54) notes several main-clause properties that subordinate clauses do display, such as the free use of evidentials like *añgá* ‘indeed’ and the lack of a *consecutio temporum*, by which the main-clause verb constrains the subordinate-clause verb in terms of its tense and mood. While the relative clause in Latin and Ancient Greek typically appears adjacent to its antecedent noun, its Sanskrit equivalent stands at a distance. Two co-referential (pro)nominal elements are therefore necessary to relate the two clauses to each other.

- (14) *Ágne* [*yáñ* *yajñám* *adhvarám* *viśvátah*
 Agni.M.VOC.SG which.M.ACC.SG sacrifice.M.ACC.SG worship.M.ACC.SG all-sides
paribhūr *ási*] [*sá* *íd* *devéṣu*
 encompassing.M.NOM.SG be.PRS.IND.2SG he.NOM.SG indeed god.M.LOC.PL
gacchati]
 go.PRS.IND.3SG
 ‘Agni, what sacrifice and worship you encompass on all sides, that certainly goes
 among the gods’
 (Sanskrit. RV 1.1.4, from Viti (2007, 64))⁶

(14) is a correlative construction, in which the accusative phrasal object and the nominative subject pronoun are coreferential. The phrase *yajñám adhvarám* is akin to the head noun of an embedded relative clause, being the entity common to both propositions, and for which both clauses offer information. Unlike the embedding grammar of Latin and Ancient Greek, the head noun does not necessarily appear in the main clause; indeed, the order in (14), with the preposed *yá-* clause containing the head noun phrase, is the “diptyque normal” in nation, treating each clause as independent and arranging multiple clauses paratactically (see Lord (1987) and Probert (2015, 14) for discussion).

⁶In the Vedic examples used, *sandhi* between words has mostly been left undissolved, with the exception of final words and words in isolation discussed in the text body.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

Vedic Sanskrit according to Minard (1936). The *yá*-clause may follow the main clause, in which case the head noun appears in the preceding main clause, although this post-position is less frequent and is interpreted as “expressive or rhetorical” (Viti, 2007, 55).

Nonetheless, there is evidence for other indicators of clausal subordination; this evidence is prosodic, lexical and semantic. Prosody offers evidence, since the finite verb in a Vedic subordinate clause is accented, in contrast with the usually unaccented verb of a main clause (Viti, 2007, 25). Lehmann (1974, §4.1.4) states that this was part of an intonation pattern that “indicated incompleteness” and thus anticipated a following clause. Moreover, there are lexical items that indicate a subordinate status, and therefore anticipate an superordinate clause. The stem *yá-* used in (14) “unambiguously signals subordination” (Viti, 2007, 57). Sanskrit displays both inflecting relative pronouns and various indeclinable adverbial subordinators derived from *yá-*. *Yád* has numerous functions, creating causal, final and explicative clauses; *yádi* is principally conditional; *yávat* is temporal; *yáthā* is comparative and purposive (Ramat and Ramat, 1998, 118-120). These elements clearly convey the function of their clause and signal its relationship with another; in doing so, they indicate a lexical and semantic kind of subordination. If a constituent of a clause like *yádi* marks it as the protasis of a conditional construction, that clause expects or requires another clause to be its apodosis. This expectation of the *yádi*-clause is a kind of clausal dependence.

- (15) a. *yádi* *ghṛtébhir* *áhuto* *vásīm*
if ghee.N.INS.PL sacrifice.PRSP.PASS.M.NOM.SG axe.F.ACC.SG

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

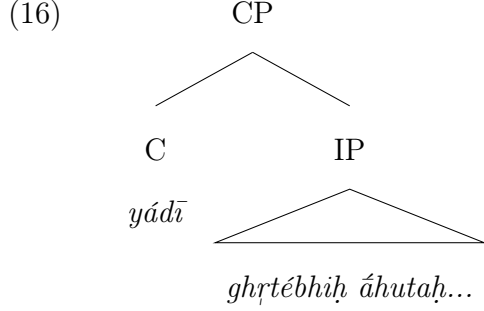
Agnír *bhárata* *úc =ca áva =ca*
 Agni.M.NOM.SG bear.PRS.IND.ACT.3SG up and down and
 ‘If Agni, worshipped with offerings of ghee, bears his axe up and down’
 (Sanskrit. RV 8.19.23)

- b. *yátaḥ* *pra jajñá* *Índro* *=asya*
 whence PV be-born.PRF.IND.MED.3SG Indra.M.NOM.SG it.GEN.SG
veda
 know.PRF.IND.ACT.3SG
 ‘Where he has sprung forth from, Indra knows’
 (Sanskrit. RV 10.73.10, from Viti (2007, 140))

These *yá*-elements offer evidence for the functional category of C. It is unclear though to what extent they have grammaticalised into a distinct syntactic category, akin to the category of C proposed for European languages. Members of this set still decline like *yá*-pronouns, such as *yāvat* (Hettrich, 1988, 564-5), with a correlative partner derived from the stem *tá-* in the main clause. In this, *yá*-elements continue to act like the pronominal stem from which they derive (cf. Pokorny, 1959, 281-6). Some though, like *yádī*, may be already further down the path of grammaticalisation into subordinators.

However we categorise individual *yá*-subordinators, a CP-syntax derives their behaviour. Some are straightforwardly best explained as a C head, while Spec,CP offers a position for those *yá*-elements still best considered to be pronouns or adverbs in syntactic status. Regardless of when we consider indeclinable subordinators to have emerged from *yá-*, they each come to settle in a clause-initial position, even from the earliest sources, as illustrated in (15). There they lose their old nominal inflection and become indeclinable.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C



In the creation of *yá*-subordinators, something attracts these markers of clausal status to the left periphery; this gravity we can explain as due to the pre-existing functional category of C.

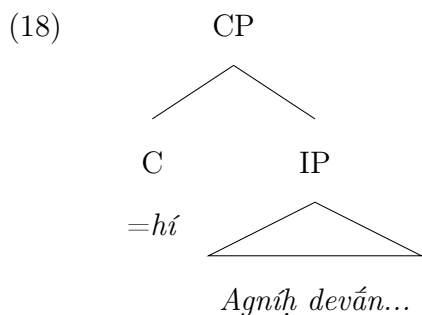
Moreover, older lexical members of the category can even be identified; Vedic Sanskrit makes use of the particle *hí*, which tends to be clause-second. *Hí* is used specifically to emphasise an action or to mark causal clauses. In the latter function it is much more common in the Rigveda than causal *yád*. It declines in frequency in Classical Sanskrit, before disappearing altogether (Viti, 2007, 139).

- (17) *Agnír* = *hí devāñ* *amṛto*
 Agni.M.NOM.SG for god.M.ACC.PL immortal.M.NOM.SG
duvasyáti
 serve.PRS.IND.ACT.3SG
 ‘For immortal Agni serves the gods’
 (Sanskrit. RV 3.3.1, from Viti (2007, 135))

For Hale (2017), the particle *hí* behaves like a C-type element, occupying a high position within its clause. It may have emerged out of a primary function as an emphatic or assertive

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

adverbial particle; since this particle could signal that its clause was the justification or motivation for an adjacent clause. In tandem with the accentuation of the verb, *hí* came to be interpreted as a marker of subordination (Viti, 2007, 134-6). However we explain its diachrony, we can consider *hí* and also the subordinating particles *ca* and *céd* to be “in C itself” (Hale, 2017, 16). The second position of *hí* in the clause is prosodic in origin. While high in the syntax, *hí* has grammaticalised and reduced phonologically to the point that it needs another word to host it. It therefore triggers ‘prosodic inversion’ (Halpern, 1992, 1995). This is a post-syntactic operation, which shifts the enclitic element to follow the first viable host. In (17), this is the noun *Agníḥ*, being accented and therefore the first prosodic word.



By positing a C component to the Sanskrit clause, we therefore account for the observed behaviour of both *hí* and the subordinators derived from the *yá-* stem. Hale’s analysis also accounts for the minority of cases in which *hí* is not the second (orthographic) word overall, such as (19).

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

- (19) *Índro* *vidvām̃* *ánu = hí = tvā*
 Indra.M.NOM.SG know.PRSP.M.NOM.SG PV for you.ACC.SG
cacákṣa...
 look.PRF.IND.ACT.3SG
 ‘Because knowing Indra has looked out for you...’
 (Sanskrit. RV 5.2.8, from Hale (2017, 17))

We can locate the initial phrase *Índro vidvām̃* within a higher position in the syntax than *hí*. This syntactic ‘distance’ is reflected in prosody by a “intonational reset (or pause)” (Hale, 2017, 305), and so the phrase does not offer a viable prosodic host for *hí*. Instead, *hí* (with the enclitic pronoun *tvā*) triggers inversion with the preverb *ánu*. Evidence for this two-layered structure also comes from *yá*-clauses. Examples in which one or more constituents precede the *yá*-subordinator are rare, and of those clauses in which the *yá*-subordinator is not initial, the most common arrangement is for it to follow only one constituent, as exemplified by (20).

- (20) a. *mahī* ***yádi*** *Dhiṣāṇā* *śiśnáthe*
 great.NOM.SG when Dhiṣāṇā.F.NOM.SG piercing.M.LOC.SG
dhāt
 put.AOR.INJ.ACT.3SG
 ‘When the great Dhiṣāṇā had determined to pierce (him)’
 (Sanskrit. RV 3.31.13, from Viti (2007, 117))
- b. *bódhad* ***yán*** = *mā* *háribhyām̃* *kumārāḥ*
 wake.IPRF.INJ.ACT.3SG when I.ACC.SG horse.M.INS.DU prince.M.NOM.SG
sāhadevyāḥ
 sāhadeva.M.NOM.SG
 ‘When the prince, the son of Sāhadeva, woke me with two horses’

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

(Sanskrit. RV 4.15.7, from Viti (2007, 94))

If only one constituent ever seems to precede the *yá*-subordinator, we may locate *mahí* and *bódhad* in a higher discourse-informational position, and *yádi* and *yád* in C. These data are discussed in Chapter 3.

As an aside, while Sanskrit seems unable to embed one finite clause into another, it does not have the same problem with their non-finite counterparts. The plethora of non-finite verbal forms in Sanskrit, including participles, infinitives and absolutes, can have their own arguments; Sanskrit is therefore capable of forming non-finite clauses with full propositional content, and these clauses can be embedded within a higher finite clause as an object or an adjunct. (21a) exhibits a finite clause containing a Locative Absolute. Note that in (21b), the object of the dative noun *ūtáye* ‘for helping’ is *nṝṇ*, accusative like the object of a finite verb, not genitive like a dependent adnominal noun phrase. In (22), the verb *śṝṇómi* has two complex objects, both of which have the argument structure of a full clause, with *tvám* as their common subject.

- (21) a. *yád adyá [súrya udyatí] príyaksatrā*
when today sun.M.LOC.SG rise.PRSP.ACT.M.LOC.SG dear-ruler.M.VOC.PL
ṛtám dadhá
rite.N.ACC.SG set.PRF.IND.ACT.2PL
‘When, at the sun’s rising, you, dear rulers, set the rite’

(Sanskrit. RV 8.27.19, from Ruppel (2012, 1))

- b. *evā Vásiṣṭha Índram [ūtáye nṝṇ]...*
so Vasiṣṭha.M.NOM.SG Indra.M.ACC.SG help.F.DAT.SG man.M.ACC.PL

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

gr̥ṇāti
 praise.PRS.IND.ACT.3SG
 ‘So Vasiṣṭha praises Indra to help men’

(Sanskrit. RV 7.26.5, from Disterheft (1981, 8))

- (22) *evā =hí [tvām ṛtuthā yātáyantam] [maghā*
 so for you.ACC.SG right-season strive.PRSP.M.ACC.SG wealth.N.ACC.PL
víprebhyo dádatam] śṛṇómi
 singer.M.DAT.PL give.PRSP.M.ACC.SG hear.PRS.IND.ACT.1SG
 ‘For indeed I hear that you are striving in due order and giving riches to the singers’

(Sanskrit. RV 5.32.12, from Hock (1982, 43))

As (21) and (22) exemplify, the non-finite clause lacks lexical markers that might offer a diagnostic for its underlying structure and operations like movement. The point here is that Vedic Sanskrit is syntactically capable of complex structures with clausal embedding; this point is picked up in Chapter 5.

One final word merits its own mention, since it goes against much of this account of subordinate clauses and subordinators; this is *íti*, the Sanskrit quotative particle. It is distinct because of its clause-finality; it typically follows what it quotes, as in (23a). An *íti*-quote may appear not only adjacent to but also sometimes within a main clause (23b).

- (23) a. [*hánāma =enāñ íti] Tváṣṭā yád*
 kill.PRS.SBJ.ACT.1PL they.ACC.PL QUOT Tváṣṭā.M.NOM.SG when
ábravīt
 say.IPRF.IND.ACT.3SG
 ‘When Tváṣṭā said “Let us kill them”’

(Sanskrit. RV 1.161.5, from Viti (2007, 225))

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

- b. *yá* [*Índrāya* *sunāvāma* *ity*]
 who.M.NOM.SG Indra.M.DAT.SG press.PRS.SBJ.ACT.1PL QUOT
āha
 say.PRF.IND.ACT.3SG
 ‘Who said “Let us press out (some) for Indra”’
 (Sanskrit. RV 5.37.1, from Hock (1982, 45))
- c. *īti vā īti =me máno* [*gām*
 thus indeed thus I.GEN.SG thought.N.NOM.SG cow.M.ACC.SG
áśvaṃ sanuyāma itī] [*kuvīt sómasya*
 horse.M.ACC.SG win.PRS.OPT.ACT.1SG QUOT surely soma.M.GEN.SG
ápāma itī]?
 drink.AOR.IND.ACT.1SG QUOT
 ‘This, this indeed, is my thinking: may I win a cow and a horse — did I not
 drink some soma?’
 (Sanskrit. RV 10.119.1, from Viti (2007, 235))

The development of the quotative can be seen in the Vedas. *Íti* starts out as an adverb, roughly akin to English *thus*. In the Rigveda, it appears in various positions with respect to the main clause and the quote. Although at first not obligatory, *īti* comes to be used so frequently in between the quote and a main-clause verb of saying that reanalysis occurs; “*īti* changes from being a member of the SPEAK clause to being one of QUOTE” (Hock, 1982, 48).

- (24) Earlier structure: [[QUOTE] *īti* SPEAK]
 Innovative structure: [[QUOTE *īti*] SPEAK]

From the innovative structure in (24), the usage of *īti* then expands, going beyond solely

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

direct-speech quotation, with *īti*-clauses and phrases appearing with verbs and nouns of hearing and thinking, as in (23c). Its expansion into new areas goes hand in hand with its increasing obligatoriness; in later Vedic sources, in terms of its presence and final position, *īti* “is well on its way toward becoming quasi-obligatory” (Hock, 1982, 42). In those later sources, *īti* is so versatile that it is comparable with a declarative subordinator, like English *that*. It is therefore tempting to view *īti* as another subordinator of category C, except one with a head-final syntax. Head-finality is by no means inconceivable for Sanskrit; it is observable across various syntactic categories, such as adpositions.

However, it does not seem to be a subordinator in the syntactic sense, used to embed one clause inside another. *Īti* places no restrictions on the material that it quotes. It can be a complete clause, a shorter phrase or a single word; a declarative statement, an imperative order or a question. The main clause does not seem to place any constraints on the *īti*-material. The ability of *īti* to subordinate anything in its domain, such as questions, means that the *īti*-phrase could itself include a full clause with a subordinator. In other words, if *īti* were a C-element, its quoted material could also have all the properties of a full and separate subordinate clause, with its own C-element. Hence, *īti* is best considered a quotative particle, which is of its own distinct category, and which does not embed its material in the same way as *that*-subordinators do. The lack of restriction on *īti*-material suggests instead that *īti* stands instead immediately outside it in the syntax, perhaps as the head of a categorially distinct ‘Quotative Phrase’.

OLD ENGLISH AND OLD NORSE SUBORDINATORS

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

Old English and Old Norse both attest markers of subordination in their finite clauses. These include *ġif/ef* ‘if’ and *þæt/at* ‘that’, and both have indeclinable markers of relative status, *þe* and *er* respectively. These consistently stand in the first position of the subordinate clause. Yet the two languages differ when it comes to the word order of finite subordinate clauses. In Old English, the finite verb itself is usually late or last, regardless of the type of the clause or the subordinator.

- (25) a. *ġif þū nū ænigne mon cūþe,*
 if you.NOM.SG now any.M.ACC.SG man.M.ACC.SG know.PST.SBJ.2SG
þāra þe hæfde ælcēs þinces
 that.M.GEN.PL that have.PST.SBJ.2SG every.N.GEN.SG thing.N.GEN.SG
ānwald, ælcne weorðscipe hæfde,
 sole-power.N.ACC.SG every.N.ACC.SG honour.M.ACC.SG have.PST.SBJ.2SG
swā forð þæt hē nā māra ne ðorfte...
 so forth that he.NOM.SG no more.N.GEN.SG not need.PST.SBJ.2SG
 ‘If you now knew some man, who held sole power over everything, who had
 all honour, so much that he needed no more...’

(Old English. OE *Boethius* 33)

- b. *ond þæt hē ġeornlīce bihealdan sōolde, þæt*
 and that he.NOM.SG diligently guard.INF should.PST.SBJ.3SG that
hē nōwiht wiðerweardes þære
 he.NOM.SG nothing.N.ACC.SG contrary.N.GEN.SG the.F.DAT.SG
sōðfæstnesse þæs ġelēafan Grēca
 truth.F.DAT.SG the.M.GEN.SG belief.M.GEN.SG Greek.M.GEN.PL
þēawe in Ongolcynnes ċirican inn
 custom.M.DAT.SG in English.N.GEN.SG church.F.ACC.SG in
ġelædde, þe hē ofer bēon sēolde
 bring.PST.SBJ.3SG that he.NOM.SG over be.INF should.PST.SBJ.3SG
 ‘... and that he should diligently make sure that he introduce nothing contrary

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

to the truth of the faith in the manner of the Greeks into the English Church,
which he was to govern’

(Old English. Bede’s *Historia* 4.1)

- c. *ond cwædon, þæt hēo nænigne incan*
and say.PST.IND.3PL that they.NOM.PL no.M.ACC.SG ill-will.M.ACC.SG
tō him wiston, ac hēo ealle
towards he.DAT.SG know.PST.IND.3PL but they.NOM.PL all.M.NOM.PL
him swiðe bliðemōde wæron
he.DAT.SG very kindly-disposed.ST.M.NOM.PL be.PST.IND.3PL
‘... and said that they knew of no ill will towards him, but that they all were
very kindly disposed’

(Old English. Bede’s *Historia* 4.26)

On the basis of clauses such as these, Salvesen and Walkden (2017, 13) conclude that Old English “basically disallows embedded [Verb Second] entirely”, that is, the consistently early position of the verb observed in main clauses (see §3.2.2). The movement of the verb that produces the phenomenon of Verb Second (V2) is therefore presumed to be absent in subordinate clauses; Old English therefore fits well with the analysis proposed (after den Besten (1977)) for Modern German, wherein subordinators and finite verbs compete for the same C-head position in the clausal syntax. If a subordinator is present, C will be lexically overt and the verb must stay low.

Meanwhile, in Old Norse, finite subordinate clauses likewise usually begin with an overt subordinator, yet when one is present, Old Norse maintains the V2 pattern seen in its main clauses. (26) includes clauses of various kinds, both main and subordinate, each with the

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

finite verb second after an unspecified constituent, not counting the initial subordinators *er*, *at*, *þá er* and *fyrr en*, or the conjunction *en* ‘but’. In the case of (26a), we can consider the subordinate clause [*er þeir hyggja, at soðit mun vera*] to be the first constituent of the main clause, while the verb *raufa* is the second.

- (26) a. *en er þeir hyggja, at soðit*
 and when they.NOM.PL think.PRS.IND.3PL that cook.PSTP.N.NOM.SG
mun vera, raufa þeir
 must.PRS.IND.3SG be.INF break-up.PRS.IND.3PL they.NOM.PL
seyðinn
 cooking-fire.DEF.M.ACC.SG
 ‘And as they thought that it must be cooked, they broke up the fire’
 (Old Norse. *Prose Edda* 3.2)

- b. *á heitir Vimur, er*
 river.F.NOM.SG be-called.PRS.IND.3SG Vimur.M.NOM.SG which
Þórr óð, þá er hann
 Thor.M.NOM.SG wade.PST.IND.3SG when that he.NOM.SG
sótti til Geirröðgarða
 seek.PST.IND.3SG to Geirrödgarth.M.GEN.PL
 ‘There is a river called Vimur, which Thor waded through when he was journeying to the halls of Geirröd’
 (Old Norse. *Prose Edda* 3.11)

- c. *þá gerðu jötnar mann á*
 then make.PST.IND.3PL giant.M.NOM.PL man.M.ACC.SG at
Grjóttúnagörðum af leiri ... en ekki fengu
 Grjóttúnagarth.M.DAT.PL from clay.M.DAT.SG but not get.PST.IND.3PL
þeir hjarta svá mikit, at honum
 they.NOM.PL heart.N.ACC.SG so big.ST.N.ACC.SG that he.DAT.SG
sómði, fyrr en þeir tóku úr
 fit.PST.SBJ.3SG, before but they.NOM.PL take.PST.IND.3PL from

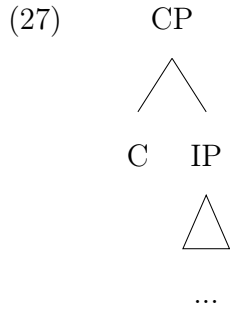
2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

meri *noġkkurri*
mare.F.DAT.SG some.F.DAT.SG
'Then the giants made a man from clay at Grjótúnagarth ... but they could
not find a heart big enough to fit him, until they took one from some mare'
(Old Norse. *Prose Edda* 3.24)

Compared with Old English, the Old Norse finite subordinate clause poses more of a challenge to the common analysis of Verb Second, since it requires an underlying derivation that can produce a surface V2 order, yet it does not offer C as a structural position to meet that end, because C is occupied by the subordinator. In other words, the component parts of the syntax in which this order is based must be below the level of C. It may therefore be the case that when Old Norse generalised the V2 pattern from main to subordinate clauses, the same surface order was produced by other syntactic means, namely the various positions of the IP and VP. This is the analysis of Faarlund; in the presence of a subordinator, “the finite verb stays in the I-position, following the subject, and preceding a sentence adverbial” (Faarlund, 2004, 251).

In sum, we can successfully derive the finite subordinate clause in Old English and Old Norse through a syntactic configuration that contains a C head. This C may be realised by one of the languages’ many markers of subordination (including the relative particles *þe* and *er*), and its lexical realisation has consequences for the position of the finite verb.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C



Turning now to non-finite clauses, both languages also attest constructions with non-finite verbal morphology. Both have infinitives; Old English specifically has two types of infinitive: bare, ending in *-an*, and inflected, ending in *-enne* and used with the word *tō*. Bare infinitives may either precede or follow their main-clause verb, like nominal objects, whereas *tō*-infinitives “invariably” follow, usually positioned at the clausal right edge like finite *þæt*-clauses (Los, 2005, 167).

- (28) a. *ðā ongan hē eft [spellian] and þus*
 then begin.PST.IND.3SG he.NOM.SG again take.INF and thus
cwæð ... gif þū nū sweotole [gecnāwan]
 say.PST.IND.3SG if you.NOM.SG now clearly know.INF
meaht [þā anlīcnesse þāere
 be-able.PRS.IND.2SG the.F,ACC.SG likeness.F.ACC.SG the.F.GEN.SG
sōðan gesælðe]
 true.WK.F.GEN.SG happiness.F.GEN.SG
 ‘Then he began to proclaim again: ... if you now can clearly recognise the
 image of true happiness’
 (Old English. OE *Boethius* 33)
- b. *ða sāde hē Sompeius ... þæt hē*
 then say.PST.IND.3SG he.NOM.SG Sompeius.M.NOM.SG that he.NOM.SG

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

inflected infinitive), Los analyses *tō* as a prefix and a non-finite modal that realises a functional syntactic category like Tense or Mood (Los, 2005, §8.4). This lack of lexical exponents for non-finite C leaves us with movement as the only possible diagnosis for a CP in the Old English non-finite clause. However, while there is evidence for some word-order variation in these non-finite constructions (such as the position of the direct and indirect objects in (28)), clear evidence for movement specifically to a CP is lacking.

Old Norse does have at least one convincing candidate for a non-finite exponent of C. Its infinitive verbs are typically accompanied by *at*. Following Faarlund, we can understand non-finite *at* to be another subordinator and lexical exponent of C; it matches finite-clausal subordinators, such as homophonous *at* ‘that’, in initial position.

- (30) a. *ok þá lofaði hon honum [at drekka af*
 and then let.PST.IND.3SG she.NOM.SG he.DAT.SG to drink.INF from
miðinum þrjá drykki]
 mead.DEF.M.DAT.PL three.F.ACC.PL drink.F.ACC.PL
 ‘And then she allowed him to drink three draughts from the mead’

(Old Norse. *Prose Edda* 3.5)

- b. *ok er þat mitt ráð heldr*
 and be.PRS.IND.3SG that.N.NOM.SG my.N.NOM.SG advice.N.NOM.SG rather
 [*at hætta eigi til þess [at etja hamingju við*
 to risk.INF not to this.N.GEN.SG to contend.INF luck.F.ACC.SG against
Ólaf Haraldsson]]
 Olaf.M.ACC.SG Haraldsson.M.ACC.SG
 ‘And my advice is instead to not dare to contend our luck with Olaf Haralds-
 son’

(Old Norse. *Heimskringla* 7.74)

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

Furthermore, *en* ‘than’ and *nema* ‘except, but’ likewise have an alternative function as non-finite subordinators. Like Old English *tō*, these three stand immediately adjacent to the infinitive. However, following Faarlund’s analysis, this surface adjacency is not indicative of a lower position, but rather of the movement of the infinitive upwards; his diagnostic for this is that the infinitive will precede the negator *eigi* (Faarlund, 2004, 271), as *hætta* does in (30b). Unlike Old Norse finite subordinate clauses, elements of the infinitival clause can be fronted to precede *at* and the verb.

- (31) *at þú skalt eigi kunna [frá tíðendum*
that you.NOM.SG shall.PRS.IND.2SG not know-how.INF about tidings.N.DAT.PL
at segja]
at say.INF
‘... that you shall not be able to tell the tale’
(Old Norse. *Njál’s Saga* 5)

(31) for instance includes a prepositional phrase that is an adjunct to the infinitive *segja*, and that stands initially in its clause. This looks like a case of topicalisation, if we analyse the word *tíðendum* ‘tidings, news’ as referring to events known to both interlocutors in the passage. More than one constituent is allowed in this pre-C position, which is surprising, considering what we have seen so far for main and finite subordinate clauses; Faarlund (2004, 274) states that “this is not just topicalisation of the kind we find in main sentences”, and that “it should rather be analysed as left-adjunction to CP, or possibly raising into the matrix sentence”.

If *at*, *en* and *nema* are of category C, non-finite C does not appear to block fronting and

topicalisation as its finite counterpart seems to. Further research may show that, like Old English *tō*, they are better analysed as lower elements in the syntax. Until such research though, we can continue to follow Faarlund and identify C-elements within the Old Norse infinitival clause, with at least one earlier position available for fronting in the latter type.

OLD CHURCH SLAVONIC SUBORDINATORS

Old Church Slavonic displays a wide array of subordinate clauses, finite and non-finite, marked and unmarked by overt subordinators. The elements *ěko*, *da* and *ašte* introduce finite clauses and have various functions. As a causal or resultative subordinator, which can introduce object clauses and direct or indirect speech (Lunt, 2001, 162), *ěko* closely matches the uses and distribution of Greek *hóti*.

- (32) a. *gl̑jō že vamъ [ěko ni Solomonъ vo v̑sei*
 say.PRS.1SG and you.DAT.PL that not Solomon.M.NOM.SG in all.F.LOC.SG
slavě svoei oblěče sę ěko edinъ otъ
 glory.F.LOC.SG his.F.LOC.SG dress.AOR.3SG self.ACC.SG like one of
sixъ]
 this.M.GEN.PL
 ‘I tell you that not even Solomon in all his glory was dressed like one of these’
 (OCS. *C. Mar.* Lk. 12.27)
- b. *nъ [da ne v̑'sego po ȓdu izglagol̑]*
 but that not all.N.GEN.SG along order.M.DAT.SG pronounce.PRS.1SG
vědě =bo [jako gněvaeši sę]
 know.PRS.1SG for that anger.PRS.2SG self.ACC.SG
 ‘But so that I do not speak about everything in order, for I know that you
 are growing angry’

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

(OCS. *C. Supr.* 1)

The subordinator *da* can be translated as purposive ‘in order to, so that’, negative ‘lest’ and simply ‘that’. *Da* seems limited to irrealis contexts, in which the subordinate proposition is yet to happen or hypothetical. It corresponds to and translates Greek *hína*.

- (33) a. *i reče učnikomъ svoimъ [da estъ pri
and say.AOR.3SG disciple.M.DAT.PL his.M.DAT.PL that be.PRS.3SG by
nemъ ladiica naroda radi] [da ne sьtožajotъ
he.LOC.SG boat.F.NOM.SG crowd.M.GEN.SG sake-of that not press.PRS.3PL
emu]
he.DAT.SG
‘And he told his disciples for there to be a little boat for him because of the
crowd, so that they would not press on him’*

(OCS. *C. Mar.* Mk. 3.9)

- b. *otъ togo =že dьne sьvěštašę [da =i
from that.M.GEN.SG and day.M.GEN.SG consult.AOR.3PL that he.ACC.SG
=bę ubili]
be.CND.3PL kill.RESP.M.NOM.PL
‘And from that day they consulted together to kill him’*

(OCS. *C. Mar.* Jh. 11.53)

Da is often found with conditional verbs, as in (33b) (Lunt, 2001, 161), as is the conditional subordinator *ašte* ‘if’ (34).

- (34) ... *ašte =sę =bi ne rodilъ čkъ
if self.ACC.SG be.CND.3SG not be-born.RESP.M.NOM.SG man.M.NOM.SG
tъ
this.M.NOM.SG*

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

‘... if this man had not been born’

(OCS. *C. Mar.* Mt. 26.24, from Lunt (2001, 160))

These subordinating elements are always clause-initial and there they provide a host for any unaccented clitics within the subordinate clause, such as the subordinate-clause pronouns and auxiliary verbs in (33b) and (34). The specific elements that translate Greek subordinators are all naturally Slavic, and comparative evidence for them from the rest of the family supports the claim that subordinators like *ěko* were a part of the native syntax. All in all, we have the lexical evidence to propose that *ěko*, *da* and *ašte* belonged to a common category of word in Slavic, namely C.

However, our OCS subordinators are very much in thrall to the word-order position of the Greek subordinators that they usually translate; *ašte* in (34) for example matches the position of *ei* in our Greek editions.⁷ This complicates any claims about their basic word order. Nonetheless, the word-order evidence from OCS does not at least contradict any claim that this C stands in a high position of the clausal syntax, just like what has been proposed for the three earlier languages. We can also corroborate a OCS CP-syntax through comparison with the rest of the Slavic family, which has not been influenced by Greek to the same extent; clause-initial subordinators are the norm across Slavic (Sussex and Cubberley, 2006, 350, 374) and there is no reason to think that OCS was substantially different in this regard.

Moreover, Pancheva (2008) makes the good point that nonetheless the Greek influence

⁷E.g. *ei ouk egennēthē ho ánthrōpos ekeînos* ‘if this man had not been born’ (Lk. 12.24).

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

cannot be so strong that our sources do not attest Slavic syntax as it actually existed in the minds of speakers in general, arguing that “whatever syntactic mechanisms ... must have been available in the OCS grammar, as no amount of imitation would have allowed ungrammatical orders” (Pancheva, 2008, 315). This is reasonable, since we know that one of the translators’ primary goals was to make their work comprehensible to a Slavic-speaking audience. This historical fact constrains the potential for unnatural word order, and it gives us a useful principle. However, it does not follow from this that they also took pains to render their translations in a current vernacular; OCS texts are religious, often Biblical, and devout translators would have wanted to remain close to the original and to convey the “sacral stamp” (Drinka, 2011) of the language of texts in that genre. In doing so, they may have exploited the grammatical possibilities of OCS syntax, disregarding everyday usage and pushing the boundaries of what is permissible for the sake of imitation. All this is to say that we may reasonably propose the existence of C in the syntax of OCS, but we must arrive at that conclusion via a roundabout route.

Non-finite clauses are also attested in our OCS sources, built around various participles, infinitives and the supine. As the examples in (35) below demonstrate, non-finite subordinate clauses typically do not have lexically overt subordinators; we therefore have very little to go off in ascertaining their underlying syntax.

- (35) a. *i* [*pristrašъni* *byvъše*] *mъněaxo* [*dxъ*
and terrified.M.NOM.PL be.PSTP.M.NOM.PL think.IPRF.3PL spirit.M.ACC.SG
videšte]
see.PRSP.M.NOM.PL
‘And being terrified, they thought they saw a spirit’

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

(OCS. *C. Mar.* Lk. 24.47, from Huntley (2001, 170))

- b. *idǫ* [*rybъ* *lovitъ*]
go.PRS.1SG fish.F.GEN.PL hunt.SUP
'I am going to catch fish'

(OCS. *C. Mar.* Jh. 21.3, from Lunt (2001, 176))

Danylenko (2019, 153) reports a longstanding idea that the participial subordinate clause is “a true Slavic pattern”, and thus may have been preferred by some translators for translating non-participial verb forms.⁸ Likewise, the Dative Absolute construction expresses “various types of attendant circumstance” (Lunt, 2001, 149) and comprises a dative participle and noun. There has been debate over whether it is another product of external influence on the syntax of OCS, or is an inherited construction. Collins (2011, 106-7) states on the basis of Indo-European-wide evidence that it is inherited, but also adds that, even if it did arise through contact and imitation, the Dative Absolute was “spectacularly successful in ways most other features of borrowed syntax ... were not”.

- (36) *učęštju* *emu* *ljudi* *vъ crkve*
teach.PRSP.M.DAT.SG he.DAT.SG people.M.ACC.PL in temple.F.LOC.SG
'While he was teaching people in the temple ...'

(OCS. *C. Mar.* Lk. 20.1 from Lunt (2001, 149))

Given the analysis for the Latin Ablative Absolute and Accusative with Infinitive above, we can wonder what in the syntax is the source of the dative case in this Slavic construction;

⁸The participle *videšte* in (35a) for one example translates the Greek infinitive *theōreîn* ‘to look at’. (35b) also shows the translator’s preference, since it employs a supine, a morphological category that expresses the purpose of a verb of motion, to translate another Greek infinitive, *halieûein* ‘to fish’.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

it could be evidence for an equivalent null C element. Furthermore, since the position of enclitic pronouns is understood to be a partly syntactic phenomenon (see §2.3.5), how they behave in non-finite OCS clauses may offer a diagnostic for the underlying syntax. These are avenues for further work on the OCS clause. Moreover, taken with the comparative evidence above from Latin, Ancient Greek and Sanskrit, there is a case to be made for reconstructing non-finite clauses back to the proto-language.

OLD IRISH SUBORDINATORS

Old Irish has a variety of elements that indicate the subordinate status of a finite clause. These elements display varied word-order effects, and at least one is lexically null. First to mention is the indeclinable relative particle (*s*)*a*. This is limited in its functions, appearing only in prepositional relative clauses. Old Irish forms these through the fusion of the relevant preposition with the particle, such as *úa* ‘from which...’, *ara* ‘for which...’, *forsa* ‘on which...’. This unit stands initially in its clause.

- (37) a. *di =nd riuth [forsa·robith]*
from the.M.DAT.SG course.VN.M.DAT.SG on-which-be.PRS.SBJ.2SG
‘From the course on which you may be’
(Old Irish. Wb. 20b1)
- b. *a folad [dia·n-immoln̄gaithaer*
the.N.NOM.SG substance.N.NOM.SG from-which-produce.PRS.IND.PASS.3SG
vox]
vox.F.NOM.SG
‘The substance from which *vox* is produced’
(Old Irish. Sg. 3a5)

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

As (37) shows, these prepositional relative units belong to the traditional grouping of ‘conjunct particles’. This heterogeneous group are defined by the form of the finite verb that they immediately precede; the verb appears stressed on its first preverb particle (if compound) or syllable (if simple) (Stifter, 2009, 87). See §2.2.3 for an overview of Old Irish verbal behaviour.

Out of these prepositional relative clauses, new uses as adverbial and complement clauses developed. *Ara* ‘for which’ for instance also has a purposive function of ‘so that’ or simply ‘that’, while *dia* ‘from which’ has a conditional function of ‘if’.

- (38) a. *gigeste=si* *Dia* *linn* [*ara·fulsam* *ar*
 pray.FUT.2PL-you(PL) God.M.ACC.SG with.1PL that·bear.PRS.SBJ.1PL our
fochidi]
 trial.VN.F.ACC.PL
 ‘You will ask God that we may bear our trials’

(Old Irish. Wb. 14c2)

- b. *ni·epeir=som* [*ara·ndíltad* *lasuidib*]
 not-say.PRS.IND.3SG-he that·deny.IPRF.IND.3SG with.3PL
 ‘He does not say that he denied with them’

(Old Irish. Wb. 17d11)

- c. [*dia·techtat* *briathra* *in go ar a* *cúl*]
 if·have.PRS.IND.3PL verb.F.ACC.PL in *go* for their back.M.DAT.SG
 ‘If they have verbs in *go* behind them’

(Old Irish. Sg. 112b40)

Likewise, the subordinating conjunct *co* ‘so that, that, until’ appears to derive from the preposition *co* ‘to’, although *co* does not include the particle (*s*)*a*.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

- (39) a. [*co·mbiat* *fo deod*]
 so-that.be.FUT.3PL at end.N.ACC.SG
 ‘So that they will be at the end’
 (Old Irish. Sg. 212a35)
- b. [*co·mtis* *ainmmnidi* *a triur*]
 that.be.PST.SBJ.3PL nominative.M.NOM.PL their three.M.DAT.SG
 ‘That the three of them would be nominative’
 (Old Irish. Sg. 7b2)

The subordinating elements *(s)a* and *co* can be straightforwardly understood to belong to a category of C (cf. Carnie et al., 1994). This stands high in the clausal syntax, above the position of the verb. Whether we analyse the more complex units *ara* and *dia* likewise as single C-elements is unclear; the respective syntactic positions of *(s)a* and the preposition may be different. If we wish to include the *wh*-movement of the preposition from elsewhere in the clause, they may remain separate, with only *(s)a* in C. Alternatively, they may have fused to form a new class of elements and exponents of C. Either way, they occupy the topmost parts of the clausal syntax.

Other subordinators in Old Irish display different word-order behaviour and effects on the verb, instead preceding ‘independent’ forms of the verb (Stifter, 2009, 107-9) (see §2.2.3). These include *má/ma* ‘if’, *(h)óre/húare* ‘because’ and *cía/cia* ‘although’.

- (40) a. *ma nudub-feil* *i n-ellug* *coirp*
 if PV-you(PL).be.PRS.IND.3SG in union.VN.M.DAT.SG body.M.GEN.SG
Crist
 Christ.M.GEN.SG
 ‘If you are in the union of Christ’s body’

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

(Old Irish. Wb. 19c20)

- b. *hóre at-tá in n-ar leid*
 because be.PRS.IND.3SG in our side.N.DAT.SG
 ‘Because he is on our side’

(Old Irish. Wb. 4b11)

- c. *is Dia=som dom=sa [hóre no-cretim*
 be.PRS.IND.3SG God.M.NOM.SG-he to.1SG-me because believe.PRS.IND.3SG
Ísu]
 Jesus.M.ACC.SG
 ‘He is God to me, because I believe in Jesus’

(Old Irish. Wb. 1a2)

Others introduce a clause with a relative verb, as *amal* ‘as, as if’ and *in tain* ‘when, since’ tend to. (*H*)*óre/húare* may do so also.

- (41) a. *amal ro-mbói fáilte dúib re= n-a galar=som*
 as be.PRF.3SG.REL joy.F.NOM.SG to.3PL before his illness.M.DAT.SG-he
 ‘As joy has been yours before his sickness’

(Old Irish. Wb. 23d5)

- b. *amal rondob-carsam=ni*
 as PV-you(PL)-love.PRF.1PL.REL-we
 ‘As we have loved you’

(Old Irish. Wb. 25a35)

- c. *in-tain bes mithich*
 when be.PRS.SBJ.3SG.REL timely
 ‘When it may be time’

(Old Irish. Wb. 20c15)

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

The relative feature of subordinate verbs like those in (41) permits us to consider such examples to be cleft-like constructions, formed through two clauses and a silent copula. In other words, (41c) really has the structure ‘*when it is [that it may be time]*’. The subordinator stands somewhere in the higher clause, while the verb occupies the lower relative clause.

Subordinate clauses like the three in (40) pose more of a challenge to derive. The verb is not relative, so we have no grounds to claim that they comprise more than one clause. Because, as is argued in §2.2.3, independent-form verbs are the result of the interactions between the verb and a null C head, subordinators like *má* and *amal* cannot be in C, elsewise the verb would appear in its dependent form. The object pronoun *dub* in (40a) would moreover be enclitic on *ma*. They must therefore be somewhere higher in the syntactic structure. *Má* and *cía* must therefore stand somewhere in a higher position than C; such a syntactic component is discussed in Chapter 3.

Moreover, Old Irish relative clauses give us evidence for a lexically null element of the category C. Specifically, the behaviour of subject and object relative clauses support the proposal of the null C-element “=Ø^{L/N}” (Bate, 2024). Though null, it still makes its presence felt through its effects on the finite verb. The first effect is dedicated relative verbal endings. These are available only to simple verbs and for only some grammatical persons; for active verbs, these are third-person singular *-as/-es*, first-person plural *-mae/-me* and third-person plural *-tae/-te*.⁹

⁹The 3sg. relative ending *-as* is harder to derive than *-mae* and *-tae*, and has been explained as the result of analogy or fusion with the relative copula (cf. Ahlqvist, 1985).

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

- (42) a. *in fualascach bec [ásas*
 the.M.NOM.SG bush.M.NOM.SG small.M.NOM.SG grow.PRS.IND.3SG.REL
as in chrunn]
 from the tree.N.DAT.SG
 ‘The small branches that grow from the tree’
 (Old Irish. Sg. 65a28)
- b. *iss ed a folad cétnae*
 be.PRS.IND.3SG it the.N.NOM.SG substance.N.NOM.SG same.N.NOM.SG
 [*sluinditae*]
 signify.PRS.IND.3PL.REL
 ‘This is the same substance that they signify’
 (Old Irish. Sg. 9b6)

For verbs that do not meet these restrictions, such as compound verbs and simple verbs with a different person, Old Irish utilises another relativising technique: leniting and nasalising relative clauses. This involves the mutation¹⁰ of the initial sound of the second component of the verbal complex, which may be a preverb or the first syllable of the verbal root.

- (43) a. *ní ail aicsu*
 be-not.PRS.IND.ACT.3SG wish.M.NOM.SG wish.VN.F.NOM.SG
for=sa n-i [no·thechti]
 on-the.N.ACC.SG PRT possess.PRS.IND.ACT.2SG.REL
 ‘It is not a wish to wish for that what one has’
 (Sg. 148a9)
- b. *amal nahí [nád·chiat] ... amal*
 like that-which.NOM.PL not·weep.PRS.IND.3PL.REL like

¹⁰Lenition ‘softens’ an initial consonant, turning voiceless and voiced plosives into voiceless and voiced fricatives respectively. Nasalisation originally added a homo-organic nasal consonant to the word, although this later developed into a voicing effect too. Neither is consistently indicated in the written record.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

nahí [nad·**ch**renat]
 that-which.NOM.PL not·buy.PRS.IND.ACT.3PL.REL
 ‘Like those who do not weep ... like those who do not buy’

(Wb. 10b6-7)

c. *ethemlagas* [do·**n**-adbat] *híc* *o=nd*
 etymology.F.NOM.SG show.PRS.IND.ACT.3SG.REL here from-the.M.DAT.SG
sun *greccdu*
 word.M.DAT.SG Greek.DAT.SG
 ‘An etymology which he shows here from the Greek word’

(Sg. 27b15)

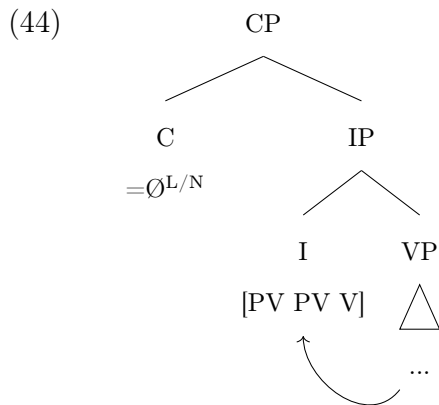
The relative clauses in (43) show the two processes. In (43a) and (43b), the root of the verb occupies the second position, and its first consonant is lenited. In (43c), a nasal *n*- is added to *ad*, the second preverb of the compound verb *do-adbat* ‘to show’. Lenition and nasalisation originally served separate grammatical purposes; a leniting relative clause was used for when the antecedent of the clause was its subject, while a nasalising clause was for object antecedents. By the time of our Old Irish sources, however, this distinction was breaking down, as leniting relative clauses took over the functions of the nasalising construction (Stifter, 2009, 107).¹¹

The account of Bate (2024) is that this behaviour of subject/object relative clauses is again the product of the lexically null element of category C, =Ø^{L/N}. Being enclitic, it triggers the phonological process of prosodic inversion (cf. Halpern (1992, 1995)), mentioned for Sanskrit above. Through this, clitics will surface after the first available host in the

¹¹Neuter nouns, moreover, were always modified by a leniting relative clause, regardless of the grammatical role of the neuter antecedent (McCone, 1980, 17).

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

linear order. This will be the next-highest lexical element in the syntax, such as the first preverb of a compound verb, which has moved up to a position in the IP layer. Its enclisis on the first preverb creates a prosodic break within compound verbs, and also mutates the initial sound of what immediately follows; this produces leniting and nasalising relative clauses. (44) illustrates this derivation for clauses with a relativised compound verb.



Being a phonological process, the suitability of a host is phonologically determined. In response to the view that Old Irish preverbs are proclitic (cf. Thurneysen (1946, 30) and McCone (1997, 4-5)), Bate (2024) suggests a redefinition of prosodic inversion in Old Irish, operating around syllables, not accented words. Since nearly all preverbs in Old Irish are monosyllabic, the clitic will intervene between the first preverb and the remainder of a compound verb. Meanwhile, for simple verbs, there is the option of the movement of the verb to the high C position, where it bears =Ø^{L/N} as an affix; this produces relative inflection.

As with many aspects of Old Irish grammar, this behaviour is a product of phonological

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

decay, with older morphology gradually disappearing, and the phonological impressions that it leaves behind being ‘morphologised’ in turn to pick up the slack. In the philological tradition, both relativised simple verbs and lenited compounds are derived through the idea of a lost relative particle **yo(m)* (Thurneysen, 1946; McCone, 1980, 364). This pre-Old Irish subordinator is responsible through affixation for the relative inflections, and its clisis on the first preverb of a compound verb created the necessary contexts for relative mutations to develop.¹² The nominative form was **yo*, hence the continued use of the lenition effect of its final for creating subject relative clauses. This C-particle must have remained capable of minimal inflection until the pre-Old Irish stage, because object relative clauses look to be a product of its (masculine and feminine) accusative counterpart, **yom* (McCone, 1980; Ahlqvist, 1985). The final nasal consonant of accusative **yom*, being enclitic on the first preverb, nasalised the first sound of the rest of the verb; through this, nasalising relative clauses like (43c) could develop.

This relative particle is cognate with the Sanskrit relative stem *yá-* (Matasovic, 2009, 436). We can therefore posit that the *pre-pre-Old Irish* starting point for the development of $=\emptyset^{L/N}$ was as a declinable relative pronoun. As in Sanskrit (see §2.3.1), this pronoun was located in the syntax in Spec,CP. By the immediately pre-Old Irish stage, it had undergone some grammaticalisation and Spec-to-Head reanalysis (van Gelderen, 2004) to become the particle **yo(m)*. Eska’s analysis (2012, 54) is that **yo(m)* by the time of Pre-Old Irish had become an element of the category C, and had to be either enclitic on

¹²Independent evidence from Gaulish supports the existence of such a relative particle in a earlier stage of Irish and across Celtic in general, namely the *io* on the verb *dugiiontio* ‘who serve’ (cf. Watkins, 1969). Note that *io* is clause-second in this Gaulish example.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

or suffixed to another component of the clause, namely the verb or its first preverb. It continued to decay phonologically, becoming null = $\emptyset^{L/N}$, though not without leaving a morphophonological mark. The C-element that it realised continues its presence in the morphophonology through what were once only the sandhi effects of **yo(m)*.

A further possible null C may be present in cases of embedded declarative clauses, in which the subordinate clause is marked simply by the nasal mutation of the initial verb. No overt marker of subordination appears to trigger the mutation (Ó hUiginn, 1998).

- (45) a. *arna ro·chretea [mbias icc*
 so-that-not PV-believe.PRS.SBJ.3SG be.FUT.3SG.REL salvation.VN.F.ACC.SG
do hua Dia]
 to.3SG from God.M.DAT.SG
 ‘So that he does not believe that he will have salvation from God’

(Old Irish. Ml. 127a7, from Ó hUiginn (1998, 123))

- b. *con·nic [do·mberthar forcell]*
 can.PRS.IND.3SG give.PRS.IND.3SG.REL testimony.VN.N.NOM.SG
 ‘It can be that a testimony is given’

(Old Irish. Ml. 24d14, from Ó hUiginn (1998, 123))

This construction looks again like a product of the old accusative relative particle **yom*. In the synchronic syntax of Old Irish, this construction may arise from an underlying C head that creates subordinate declarative, not relative, clauses.

In sum, Old Irish provides a variety of evidence for the syntactic category C across different eras of the Irish language. This includes the overt elements (*s*)*a* and *co*, but also the null = $\emptyset^{L/N}$, which builds relative clauses, and which continues the pre-Old Irish relative

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

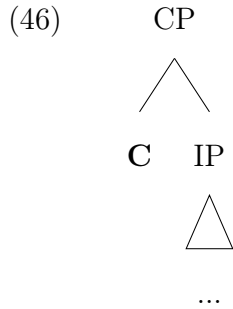
particle **yo(m)*. Other subordinating elements and phrases (such as *ma* and (*h*)*óre*) suggest further components to the left periphery, beyond a simple CP. What exactly they can tell us is discussed in Chapter 3.

Lastly, it should be noted that Old Irish additionally has non-finite verbal morphology and constructions. A given verb can form a past passive particle and a verbal form of necessity, the latter akin to the Latin gerundive, which are used with finite verbs such as the copula (Thurneysen, 1946, 441-4). Perhaps the most common and most productive of the non-finite forms is the verbal noun. These are diverse in their derivation, yet united in their behaviour; they act like abstract nouns, in that they have nominal inflection, may be governed by a preposition and may have dependent genitive nouns, possessive pronouns and prepositional phrases. Verbal-noun constructions tell us little though about their CP layer, if they even have one. Many can be straightforwardly treated as noun phrases. Their productive use for expressing subordinate propositions may be a sign of their reanalysis as clauses, but there is nothing in the way of subordinators or word-order variation that can tell us about the presence and features of a non-finite CP.

INTERMEDIATE SUMMARY: THE EVIDENCE FROM SUBORDINATORS

Across the seven languages considered, a picture emerges of a common syntactic category that is stable in both its presence and its behaviour. Various kinds of subordinate finite clause attest to a C-component of the underlying syntax, one very early in the linear order, and therefore high in the hierarchical structure and forming its own left-headed CP.

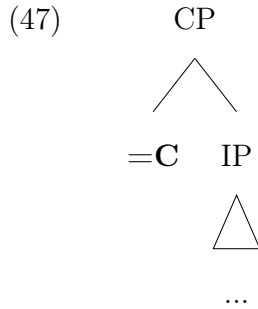
2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C



To this functional category belong diverse lexical items like relative particles (e.g. Old English *þe* and Old Irish *(s)a*) and conditional subordinators (e.g. Latin *sī* and Ancient Greek *ei*).

Even in Vedic Sanskrit, in which the common *yá-* elements language do not uncontroversially belong to the category C, they still appear consistently early in the clause. Moreover, the Vedic Sanskrit particle *hí* seems to be an early exponent of C; its stable word-order behaviour suits its analysis as an exponent of C, while its weak prosodic status and second position may result from its grammaticalisation, having changed from an emphatic particle into a marker of subordination. The syntax of both Sanskrit *hí* and Old Irish $=\emptyset^{L/N}$ works just as it does in (46); the key difference is in their phonological features, namely their enclitic status, which modifies the order that the syntax generates.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C



The enclitic feature of C in (47) will cause what realises the C (be it Old Irish $=\emptyset^{L/N}$ or Sanskrit hi) to surface after the next suitable host; in Old Irish, this is often a preverb particle belonging to the verb in the lower IP.

The evidence for a corresponding C in non-finite subordinate clauses is much weaker. There are possible cases, such as the possible case-assigning null C of the Accusative with Infinitive construction in Latin and Ancient Greek. All things considered, we can at least conclude that the abstract syntactic category of C is part of the syntax of early Indo-European subordinate finite clauses. From a comparative perspective, this Indo-European behaviour is too consistent to be put down as a series of unrelated coincidences. On this basis, we have plentiful evidence and good grounds to follow the ‘majority rule’ of traditional linguistic reconstruction (see §1.2.1) and conclude that subordinate C must have been present in the common ancestor. However, before attempting that reconstruction, let us consider further lexical evidence.

2.2.2 INTERROGATIVE ELEMENTS IN EARLY INDO-EUROPEAN

Subordinators are not the only lexical indicators of the kind of the clause that we find across the seven Indo-European languages. Though they appear with much less consistency, there are also lexical items that mark a main interrogative clause, specifically a polar question that expects a *yes/no* answer. This section argues that these elements in early Indo-European should again be derived through a functional category high in the underlying syntax. This component is furthermore identified as the same as that in (46), namely C.

LATIN INTERROGATIVE PARTICLES

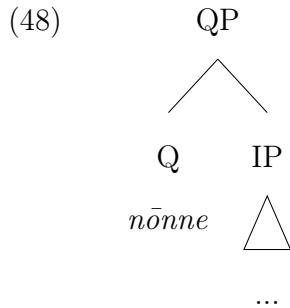
In Latin, for one, dedicated lexical elements are employed to express an interrogative status and communicate a request for information; these include *nōnne*, *num*, *utrum*, *an* and enclitic *ne*. They are not strictly necessary to form polar questions, but without them, only context can help us to identify such questions, since our sources lack punctuation and written indications for interrogative intonation.¹³ The traditional explanation for their use is that they differed according to what answer was expected (i.e. affirmative or negative), although the reality of their distribution is more complicated (Pinkster, 2015, 316).

In terms of their word order, *nōnne*, *num* and *utrum* are frequently clause-initial, while enclitic *ne* likewise appears early (Pinkster, 2015, 316-334). On this word-order basis, we should consider Latin's interrogative particles to be exponents of a stable interrogative

¹³We should therefore bear in mind the possibility of their exaggerated use in written contexts, yet their consistent presence across writers and genres suggests that they were part of the spoken language too.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

syntactic head. The syntax in (48) will produce a question with the question particle *nōnne* in initial position.



The syntax in (48) can even successfully derive enclitic *ne*, with the addition of the post-syntactic process of prosodic inversion. Unlike *nōnne*, the common particle *ne* cannot be initial in the clause, but nonetheless is “regularly attached to the first word of the sentence or clause to emphasize it” (Pinkster, 2021, 999).

- (49) a. *sed potest =ne rērum maior esse*
 but can.PRS.IND.3SG Q thing.F.GEN.PL greater.F.NOM.SG be.INF
dissēnsiō?
 disagreement.F.NOM.SG
 ‘But can there be a greater difference of opinions?’
 (Latin. Cicero, *On Ends* 3.13)
- b. *quem =ne ego servāvī in campīs*
 who.M.ACC.SG Q I.NOM.SG save.PRF.IND.ACT.1SG in field.M.ABL.PL
Curculionīs?
 Gorgonidonian.M.ABL.PL
 ‘The one I saved in the Gorgonidonian fields?’
 (Latin. Plautus, *Vainglorious Soldier* 1.1)

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

In (49), *ne* follows the first word of its clause.¹⁴ This very common second position of *ne* can be derived either through the movement of another constituent up to the functional projection of which *ne* is the head, or through the prosodic inversion of enclitic *ne* with the next word in the post-syntactic order.

This combined syntacto-prosodic analysis can derive clauses in which *ne* comes later. In (50b) and (50a), it is attached to the verbs *licet* and *videntur*, the sixth and fourth words of their clauses.

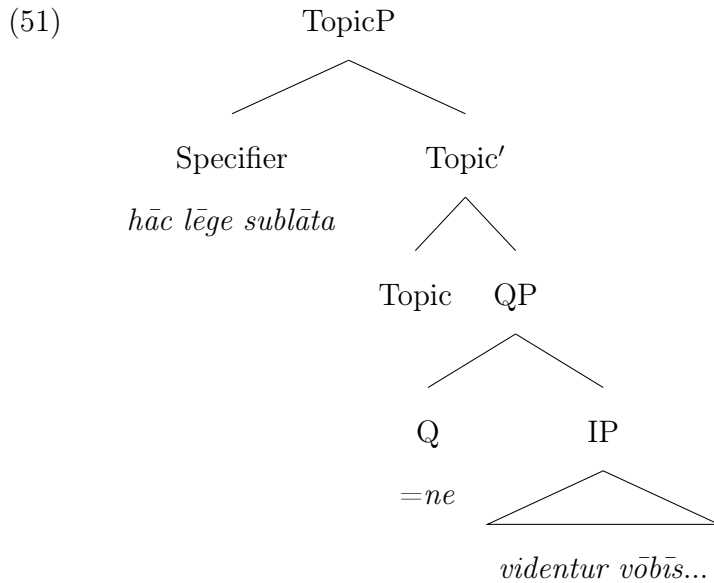
- (50) a. *dē reliquīs rei publicae malīs*
of remaining.F.ABL.PL thing.F.GEN.SG public.F.GEN.SG evil.N.ABL.PL
licet = **ne** *dīcere*?
be-allowed.PRS.IND.3SG Q say.PRS.INF.ACT
‘Is it permitted to talk about the remaining evils of the republic?’
(Latin. Cicero, *Philippics* 1.6, from Pinkster (2021, 999))
- b. *hāc lēge sublāta videntur*
this.F.ABL.SG law.F.ABL.SG abolish.PRFP.F.ABL.SG see.PRS.IND.PASS.3PL
= **ne** *vōbīs posse Caesaris ācta*
Q you.DAT.PL be-possible.PRS.INF Caesar.M.GEN.SG act.N.ACC.PL
servārī?
save.PRS.INF.PASS
‘With this law abolished, do Caesar’s acts seem to you able to be maintained?’
(Latin. Cicero, *Philippics* 1.9, from Pinkster (2021, 1000))

These later positions for *ne* do not contradict a high underlying position. In (50b) and (50a), [*hāc lēge sublāta*] and [*dē reliquīs rei publicae malīs*] each make up a single con-

¹⁴Excluding the clausal conjunction *sed* ‘but’. The working assumption throughout the thesis is these stand outside the syntax of the clause proper.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

stituent, respectively an Ablative Absolute and a prepositional phrase. Seen this way, *ne* seems to respect syntactic constituency, and its position is not purely determined by prosodic requirements. In cases like (50a), the topical phrase *dē reliquīs reī pūblicaē malīs* we can locate in a topical projection even higher than *ne* (see Chapter 3). Being enclitic, *ne* requires another element to bear it; this is done through prosodic inversion with the following verb *videntur*.



Ne inverts with *videntur* and not the initial topical phrase perhaps because such topics were intonational phrases separated from the remainder of the clause by a brief pause. This prosodic behaviour of *ne* also explains why it is “almost never” attached to prepositions, standing instead with another element of the prepositional phrase, such as the dependent noun that follows the preposition (Pinkster, 2021, 1000); if Latin prepositions are proclitic,

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

they do not provide a viable host for *ne*.

Even as new interrogative particles develop, they gravitate towards the same early position in the clause; Pinkster (2015, 333) notes that *utrum* comes to mark an interrogative simple clause from the time of Cicero onwards, and it too tends to stand initially.

- (52) *utrum* = *enim in clārissimīs est cīvibus*
Q for in clearest.M.ABL.PL be.PRS.IND.3SG citizen.M.ABL.PL
is...?
he.NOM.SG
'For is he among the most illustrious citizens?'
(Latin. Cicero, *Pro Flacco* 19.45)

Whether this underlying head is solely interrogative in function is another matter, since elements like *ne* have other functions; *ne* also appears in exclamatory clauses (Bennett, 1910, 423) like (53).

- (53) *hunci =ne sōlem tam nigrum surrēxe*
this.M.ACC.SG Q sun.M.ACC.SG so unlucky.M.ACC.SG rise.PRF.INF.ACT
mihī
I.DAT.SG
'How ill-omened this sun rose for me!'
(Latin. Horace, *Sermones* 1.9)

This broader function of *ne* is key to the identification of its category as the same as that of subordinators, discussed in the conclusion to this subsection.

ANCIENT GREEK INTERROGATIVE PARTICLES

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

As in Latin, polar questions in Ancient Greek may be difficult to identify in the written record, which also marks neither interrogative intonation nor interrogative punctuation until a later date. However, the questions in (54) demonstrate how Ancient Greek employs particles for indicating interrogativity, most notably *ê* and its derivative *âra*, although we should again acknowledge that their frequency may be in part a literary mechanism.

- (54) a. *ê téthnēke* *Pólubos?*
 Q die.PRF.IND.ACT.3SG Polybus.M.NOM.SG
 ‘Is Polybus dead?’
 (Ancient Greek. Sophocles, *Oedipus Rex* 943, from Smyth (1920, §2650))
- b. *âr’ Odusséōs* *klúō?*
 Q Odysseus.M.GEN.SG hear.PRS.IND.ACT.3SG
 ‘Do I hear Odysseus?’
 (Ancient Greek. Sophocles, *Philoctetes* 976, from Smyth (1920, §2650))
- c. *kaī toúto* *poiōnti* *âra áxiōn* = *soi*
 and this.N.ACC.SG do.PRSP.M.DAT.SG Q worthy.N.NOM.SG you.DAT.SG
zên *éstoi.* *ê plēsiáseis*
 live.PRS.INF.ACT be.FUT.IND.3SG Q approach.FUT.IND.ACT.2SG
toútois...?
 this.DAT.PL
 ‘And by doing this, will your life be worth living? Will you approach them...?’
 (Ancient Greek. Plato, *Crito* 53c)

These two on their own anticipate no particular answer, but may do so in compound phrases (such as *âr’ ou’*, which anticipates affirmation). Moreover, *âra* can also appear in subordinate clauses, functioning as an interrogative subordinator, akin to English *whether*.

Both of these elements have a fairly strict clause-early position, preceding all material

in (54), with the exception of the anaphoric phrase in (54c). It is therefore very plausible to view \hat{e} and \hat{ara} as two lexical exponents of an underlying functional head, situated above the lexical domain; this is the hypothesis of Arad and Roussou (1997), who identify \hat{e} and \hat{ara} as exponents of the topmost head in the clause, labelled Force by Rizzi (1997). The function of Force is clause-typing; the functional head that \hat{e} and \hat{ara} represent sets the clause as ‘interrogative’. Ancient Greek polar questions therefore can be derived with the same functional structure proposed for Latin in (48), containing a high “QP”.¹⁵

SANSKRIT INTERROGATIVE PARTICLES

Polar questions are rare in the earliest Sanskrit sources; of those few identified, they either show no difference from declarative statements, or display *pluta* (lengthened) vowels in the last syllable of the question (Clackson, 2007, 160), which is likely written evidence for marking interrogativity through intonation in the spoken language. However, later Sanskrit tends to use indeclinable particles like *ápi*, *kím*, *kaccít*, *utá* and *svid* (Speyer, 1886, §412). These words each have numerous uses and form an etymologically disparate group; *kím* is also the neuter nominative/accusative singular form of the interrogative *ká-* stem, *utá* means ‘and, also’ in Vedic Sanskrit, while *ápi* can be a preposition meaning ‘on’ and can emphasise an adjacent noun, akin to English *also* and *even*. It is through a distinct word order that this diverse set of words create polar questions; in this function, they tend to stand first in the clause.

¹⁵Similarly, the common particle *án* can likewise be treated as C-element, though it is not limited to interrogative contexts. It is an ‘irrealis’ particle, indicating the contingency or potentiality of its clause.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

- (55) *ápi śivám bhavataḥ?*
Q welfare.N.ACC.SG be.PRS.IND.2DU
'Are you in good health?'

(Sanskrit. *Panchatantra* 35, from Speyer (1886, §412))

Negative elements like *ná* and *nánu* can also form questions that expect affirmation by standing initially (Macdonell 1916, 237; Hackstein 2016, 14).

What we therefore see in Sanskrit is that in the development of interrogative particles, a left-peripheral position seems to have offered a means to convey a distinct interrogative function, as in Latin and Ancient Greek. This surface position corresponds to a high functional projection in the syntax, analogous to Latin and Ancient Greek “QP” proposed previously. Sanskrit interrogative particles may stand in the same structural position as moved *wh*-elements of the *ká*- stem (see §2.3.1), although their initial position appears to be stricter; *wh*-words can follow fronted material (see §3.2.1), but examples of fronting to before *ápi* or other particles are not forthcoming. Given their diverse origins and multiple meanings, it may be that interrogative particles do occupy the same structural position as *wh*-words, but that the option of fronting other material even higher than them is not utilised, to avoid ambiguity and keep their interrogative function clear.

THE OLD CHURCH SLAVONIC INTERROGATIVE PARTICLE

The Old Church Slavonic particle *li* serves to create polar questions. It behaves much like the clausal conjunctions *bo* and *že*, being most often enclitic on a clause-initial word, although Večerka (1989, 44) notes that it occupies the second position with less consistency.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

- (56) a. *dostoitъ* = *li dati kinъsъ Kesarevi ili ni.*
 be-proper.PRS.3SG Q give.INF tax.M.ACC.SG Caesar.M.DAT.SG or not
damъ = li ili ne damъ?
 give.PRS.1PL Q or not give.PRS.1PL
 ‘Is it right to give tax to Caesar or not? Do I give or not give?’
 (OCS. *C. Zogr.* Mk. 12.14)
- b. *xošteši = li ubo da šedъše isplēvemъ*
 want.PRS.2SG Q then that go.PSTP.M.NOM.PL gather.PRS.1PL
ję?
 they.ACC.PL
 ‘Do you then want us to go and round them up?’
 (OCS. *C. Mar.* Mt. 13.28)
- c. *slyšiši = li filisofe rěčъ siju?*
 hear.PRS.2SG Q philosopher.M.VOC.SG speech.F.ACC.SG this.F.ACC.SG
 ‘Do you hear, oh philosopher, this request?’
 (OCS. *Life of Methodius* 5)
- d. *ne i mytare = li tožde tvoreťъ?*
 not and tax-collector.M.NOM.PL Q also do.PRS.3PL
 ‘Do not even tax-collectors do likewise?’
 (OCS. *C. Mar.* Mt. 5.46)

Examples (56a) to (56c) each contain polar questions and *li* stands in second position. The exception is (56d), in which *li* appears to be the fourth element of its clause. However, it may be that *ne* and *i* are in fact proclitic on *mytare*, together forming the first prosodic unit of the clause, on which *li* may then depend. Alternatively, *ne i mytare* may instead be a topical phrase, one that stands higher than *li* in the underlying syntax, and so is not interrupted by it through prosodic inversion. Either of these derivations would conform

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

to the second-position rule of OCS clitics, while also matching the order of the original Greek.¹⁶

In none of these four examples does *li* correspond to a Greek counterpart; it is a Slavic feature, and exists to this day in languages like Russian, Czech and Bulgarian (Sussex and Cubberley, 2006, 317). There are grounds therefore to propose again that *li* is an exponent of the native syntax, specifically of a hierarchically high functional category. Yet, being enclitic like Latin *ne*, *li* triggers prosodic inversion and thus surfaces after the next highest constituent of the clause or the first word of it. This functional component of the syntax may go all the way back to Proto-Slavic; this would explain the distribution of *li* across the Slavic languages, as well as non-enclitic and therefore clause-initial interrogative particles like Polish *czy*.

OLD IRISH INTERROGATIVE PARTICLES

The particle for forming positive interrogative clauses in Old Irish is *in*, and it has the negative counterparts *in-nád* and *ca-ni*. All three are part of the traditional grouping of conjunct particles, meaning that they precede the verb (be it simple or compound) and that the verb appears with initial-syllable stress and with its shorter conjunct inflection.

- (57) a. *in·bértar* *epistli* *uaín...?*
Q-carry.FUT.PASS.IND.3PL epistle.M.NOM.PL from.1PL
'Will epistles be brought from us?'

(Old Irish. Wb. 15a3)

¹⁶Mt. 5.46 “*Oukhì kai hoi telônai...*”

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

- b. *in·mesor=sa?*
Q-judge.FUT.ACT.1SG-I
'Will I judge?'

(Old Irish. Sg. 179a2)
- c. *innád·cualaid=si* *a geinti?*
QNEG-hear.PRF.ACT.2PL-you(PL) oh gentile.M.VOC.PL
'Have you not heard, O gentiles?'

(Old Irish. Wb. 5a21)
- d. *cani·epir?*
QNEG-say.PRS.IND.ACT.3SG
'Does it not say?'

(Old Irish. Wb. 10d5)

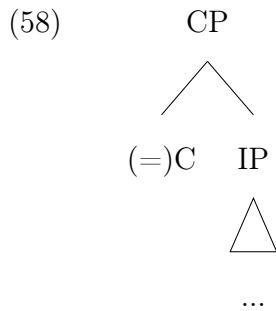
Crucially, their integration into the overall 'verbal complex' and interaction with the verb mean that Old Irish interrogative particles behave exactly like certain subordinators (see §2.2.1 on Old Irish). We therefore have grounds to derive these interrogative particles from a similar structural position in the syntax; this is the analysis of Bate (2024).

INTERMEDIATE SUMMARY: THE EVIDENCE FROM INTERROGATIVE PARTICLES

Five of the Indo-European languages reviewed make use of indeclinable elements to mark their clause as a polar question. These particles, like subordinators, tend to occur very early in the clause; Latin *nōnne/num*, Sanskrit *ápi* and Old Irish *in* are consistently first, while Latin *ne* and Old Church Slavonic *li* can be analysed as initial as far as the syntax is concerned, with their linear order modified by the requirements of enclisis.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

In the languages with more than one interrogative particle, the set may be etymologically very diverse, yet united in word-order behaviour. Out of various origins, these particular elements end up in a clause-initial position in the linear order. This is reminiscent of the development of *yá*-subordinators in Vedic Sanskrit (§2.2.1). Given the word-order commonalities between interrogative particles and subordinators, and their complementary distribution, it is most theoretically parsimonious to collapse the two broad groups of elements into one. This is to say, rather than separate CP and QP, subordinators and interrogative particles arise from one common functional category, which we may continue to label C.



In doing so, the defining quality of this functional category must broaden beyond only subordination. Since it is present in the syntax of both subordinate clauses and interrogative main clauses, it is responsible for setting the type, status and illocutionary force of these finite clauses.

2.2.3 DECLARATIVE ELEMENTS IN EARLY INDO-EUROPEAN

Before moving on to different sources of evidence for C, a section is necessary on elements that mark their clause as a declarative main one. Such elements have been identified in Latin; Pinkster (2015, 309) understands clause-initial *ne* ‘indeed’ and *equidem* ‘indeed’ as providing “assertive illocutionary force”. Such diverse elements are by no means obligatory though, and most declarative main clauses in Latin contain no element that can be said to perform an assertive function. The lack of declarative-marking elements across early Indo-European is not too surprising, if we assume that this type of clause is somehow the default, while other types utilise clause-type markers in opposition to it. Yet at least one obligatory declarative element can be detected in the language set, namely the Old Irish null element = \emptyset .

As with the Old Irish relative clauses in (43) in §2.2.1, the Old Irish declarative main clause typically begins with the verbal complex, preceding any nominal arguments of the verb. This complex contains conjunct particles (e.g. *ní* ‘not’, *co* ‘so that’), preverbs (e.g. *dí* ‘of’) and the finite verb, which are arranged to meet strict requirements for the position of stress and the hosting of infixes (see §2.3.5). The abstract surface order of the complex is as follows:

- (59) [X]·[Y Z W]
(Adger, 2006, 610)

The components of the complex are placed into these positions, following certain restric-

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

tions. For one, [X] must be filled. If a conjunct particle is present, it takes precedence and occupies [X]. If not, the first preverb particle present stands in [X]. In the absence of both conjunct particles and preverbs, the simple verb itself stands initially, and is not divided into two. (60) illustrates this chain of precedence:

- (60) a. [*ní-derscigem*] *nech* *di alailiu*
 not-surpass.PRS.IND.1PL someone.NOM.SG to other.M.DAT.SG
 ‘We do not surpass each other’
 (Old Irish. Wb. 2a14)
- b. [*co-mbad*] *etarcnad* *doib*
 ‘so-that-be.PST.SBJ.3SG known.N.NOM.SG to.3PL
 ‘so that it may be known to them’
 (Old Irish. Wb. 26d12)
- c. [*do-rigénsat*] *in* *discipuil* *dechor*
 make.PRF.3PL the.M.NOM.PL disciple.M.NOM.PL distinction.N.ACC.SG
 ‘The disciples have made a distinction’
 (Old Irish. Wb. 7d10)
- d. [*do-beir=som*] *ainm* *bráthre* *doib*
 give.PRS.IND.3SG-he name.N.ACC.SG brother.M.GEN.PL to.3PL
 ‘He gives the name of brothers to them’
 (Old Irish. Wb. 7d8)
- e. [*caraid*] *cesin a macdán*
 love.PRS.IND.3SG himself his childlike-art.M.ACC.SG
 ‘He loves his childlike art’
 (Old Irish. *Pangur Bán*)

No matter what constitutes it, the [Y Z W] domain acts like a typical Old Irish phonological

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

word, since stress is always allocated to the syllable in its initial [Y] slot. Meanwhile, whatever stands in [X] is cut off from and unaffected by these phonological processes by an apparent prosodic break (McCone, 1997, 4).

If it is the first preverb that fills [X], this leaves stress to be borne by the second element of the compound verb, be it the verbal root or another preverb. This arrangement produces the *deuterotonic* forms of the verb.¹⁷ Examples (60c) and (60d) include deuterotonic forms, with the first preverb of the compound verb in [X] and the second element bearing the stress after the prosodic break. However, if a conjunct particle in [X] confines a compound verb entirely to the [Y Z W] domain, the first preverb instead bears the stress in the [Y] position, and so the verb appears in its *prototonic* forms¹⁸, as in (60a). Meanwhile, if a simple verb stands in [X], it appears with the longer forms of ‘absolute’ inflection. If it stands in any other position in the complex or the clause, the verb bears ‘conjunct’ inflection. Because they follow conjunct particles, simple verbs with conjunct inflection and prototonic compound verbs are known as the *dependent* forms of a verb. In the absence of a conjunct particle, simple verbs with absolute inflection and deuterotonic compound verbs are the *independent* forms.

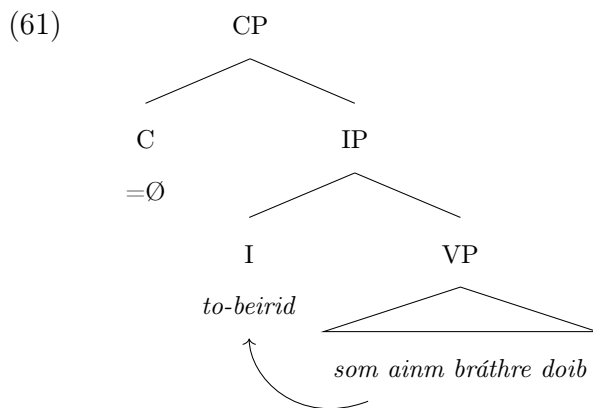
Inspired by Adger’s (2006, 632) enclitic Force head, the analysis of Bate (2024) and this thesis is that this behaviour is the product of interactions with the verb and $=\emptyset$, another C-type element, which in declarative main clauses is lexically null and enclitic. As with relative $=\emptyset^{L/N}$, this null $=\emptyset$ triggers prosodic inversion with preverbs and creates the

¹⁷From Ancient Greek *deúteros* ‘second’ and *tónos* ‘accent’.

¹⁸From Ancient Greek *prōtos* ‘first’

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

prosodic break within deuterotonic verbs. In (60d), the preverb *do* occupies the pretonic [X] position, because it is separated from the remainder of the verb and the prosodic word by the enclisis of =Ø. This is a post-syntactic operation; in the syntax, =Ø is higher than the compound verb in the IP.



In the case of simple verbs, these undergo syntactic movement to C; through this the null C head surfaces as absolute inflection.

This null C-element has strong philological support, specifically from the established Cowgill particle (Cowgill, 1975). This is the pre-Old Irish enclitic element **eti*.¹⁹ The Cowgill particle accounts for the independent-form verbs in declarative clauses; it produced both simple verbs with absolute inflection and deuterotonic compound verbs with their prosodic break. Adopting Eska's diachrony (2012, 54-55), we can identify the origins

¹⁹The exact shape of **eti* and its origins have been contested. Cowgill himself posits **es* (**s* after vowels) as its form, yet Schrijver (1994) and Schumacher et al. (2004) have more recently reconstructed it as **eti*. Thurneysen (1946) and Kim (2000) derive the particle from the copula verb **esti* 'it is'. Schrijver (1994) and Eska (2012) instead trace it back to a connective particle **eti*, which would introduce and conjoin whole clauses.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

of $=\emptyset$ in the pre-Old Irish connective particle **eti*. This over time diminished in its strength and shape, becoming an enclitic; a key step was that **eti* was reanalysed as an exponent of C. It developed into a generalised “affirmative particle”, marking any clause as a declarative, non-subordinate one. It motivated the movement of the simple verb to C to bear it as a declarative-typing affix. In time, its fusion with the simple verb produced absolute inflection, which continues the association with declarativity.²⁰ In the case of preverbs, these too once used to bear enclitic **eti* (Kortlandt, 2007, 2), moving up to a high functional projection, such as the CP, to do so. The short element itself continued its phonological decay; it became a lexically null item, only reflected phonologically in the prosodic break in deuterotonic compound verbs by the time of Old Irish. A stronger connection between preverbs and their verb meant that the syntactic movement of preverbs to C gave way to a non-syntactic derivation through prosodic inversion.²¹

Old Irish thus gives evidence for a lost particle whose function it was to mark its clause as a declarative main one. Given the consistency with which its effects (independent verb forms) appear in Old Irish, we may conclude that the Cowgill particle was an obligatory part of the language at a prehistoric point, having developed into a head of category C. Similar to relative $=\emptyset^{L/N}$, Old Irish $=\emptyset$ signals the declarative type and superordinate status of its clause.

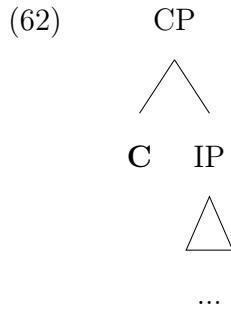
²⁰More specifically, its effect on the simple verb was “protecting the ending from atrophy” (Russell, 2014, 53).

²¹The V-to-C movement of simple verbs in declarative and relative clauses was by this point a syntactic outlier, which may offer a reason for its decline in Irish since.

2.2. LEXICAL ELEMENTS OF THE SYNTACTIC CATEGORY C

2.2.4 SUMMARY: THE LEXICAL EVIDENCE FOR C

The seven early Indo-European languages reviewed each contribute evidence for the claim that the abstract syntactic category of C was a common feature of Indo-European syntax. Its presence and clause-early position are very consistent, at least in finite clauses (less so in non-finite ones), especially if we do not limit C to markers of subordination, but rather expand it to include clause-typing elements in general. The syntactic left periphery of early Indo-European clauses can be said to have in common at least the following abstract structure:



This consistency is to such a degree that it seems reasonable to reconstruct a common origin for this feature of Indo-European languages. Even in cases in which the grammaticalisation of subordinators is still underway in the historical record (cf. the Sanskrit *yá-*stem), lexical elements that indicate subordination stand in the clausal left periphery.²² This repeated convergence on the left periphery lends support to the claim that C was

²²Even if we do not consider *yá*-elements as belonging to the functional category of C, the element *hí* is a clear candidate for membership.

2.3. SYNTACTIC MOVEMENT TO CP

present in the inherited syntax. We can bolster this proto-C if we have first expanded its definition to involve clause typing in general. C may be present in all early Indo-European (finite) clauses, and it is towards this position that new lexical markers of clause type and sub-/superordinate status are attracted, be they declarative, interrogative, relative, subordinating, etc. elements. Once there, they may undergo grammaticalisation to become exponents of the category C.

To reconstruct a functional category of C present in the Proto-Indo-European subordinate clause, specifically within its left-peripheral syntax, therefore accounts not only for the common behaviour of subordinators and clause-typing particles across the descendent languages, but also for the similar paths of development that certain lexical elements follow, as we see with the Sanskrit *yá-* stem and Old Irish $=\emptyset^{L/N}$ and $=\emptyset$. To strengthen that reconstruction though, there is a separate source of evidence to consider: the evidence from movement.

2.3 SYNTACTIC MOVEMENT TO CP

As mentioned in Chapter 1, the syntactic operation of movement is considered to be an additional feature of the syntax of Indo-European languages, and a further means by which we may identify underlying functional heads. Movement can therefore offer more diagnostic support for C in the left periphery; this section demonstrates that support, beginning with *wh*-movement in questions and relative clauses.

2.3.1 WH-MOVEMENT IN EARLY INDO-EUROPEAN

The view that the distinct word order of questions in certain languages is the product of movement to a structurally high position is well established in generative grammar (cf. Ross, 1967; Chomsky, 1977; Bresnan, 1970). In a language like English, question words like *what*, *who* and *when* are thought to have a *wh*-feature, which triggers their movement up to a functional node, thereby disrupting the basic arrangement of verbal arguments and adjuncts. *Wh*-movement has also been proposed in the derivation of English relative clauses too, since they too display the same patterns of word order with the same lexical items, excluding relative clauses with *that* (cf. Chomsky, 1977). All of this theoretical analysis may be applied to historical languages, likewise on the basis of word order.

WH-MOVEMENT IN LATIN

In Latin, we can identify the movement of its *wh*-elements (e.g. *quis* ‘who’, *quid* ‘what’) to the left periphery, since its *qu*-words precede the rest of their clause with overwhelming frequency (Pinkster, 2021, 982-3), regardless of their syntactic category. They may be pronominal (63a) or adverbial (63b) in status, but they crucially do not maintain the typical positions of their non-interrogative counterparts. The *qu*- stem can also form determiners within a noun phrase, as in (63c).

- (63) a. *quid* *is* *dīcit?*
 what.N.NOM.SG he.NOM.SG say.PRS.IND.ACT.3SG
 ‘What does he say?’

2.3. SYNTACTIC MOVEMENT TO CP

(Latin. Cicero, *For Quintus Roscius the Actor* 42, from Oniga (2014, 278))

- b. *ubi illum quaeram gentium?*
 where that.M.ACC.SG seek.PRS.SBJ.ACT.1SG people.F.GEN.PL
 ‘Where in the world am I to look for him?’

(Latin. Plautus, *Epidicus* 5.2)

- c. *quī = enim cantus moderatā*
 which.M.NOM.SG for song.M.NOM.SG moderate.PRF.P.F.ABL.SG
oratiōne dulcior inveniri
 oratory.F.ABL.SG sweeter.M.NOM.SG find.PRS.INF.PASS
potest?
 be-able.PRS.IND.3SG
 ‘For what song can be found sweeter than restrained oratory?’

(Latin. Cicero, *Orator* 2.34)

The same *qu-* stem also has a relative function, and relative material matches interrogative words and phrases in their early position in the clause (Pinkster, 2021, 978).

- (64) *Ubi sē diūtius dūcī intelligit et*
 When self.ACC.SG long lead.PRS.INF.PASS perceive.PRF.IND.ACT.3SG and
diem instāre quō diē
 day.M.NOM.SG near.PRS.INF.ACT which.M.ABL.SG day.M.ABL.SG
frūmentum militibus mētiri
 grain.N.ACC.SG soldier.M.ABL.PL measure.PRS.INF.ACT
oporteret, convocātis eōrum
 be-proper.IPRF.SBJ.ACT.3SG convoke.PRF.P.M.ABL.PL they.GEN.PL
principibus, quōrum magnam cōpiam in
 chief.M.ABL.PL which.M.GEN.PL great.F.ACC.SG plenty.F.ACC.SG in
castris habēbat, in hīs Diviciacō
 camp.M.ABL.PL have.IPRF.IND.ACT.3SG in this.N.ABL.PL Divitiacus.M.ABL.SG
et Liscō, quī summō magistratū
 and Liscus.M.ABL.SG who.M.NOM.SG highest.M.DAT.SG magistracy.M.DAT.SG

2.3. SYNTACTIC MOVEMENT TO CP

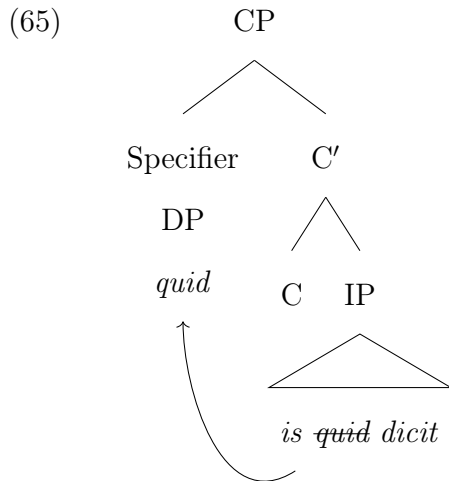
praeerat, *quem* *‘Vergobretum’*
 be-before.PPRF.IND.3SG who.M.ACC.SG Vergobretus.M.ACC.SG
appellant *Haeduī,* *quī*
 call.PRS.IND.ACT.3PL Aedui.M.NOM.PL who.M.NOM.SG
creātur *annuus et vitāe* *necisque* *in*
 create.PRS.IND.PASS.3SG annually and life.F.GEN.SG death.F.GEN.SG-and in
suōs *habet* *potestātem,* *graviter eōs*
 own.M.ACC.PL have.PRS.IND.ACT.3SG power.F.ACC.SG gravely they.ACC.PL
accūsāt
 reproach.PRS.IND.ACT.3SG

‘When he realised that he was put off too long, and that the day was near on which he had to mete out the corn to his soldiers, having summoned their chiefs, of whom he had a great number in the camp, among them Divitiacus and Liscus, who held the chief magistracy, whom the Aedui call the *Vergobretus*, and who is appointed annually and has power of life or death over his own people, he reprimanded them harshly.’

(Latin. Caesar, *Gallic War* 1.16)

As with modern languages, the preference for this far-leftward position for Latin’s *qu-*material has been analysed as the result of movement from a more basic position to the specifier of CP (*inter alios* Du Toit, 1987; Ram-Prasad, 2022). Since no one would posit that the Object-Subject-Verb order in (63a) is basic, we can instead consider the object *quid* to have moved up at least past the subject, the pronoun *is*.²³

²³The basic position of *ubi* ‘where’ may even be identifiable in (63b). In this, *ubi* appears at a distance from its intensifying partner word, *gentium* (literally ‘of peoples’). If we first assume that *ubi gentium* are base-generated as a phrase, meaning ‘where in the world’, we can derive (63b) through the movement of *ubi* alone to Spec,CP, thereby stranding *gentium*.



In (64), the relative elements likewise appear first in each clause, regardless of their function or more usual position; something in the syntax attracts them to the clausal edge. Salvi (2005) notes though that the rigidity of this initial position does vary across time, since in earlier Latin, prior to the classical language in (64), one or even two constituents sometimes precede the *qu*-word. This he takes to be a “a relic of the archaic age where the movement of constituents to the front of an embedded clause was freer”.

The clause-early position for *qu*-material in interrogative clauses also serves to distinguish between different interpretations of the same lexical item; the interrogative determiner *quī* ‘which’ and interrogative adverbs like *quandō* ‘when’ can also express an indefinite function (i.e. ‘something’, ‘some time’). When so used, they cannot occupy the clause-initial position (Pinkster, 2021, 983). Although not a unified syntactic class and not base-generated in a high position, this distinction in the interpretation of *qu*-words in

2.3. SYNTACTIC MOVEMENT TO CP

Latin offers further evidence for the CP within the syntax of finite clauses, to which they are attracted only when they express interrogativity or relativity.

WH-MOVEMENT IN ANCIENT GREEK

What pertains to Latin also applies to Ancient Greek interrogative material, such as pronominal *tis* ‘who’ or adverbial *poû* ‘where’. This typically occurs in the initial position of its clause. This is a “strong tendency” (Goldstein, 2015, 71), regardless of the syntactic category of the interrogative item.

- (66) a. *tí* = *d’ éstin* *ô geraié?*
 what.N.NOM.SG and be.PRS.IND.3SG oh old.M.VOC.SG
 ‘What is it, old man?’
 (Ancient Greek. Euripides, *Medea* 49)
- b. *poîon* = *ti kindúneuma?* *poû*
 what-kind.N.NOM.SG some.N.NOM.SG danger.N.NOM.SG where
gnómēs pot’ ei?
 thinking.F.GEN.SG ever be.PRS.IND.2SG
 ‘What kind of danger? Where on earth is your thinking?’
 (Ancient Greek. Sophocles, *Antigone* 42)

Ancient Greek also makes use of finite relative clauses. The usual relative pronoun is declinable *hós*. This, fused with the suffix *-per* or the indefinite pronoun *tis*, forms the emphatic and indefinite relative pronouns *hósper* and *hóstis*.

- (67) a. *hōs =dè tōi andri [hōn àn*
 that but the.M.DAT.SG man.M.DAT.SG who.M.ACC.SG PRT

2.3. SYNTACTIC MOVEMENT TO CP

However, there is a minority of questions in which the *wh*-word comes later; in (68), the *wh*-word follows two contrasted nouns.

- (68) *puretòs* = *dè tí* *poieî?* ... *keramìs*
 fever.M.NOM.SG and what.N.ACC.SG do.PRS.IND.ACT.3SG tile.F.NOM.SG
 = *dè tí* *poieî?*
 and what.N.ACC.SG do.PRS.IND.ACT.3SG
 ‘And a fever, what does it do? ... And a roof tile, what does it do?’
 (Ancient Greek. Epictetus, *Discourses* 4)

Interrogative material in Ancient Greek does not have the right to occupy the clause-initial position, only an early one. Fraser’s (non-generative) view (1999) is that the “regular position” of interrogatives is what he labels “P2”, which is the second stressed constituent of the clause; “they appear first in the clause only by being emphasised” (Fraser, 1999, 13). Likewise, the movement of relative pronouns or adverbs targets a very high position in the structure, but not the highest, since rare examples of constituents preceding the relative element within the relative clause are attested (69).

- (69) *autàr egò* *theós* *eimi* [*diamperès* ***hḗ***
 but I.NOM.SG god.M.NOM.SG be.PRS.IND.3SG continuously who.F.NOM.SG
 = *se* *phulássō* *en pántessi* *pónois*]
 you.ACC.SG guard.PRS.IND.ACT.1SG in all.M.DAT.PL toil.M.DAT.PL
 ‘But I am a god, who guards you unfailingly in all your struggles’
 (Ancient Greek. Homer, *Odyssey* 20.47-8)

In (69), the adverb *diamperés*, a dependent of the subordinate verb *phulássō*, precedes the relative pronoun *hḗ*. Since this is Homeric Greek, composed in dactylic hexameter,

2.3. SYNTACTIC MOVEMENT TO CP

this unusual word order may be simply due to metrical patterns, yet this also could be an archaic grammar, in which a higher position is available in relative clauses for salient constituents. This position is explored in detail in Chapter 3; for now, suffice to say, it does not contradict the derivation of Ancient Greek relative clauses through *wh*-movement to CP, as with Latin.

A final point of note is that, as with Latin *quis*, the cognate *tís* in Ancient Greek has the unaccented and indefinite counterpart *tis* ‘someone, something’. Given their near-identity, word order must convey the intended interpretation; while interrogative *tís* is consistently clause-initial, indefinite *tis* is an enclitic on another word, and so cannot stand first.

- (70) a. *kaì tí án =tís sunnoôito, ô*
 and what.N.ACC.SG PRT someone.M.NOM.SG ponder.PRS.OPT.MED.3SG oh
Sôkrates?
 Socrates.M.VOC.SG
 ‘And what might someone ponder, Socrates?’

(Ancient Greek. Plato, *Alcibiades* 2.138)

- b. *epeí =tís érkato tôn*
 when someone.M.NOM.SG begin.AOR.IND.MED.3SG the.M.GEN.PL
katelēluthótōn mnēsikakeîn
 return.PRFP.M.GEN.PL bear-malice.PRS.INF.ACT
 ‘When someone began to bear malice again those returning from exile...’

(Ancient Greek. Aristotle, *Athenian Constitution* 40.2)

Since *tis* is enclitic, its word order may be (partly) determined by prosody; clitics in Ancient Greek display the second-position clustering behaviour of Wackernagel’s law (1892), discussed in §2.3.5. This is to say, where we find *tis* may not reflect a basic syntactic

2.3. SYNTACTIC MOVEMENT TO CP

position for the pronoun, but rather results from the syntactic and phonological operations discussed in §2.3.5 (cf. Taylor, 1990).

WH-MOVEMENT IN VEDIC SANSKRIT

Wh-elements (such as *káḥ* ‘who, which’, *kím* ‘what, why’ and *kútra* ‘where’) appear first or almost first in their clause as a rule, which may be taken as evidence of *wh*-movement. Hale (1987a, 39) reports that 90% of interrogative elements, built on the stem *ká-*, are clause-initial in the Rigveda.

- (71) a. ***kás*** =*te* *mātáraṃ* *vidhāvām*
who.MF.NOM.SG you.GEN.SG mother.F.ACC.SG widow.F.ACC.SG
acakrat?
do.PPRF.IND.ACT.3SG
‘Who made your mother a widow?’
(Sanskrit. RV 4.18.12, from Hale (1987a, 2))
- b. ***káḥ*** *svid vṛkṣó* *ní ṣṭhito*
who.M.NOM.SG Q tree.M.NOM.SG PV stand.PRFP.M.NOM.SG
mádhye *árṇasaḥ?*
middle.N.LOC.SG water.N.GEN.PL
‘Which tree then, standing in the midst of the waters...?’
(Sanskrit. RV 1.182.7, from Macdonell (1916, 251))
- c. ***kadā*** =*naḥ* *śuśravad* *gíra* *Índro*
when we.GEN hear.PRF.SBJ.ACT.3SG prayer.M.ACC.PL Indra.M.NOM.SG
aṅgá?
PRT ‘When might Indra have truly heard our prayers?’
(Sanskrit. RV 1.84.8, from Macdonell (1916, 354))

2.3. SYNTACTIC MOVEMENT TO CP

Wh-movement therefore seems to be a productive means of creating questions at this early stage of Sanskrit, as with Latin and Ancient Greek. It is not necessarily accompanied by verb movement to a distinct position; the verb continues to display the word-order patterns in declarative clauses.

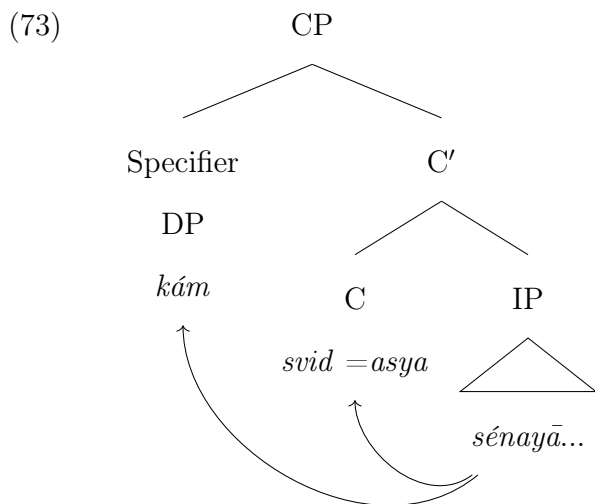
Example (71b) merits attention, since the *wh*-word *kaḍā* co-occurs with *svid*, “a particle of interrogation or inquiry or doubt” (Monier-Williams, 1899, 1284). This is not a lone example of such co-occurrence.

- (72) *kám* *u* *ṣvid* = *asya* *sénayā* ... *Pañím*
which.M.ACC.SG now Q he.GEN.SG army.F.INS.SG Pani.M.ACC.SG
góṣu *starāmahe?*
cow.M.LOC.PL fight.AOR.SBJ.MED.1PL
‘Which of the Pani might we kill with his army for the cattle?’
(Sanskrit. RV 8.75.7)

Several analyses offer themselves for this arrangement. One is that *svid* forms a phrase with the *ká*-element, modifying its meaning into that of an indefinite interrogative (i.e. *whoever*). Another is that *svid* is an enclitic particle, and forms a clitic chain with the pronoun *asya* dependent on the initial *ká*-word. Yet another possibility is that *svid* and other interrogative particles in such constructions are exponents of C, and that Sanskrit therefore allows *wh*-movement into the CP, even when the C head is lexically overt.²⁴

²⁴This derivation could also be applied to Ancient Greek and the post-interrogative particle *án* in (70a).

2.3. SYNTACTIC MOVEMENT TO CP



In light of the argument in §2.2.1 for a broadly defined C, identifying *svid* as another exponent of C is appealing, but there is nothing to decide between these different analyses of questions like (71b) and (72).

Relative clauses have likewise been treated as instances of *wh*-movement in Sanskrit (e.g. Ram-Prasad, 2022, 48). The typical relativising lexical item is the stem *yá-*, which forms a variety of relative pronouns, determiners and adverbs. For Hale (1987a, 2018), the target of movement within both the relative and interrogative clause is again a high C position; this analysis is reasonable, given that relative elements consistently appear first or second in the clause; Hale (1987a, 43) reports that *yá-* stands in first position in over 350 relative clauses in a sample of over 450 in the Rígvēda. It also frequently appears separated from the rest of its noun phrase, as in (74). As mentioned in §2.2.1, the *yá*-clause altogether stands at either the left or right edge of another, together forming a correlative

2.3. SYNTACTIC MOVEMENT TO CP

construction.

- (74) a. *yó* =*naḥ* *Pūṣann* *aghó*
 which.M.NOM.SG we.ACC.PL Pusan.M.VOC.SG bad.M.NOM.SG
vṛko *duḥśéva* *ādídēśati,* *ápa*
 wolf.M.NOM.SG inauspicious.M.NOM.SG threaten.PRS.SBJ.ACT.3SG away
sma tám *pathó* *jahi*
 verily he.ACC.SG path.M.ABL.SG kill.IMP.ACT.2SG
 ‘The wicked, sinister wolf who threatens us, chase him, Pusan, from the path’

(Sanskrit. RV 1.42.2)

- b. *yád* *aṅgá* *dāśúṣe* *tvám* *Ágne*
 which.N.ACC.SG indeed worshipping.M.DAT.SG you.NOM.SG Agni.M.VOC.SG
bhadráṃ *kariṣyási* *táva* *ít* *tát*
 blessing.N.ACC.SG do.FUT.IND.ACT.2SG you.GEN.SG indeed that.N.NOM.SG
satyám
 truth.N.NOM.SG
 ‘Which(ever) blessing you, Agni, will grant your worshipper, that is your truth’

(Sanskrit. RV 1.1.6)

Lowe (2014, 9-11) and Ram-Prasad (2022, 10-12) note a subtle but important difference in the distribution of interrogative and relative elements. Both may be preceded by another fronted constituent of their clause, yet this is more common with relative markers, which are occasionally preceded by more than one constituent. Relative elements always follow any preverb particles that appear in the left periphery (75).

- (75) *prá yé* *śúmbhante* *jánayo* *ná*
 PV which.M.NOM.PL shine.PRS.IND.MED.3PL woman.F.NOM.PL like
 ‘Who beautify themselves like women...’

2.3. SYNTACTIC MOVEMENT TO CP

(Sanskrit. RV 1.85.1, from Lowe (2014, 27))

In contrast, interrogative elements never follow preverbs or more than one fronted constituent. To account for this difference, Ram-Prasad (2022) has argued that interrogative and relative material in Sanskrit, while both moved, occupy different positions in the syntactic left periphery. The difference and Ram-Prasad’s account are explored in detail and rejected in §3.3.3, in favour of a unitary CP.

WH-MOVEMENT IN OLD ENGLISH AND OLD NORSE

Like their modern descendant languages, Old English and Old Norse express interrogativity by placing the verb early in its clause, typically either first (in case of polar questions) or second (for *wh*-questions). If the latter, the accompanying *wh*-word or phrase stands in the initial position, immediately before the verb, as in (76). Neither language allows for intervening constituents and a V3 order, although this is permitted in other early Germanic languages, such as Old High German (Walkden, 2014, 92).

- (76) a. *hwæt* *is* *sāwla* *hāelo* *būte*
 what.N.NOM.SG be.PRS.IND.3SG soul.F.GEN.SG health.F.NOM.SG unless
rihtwīsnes? *oððe hwæt* *is* *hiora*
 righteousness.F.NOM.SG or what.N.NOM.SG be.PRS.IND.3SG she.GEN.SG
untrymnes *būte unþēawas?* *hwā* *is*
 sickness.F.NOM.SG unless vice.M.NOM.PL who.MF.NOM.SG be.PRS.IND.3SG
þonne betera *lāce* *þære* *sāwle*
 then better.WK.M.NOM.SG doctor.M.NOM.SG the.F.GEN.SG soul.F.GEN.SG
þonne sē *þe hī* *gesceōp,* *þæt*
 than he.NOM.SG that she.ACC.SG create.PST.IND.3SG that.N.NOM.SG

2.3. SYNTACTIC MOVEMENT TO CP

is *God?*
 be.PRS.IND.3SG God.M.NOM.SG
 ‘What is a soul’s health if not righteousness? Or what is its sickness, if not
 vices? Who is a better doctor for the soul than he who created it, namely
 God?’

(Old English. OE *Boethius* 39)

- b. *hversu lengi ætlar* *þú* *hér at vera?* ... *hvert*
 how long intend.PRS.IND.2SG you.NOM.SG here to be.INF whither
ætlar *þú* *þá?*
 intend.PRS.IND.2SG you.NOM.SG then
 ‘How long do you intend to be here? ... Where will you go next?’

(Old Norse. *Njáls Saga* 5)

Early Germanic questions are well accounted for with the syntax proposed for their modern descendants: namely, a high CP, containing a null C head that both attracts the finite verb, and triggers *wh*-movement towards a specifier position for *wh*-words and phrases. If no such material is present in the clause, only verbal movement occurs, producing a verb-initial order and a polar question. This movement into the CP is natural, if we assume first that the C is the locus of clause type in the Old English and Old Norse clause; if it is set to an interrogative type, it motivates *wh*-movement to express this feature.

As mentioned in §2.2.1, Old English and Old Norse have the indeclinable relative particles (*þe* and *er*) that form relative clauses. These typically stand initially in the clause, yet Old English *þe* may be preceded by *se*, a lexeme that serves as both a pronoun and a definite determiner. This two-part relative construction may be an innovation intended to disambiguate the function of the relative-clause head *þe* within its clause.

2.3. SYNTACTIC MOVEMENT TO CP

- (77) a. *Ʒæs monnes nama wæs*
 the.M.GEN.SG man.M.GEN.SG name.N.NOM.SG be.PST.IND.3SG,
 [*sē þe hī behēafdade*] *Dorotheos*
 the.M.NOM.SG REL they.ACC.PL behead.PST.IND.3PL Dorotheos.M.NOM.SG
 ‘The name of the man who beheaded them was Dorotheos’

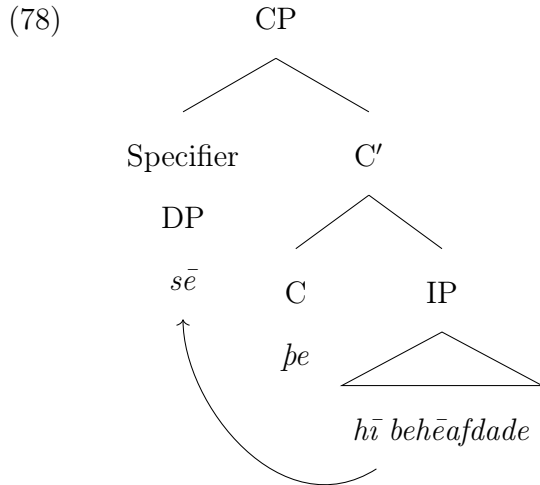
(Old English. *Martyrology* 98)

- b. *Gode āġeaf þone dēorwyrðan*
 God.M.NOM.SG give-up.PST.IND.3SG the.M.ACC.SG precious.WK.M.ACC.SG
ġym [þone þe Dēofol wolde
 gem.M.ACC.SG the.M.ACC.SG REL devil.M.NOM.SG want.PST.IND.3SG
ġerēafian]
 steal.INF

‘God gave that precious gem, which the Devil wanted to steal’

(Old English. *Martyrology* 236)

We can treat this ‘*se + þe*’ construction as a further example of *wh*-movement in Old English, if we assume that the determiner moves up into the CP structure, of which *þe* is the head.



At this early stage, *sē* has not yet become a definite article akin to modern *the*; it still maintains pronominal uses in Old English, and on this basis, it functions as a moved DP argument of the subordinate verb and marker of relativity.

WH-MOVEMENT IN OLD CHURCH SLAVONIC

Wh-questions are well attested in our Old Church Slavonic sources. In questions with *wh*-words, the polar particle *li* is not used, so the two types of interrogative marker appear to be in complementary distribution. As is usual with Old Church Slavonic, it is a challenge to tease out the native Slavic syntax from underneath the influences of Greek; interrogative pronouns, adverbs and phrases closely imitate the word order of the Greek source material.

- (79) a. **čto** = *sē* *mьnitь* *vamь?*
 what.N.NOM.SG self.ACC.SG think.PRS.3SG you.DAT.PL
 ‘What do you think?’

2.3. SYNTACTIC MOVEMENT TO CP

(OCS. *C. Mar.* Jh. 11.56)

- b. **кѣто** *отъ васъ* *обличаатъ* *мѣ* *о грѣсѣ?*
 who.M.NOM.SG of you.GEN.PL prove.PRS.3SG I.ACC.SG of sin.M.LOC.SG
 ... **по** **ѣто** *вы* *не емлете* *вѣры* *мѣнѣ?*
 after why you.NOM.PL not have.PRS.2PL faith.F.GEN.SG I.DAT.SG
 ‘Who of you can prove me guilty of sin? ... Why do you not have faith in
 me?’

(OCS. *C. Mar.* Jh. 8.46)

- c. **къде** *положисте* *и?*
 where lay.AOR.2PL he.ACC.SG
 ‘Where have you laid him?’

(OCS. *C. Mar.* Jh. 11.34)

The questions in (79) show a nearly one-to-one correspondence with the Greek word order, with the *wh*-element first in its clause.²⁵ However, it is likely that such apparent *wh*-movement may be a case of overlap between the grammars of Greek and OCS; for one reason, Slavic languages generally display *wh*-movement (Sussex and Cubberley, 2006, 360), so it is unlikely that OCS was significantly different in this regard. Moreover, the absence of *li* in *wh*-questions is systematic, suggesting that *wh*-movement was an equally productive alternative means of expressing interrogativity. If *li* realises a C head that requires lexical form, it seems that *wh*-movement into Spec,CP could likewise satisfy that same syntactic requirement.

The lexical items that introduce OCS relative clauses are also native vocabulary, such as the relative pronoun *iže* ‘who, which’.

²⁵OCS even employs a two-word phrase *po čto* to match Greek *dià tí*

2.3. SYNTACTIC MOVEMENT TO CP

- (80) a. *i malo ixъ estъ [iže =i*
 and few they.GEN.PL be.PRS.IND.3SG which.M.NOM.PL it.ACC.SG
obrětajotъ]
 find.PRS.3PL
 ‘And they are few who find it’
 (OCS. *C. Mar.* Mt. 7.14)
- b. *se otrokъ moi [egože izvolixъ]*
 behold servant.M.NOM.SG my.M.NOM.SG which.M.GEN.SG choose.AOR.1SG
vъzljubleny moi [na nъže
 beloved.PSTP.M.NOM.SG my.M.NOM.SG on which.M.ACC.SG
blagoizvoli dša moě]
 delight.AOR.3SG soul.F.NOM.SG my.F.NOM.SG
 ‘Behold my servant, whom I chose, my beloved, in whom my soul delights’
 (OCS. *C. Mar.* 12.18)
- c. *po Pilatovu pisaniju [ježe na*
 along Pilate’s.N.DAT.SG writing.N.DAT.SG which.N.ACC.SG on
krъstě gospodъni napisa]
 cross.M.LOC.SG Lord’s.M.LOC.SG write.AOR.3SG
 ‘According to Pilate’s words which he wrote on the Lord’s cross’
 (OCS. *Life of Methodius* 6)

These subordinate clauses closely imitate the word order of the Greek sources. However, examples (80a) and (81a) display a rare deviation from the Greek; these two OCS relative clauses with *iže* do not translate finite relative clauses in their Greek counterparts, but rather nominal phrases marked by the definite article.

- (81) a. *otъ otca vašego [iže estъ na*
 from father.M.GEN.SG your.M.GEN.SG which.M.NOM.SG be.PRS.3SG in

2.3. SYNTACTIC MOVEMENT TO CP

nbsxъ]
 heaven.N.LOC.PL
 ‘From your father who is in heaven’
 (OCS. *C. Mar.* Mt. 6.1)

- b. *parà tōi Patri humōn [tōi en*
 from the.M.DAT.SG father.M.DAT.SG you.GEN.PL the.M.DAT.SG in
toīs ouranoīs]
 the.M.DAT.PL heaven.M.DAT.PL
 ‘From the father of yours in the heavens’
 (Koine Greek. Mt. 6.1)

Lacking a matching definite article, the construction in (81b) is not available to OCS, which therefore resorts to *īže* and a finite relative clause. Freed from strict imitation, examples (81a) and (80a) bring us somewhat closer to the native Slavic syntax, which we may therefore presume to have had subordinate relative clauses with clause-initial markers. In conjunction with the comparative Slavic evidence (Sussex and Cubberley, 2006, 383-5), this offers a little more support for the existence of a CP component in the Slavic clause, which attracts and hosts interrogative and relative material.

WH-MOVEMENT IN OLD IRISH

Wh-elements in Old Irish fall into one of two groups: ‘*stressed*’ and ‘*unstressed*’. The unstressed *wh*-elements include *ce/ci/cía* and *ced/cid*, which are pronominal, used as the subject or the object of the question. There is also *co*, an *wh*-adverb meaning ‘how?’, although it does not appear in the language of the manuscript glosses. These elements all behave like conjunct particles in the initial position of the verbal complex; there they

2.3. SYNTACTIC MOVEMENT TO CP

precede any preverbs and the verb itself, which will appear with conjunct inflection.

- (82) a. *ce·rricc?*
 what-reach.PRS.IND.3SG
 ‘Why?’ (literally ‘what does it reach?’)
 (Sg. 199b28)
- b. *a Féidelm banfáith, co·acci*
 oh Fedelm.F.VOC.SG prophetess.F.VOC.SG how-see.PRS.IND.2SG
in slúag?
 the.M.ACC.SG host.M.ACC.SG
 ‘Oh prophetess Fedelm, how do you see the army?’
 (Old Irish. *Táin Bó Cúailnge* (I) 48)

Meanwhile, the elements of the stressed group show distinctions for number and gender (with masculine *cía*, feminine *cisí/cessi*, neuter *cid/ced* and plural *citné*), and presumably were more emphatic for speakers than their unstressed counterparts. This group not only behaves differently in terms of morphology; they are also not conjunct particles and they precede another kind of clausal construction. Although the morphology may not be obvious, the verb that they appear with must be relative, meaning that the verb belongs to its own separate relative clause. Stressed *wh*-words therefore appear in bi-clausal cleft constructions, albeit with the copula absent (García-Castillero, 2020, 192-5). In other words, a question akin to ‘*who do you see?*’ that uses a stressed *wh*-pronoun really has the structure ‘*who [is it] that you see?*’.

- (83) a. *cid atob·aig dó?*
 what PV-you(PL)-impel.PRS.IND.2SG.REL to.3SG

2.3. SYNTACTIC MOVEMENT TO CP

‘What impels you to it?’

(Old Irish. Wb. 19d10)

- b. *cisi rann* *do·géntar* *di?*
 what-part.F.NOM.SG make.FUT.PASS.3SG.REL from.3SG
 ‘What part will be made from it?’

(Old Irish. Sg. 27a21)

Both stressed and unstressed interrogative pronouns form the basis for complex *wh*-phrases. These may serve adverbial functions, such as *cia-indas* ‘how?’ and *cia-eret* ‘how long?’, literally ‘what manner?’ and ‘what length?’ (García-Castillero, 2020, 199-203). These complex elements similarly form questions with relative verbs.

- (84) a. *ce mét aimmser* *bes* *indi?*
 what-number.F.NOM.SG-time.F.GEN.PL be.PRS.SBJ.3SG.REL in.3SG
 ‘What number of times will be in it?’

(Old Irish. Sg. 25b32)

- b. *cindas* *mbias* *iar n-a cétbuid=sem?*
 what-manner be.FUT.3SG.REL after their opinion.VN.F.DAT.SG-them
 ‘How will it be according to their opinion?’

(Old Irish. Sg. 40a19)

A CP-syntax again explains well the behaviour that we see with Old Irish *wh*-questions (cf. Bate, 2024). Unstressed *wh*-words can be located either in C or in Spec,CP (the latter if we want to adopt an analysis of *wh*-movement from a lower argument position). This high position produces the pre-verbal position of subject and objects in questions, which contrasts with the default VSO order of declarative main clauses. The CP is thus lexically

2.3. SYNTACTIC MOVEMENT TO CP

realised through an C head or through *wh*-material in Spec,CP, and it has no effect on the verb in IP; unlike relative and declarative main clauses, there are no prosodic requirements that the verb in the IP below must satisfy (see §2.2.3 and Bate (2024)). This results in conjunct inflection and prototonic forms for the verb.

As for questions with stressed *wh*-words or a *wh*-phrase, we can consider these to be cleft constructions, and underlyingly to comprise two clauses, with the *wh*-word or phrase in the higher and the relative verb in the lower. It remains unclear though in which syntactic position in the higher clause the *wh*-material stands; it may similarly occupy Spec,CP, but we lack the word-order evidence to claim so confidently. We can only do so on grounds of theoretical parsimony and uniformity.

INTERMEDIATE SUMMARY: *WH*-MOVEMENT AS EVIDENCE FOR INDO-EUROPEAN C

Across the seven languages surveyed, the fronting of interrogative and relative elements to an initial or early position in the clause is consistently present. This fronting can be derived in each language, as for modern languages, through the syntactic operation of *wh*-movement, through which arguments or adjuncts of the verb are attracted to a hierarchically high part of the syntactic structure. Their attractor we may call C, since in that leftward position they resemble subordinators and other markers of clause type. *Wh*-movement therefore offers indirect support for the functional head C as a common feature of the Indo-European clause.

A key point to note is that this review of *wh*-movement has concentrated on words, not

2.3. SYNTACTIC MOVEMENT TO CP

phrases, in which both the moved material consists of the interrogative or relative element and a pied-piped head noun. This is for a good reason; phrasal *wh*-movement is poorly attested across the language set. Especially in the older three languages, it seems that early Indo-European interrogative and relative elements were unable to pied-pipe their head noun, leaving it stranded later in the clause instead (cf. Ram-Prasad, 2022, 45). (85) illustrates with Sanskrit *ká-* and *yá-*.

- (85) a. *kásminn á yatatho jáne?*
what.M.LOC.SG PV connect.PRS.IND.ACT.2DU person.M.LOC.SG
'To which people are you joining?'
(Sanskrit. RV 5.74.2)
- b. *yéna māsām ásiṣāsann ṛténa*
which.N.INS.SG month.M.ACC.PL gain.IPRF.IND.ACT.3PL order.N.INS.SG
'By which law they sought to gain the months'
(Sanskrit. RV 3.31.9)

We could treat these orders as instances of extraction, in which *ká-* and *yá-* have moved out of their origin noun phrase and up to Spec,CP. This is difficult to support though, since these could equally be explained as cases of apposition between a interrogative/relative pronoun and a syntactically independent noun. Moreover, the interaction of these elements with Wackernagel's law (see §2.3.5) suggests that *ká-*, *yá-* and their Ancient Greek cognates in general did not act as determiners and form phrases with head nouns. If confirmed by further data, this offers a window into the diachrony of Indo-European *wh*-movement. In whichever proto-stage we reconstruct it back to, it may have started by strictly attracting

2.3. SYNTACTIC MOVEMENT TO CP

only the clause-type-associated word to the CP.²⁶

Given its consistent presence in the comparative evidence, we might straightforwardly propose that the syntax of Proto-Indo-European itself included *wh*-movement. Yet there is the complicating issue of Hittite. Unlike elsewhere in early Indo-European, *wh*-questions in Hittite do not consistently display a fronted *wh*-word or phrase. Instead, as Hoffner (1995, 101) summarises, “the interrogative word sometimes precedes the finite verb as closely as possible. And since the finite verb is usually clause final, the interrogative word gravitates to the end of the clause”.

- (86) ^d*U* ^{URU}*Nerik TUKU-an* *ZI-an* *kuēzza* *KASKAL-aḫmi?*
 Stormgod of-Nerik angry.ACC.SG soul.ACC.SG what.ABL. appease.PRS.1SG
 ‘By what means can I appease the angry soul of the Stormgod of Nerik?’

(Hittite. KUB 5.1 i 92–3, oracular inquiry, from Huggard (2011, 89))

The absence of similar fronting has led some scholars (but not all) to consider Hittite to be a *wh-in-situ* language, in which *wh*-material remains in its original position, as befits its function within the clause; *kuēzza* in (86) appears in the usual relative position for such an adjunct. Others have disputed this; Huggard (2011) and Sideltsev (2014) agree that, while “there is no prototypical *wh*-movement to Spec,CP triggered by +*wh*-feature” (Sideltsev, 2014, 219), *wh*-material does not necessarily surface in its basic position, but rather is subject to other kinds of movement. For Huggard, it may be *in-situ* or stand in a dedicated position for focus. Regardless, *wh*-words and phrases do not seem to move to

²⁶This we can model as a feature, e.g. [Interrogative] or [Relative], that was inherent to the etyma of *ká-* and *yá-*, and which through movement satisfied a feature of the higher C head.

2.3. SYNTACTIC MOVEMENT TO CP

the same location as in other Indo-European languages, nor for the same reason.

Consequently, it becomes difficult to propose *wh*-movement for the ancestor of all Indo-European languages, the reconstruction of which depends on the Anatolian evidence (see §1.4.7). It seems unlikely that Hittite should lose this *wh*-movement to CP; rather, it instead appears to maintain an older grammar with *wh*-movement to a low Focus position. The Anatolian evidence does necessarily not refute the view that Proto-Indo-European syntax in its earliest state also included CP, but it deprives us of a key diagnostic for identifying CP within Anatolian and thus reconstructing an earlier CP. *Wh*-movement to CP within interrogative and relative clauses may instead have been an innovation within Proto-‘Core’-Indo-European (see §1.4.7), excluding the Anatolian branch of the family. While Hittite attests movement determined primarily by discourse-informational factors (such as focus), out of this syntax may have then emerged a new kind of movement for the purpose of clause typing, with interrogative and relative elements targeting the high position where they could signal the illocutionary force of the overall clause.

2.3.2 MOVEMENT TO CP IN IMPERATIVE CLAUSES

One further type of clause relevant to the thesis is the imperative clause, since a CP structure can derive the distinctive word order of these too in early Indo-European. As the word suggests, such clauses include verbs with imperative morphology and are typically used to express commands. In brief, the imperative verb displays a slight tendency to stand early in its clause. For some of the language set, the placement of the imperative is

2.3. SYNTACTIC MOVEMENT TO CP

not noteworthy; Old Irish imperative verbs, for instance, adhere to the language’s general order of Verb-Subject-Object, regardless of clause type, while both declarative main and imperative clauses in Old English and Old Norse place all finite verbs in second, third or first position (more on which in §3.2.2). Yet for two of the three ancient languages, the frequently clause-early position of the imperative verb stands in contrast with their clause-late declarative counterparts and an SOV order.

For example, the (non-negative) imperative verb in Latin has a strong association with the initial position of the clause (Pinkster, 2021, 1033-6). This tendency varies though according to the sub-type of imperative clause and illocutionary force (e.g. directive, optative, concessive imperatives, cf. Risselada (1993)). Directive imperative verbs, with their prototypical function of commanding, usually occupy the initial position.

- (87) a. *cavē* *canem*
 beware.PRS.IMP.ACT.2SG dog.M.ACC.SG
 ‘beware of the dog’
 (Latin. Petronius, *Satyricon* 29)
- b. *gerite* *amanti* *mihi* *mōrem* ...
 bear.PRS.IMP.ACT.2PL love.PRS.P.N.ABL.SG I.DAT.SG manner.M.ACC.SG
fīte *causā* *meā* *lūdiū*
 do.PRS.IMP.PASS.2PL sake.F.ABL.SG my.F.ABL.SG actor.M.NOM.PL
barbari, *sussilite*, *obsecrō*,
 barbarian.M.NOM.PL jump-up.PRS.IMP.ACT.2PL implore.PRS.IND.ACT.1SG
et mittite *istanc* *forās*
 and send.PRS.IMP.ACT.2PL this.F.ACC.SG out
 ‘Give to me in love your manner ... become for my sake barbarian actors, leap
 up, I beg you, and send this one outside’

2.3. SYNTACTIC MOVEMENT TO CP

In (88), initial *sĭc* anticipates the subsequent information that Cicero wants Tiro to know, while *hanc* and *hanc rem* are topical, the former referring back to *domesticā crūdēlitāte* ‘domestic cruelty’ in the previous sentence. All in all, while there is a leftward attraction of imperative verbs in Latin, there is evidence for an even higher position reserved for discourse-functionally marked constituents. This is discussed in Chapter 3.

Ancient Greek displays similar word orders. Its imperative verbs (which may bear imperative or subjunctive morphology) show a preference for an early position; it precedes its object, subject and any adverbial constituents in (89) and (91c), although it does not precede the vocative addressee in (89d). This does not necessarily entail a Subject-Verb-Object order for (89d) though, as a vocative phrase is not, syntactically or semantically, the same thing as a clausal subject.

- (89) a. *gnōthi* *sautón*
 know.AOR.IMP.ACT.2SG yourself.ACC.SG
 ‘Know yourself’
 (Ancient Greek. Plato, *Protagoras* 343)
- b. *hupolúete,* *paîdes,* *Alkibiádēn*
 loose-under.PRS.IMP.ACT.2PL child.M.VOC.PL Alcibiades.M.ACC.SG
 ‘Take off Alcibiades’ shoes, slaves’
 (Ancient Greek. Plato, *Symposium* 213)
- c. *kai skópei* *ekeinōi* *tōi* *akribēi*
 and view.PRS.IMP.ACT.2SG this.M.DAT.SG the.M.DAT.SG precise.M.DAT.SG
lógōi
 reasoning.M.DAT.SG
 ‘Consider in this precise manner of reasoning’

2.3. SYNTACTIC MOVEMENT TO CP

(Ancient Greek. Plato, *Republic* 1.342)

- d. *ô pankratès Zeû, trépson eis*
 oh almighty.M.VOC.SG Zeus.M.VOC.SG turn.AOR.IMP.ACT.2SG into
ekhthroùs bélos!
 enemy.M.ACC.PL missile.N.ACC.SG
 ‘Almighty Zeus, turn your missile against our enemies!’

(Ancient Greek. Aeschylus, *Seven Against Thebes* 255)

There seems to be a leftward attraction of the verb in the imperative clause, on the basis of the greater frequency of its early position there than in the declarative clause. Statistical data are necessary for any stronger statement, but the generalisation has been acknowledged by others (cf. Recht, 2015, 63). One constituent of the clause that consistently precedes the imperative verb immediately is its negator, usually *mè*.

- (90) a. *paûe, mèn léxēis péra!*
 stop.PRS.IMP.ACT.2SG not say.AOR.SBJ.ACT.2SG further
 ‘Stop, do not speak further!’

(Ancient Greek. Sophocles, *Philoctetes* 1276)

- b. *kai nûn mèn pephóbēsthe aûtén*
 and now not fear.PRF.IMP.MED.2PL this.ACC.SG
 ‘And now be not afraid of this’

(Ancient Greek. Thucydides, 6.17)

If there is movement within the negative imperative clause, it may involve the movement of both the verb and the negator as a unit.

Further evidence of movement is that the imperative verb may act as host for any

2.3. SYNTACTIC MOVEMENT TO CP

‘second-position’ enclitic elements. Sometimes, as in (91a), the imperative and the clitic are indeed the first and second words and constituents in their clause, yet in (91b), they stand after an initial topical phrase.

- (91) a. *ô Krittōn ... apagētō =tis*
 oh Crito.M.VOC.SG lead-away.PRS.IMP.ACT.3SG someone.M.NOM.SG
autēn oĩkade
 this.F.ACC.SG homeward
 ‘Crito, have someone take her home’
 (Ancient Greek. Plato, *Phaedo* 60A)
- b. *pròs taũta mē psauṣēi =tis*
 towards this.N.ACC.PL not touch.AOR.SBJ.ACT.3SG someone.M.NOM.SG
Argeiōn emoũ
 Argive.M.GEN.PL I.GEN.SG
 ‘Therefore, let no one of the Argives touch me’
 (Ancient Greek. Euripides, *Iphigenia in Aulis* 1559, from Hale (2017))
- c. *kai =moi mē thorubēsēi mēdeis prìn*
 and I.DAT.SG not make-noise.AOR.SBJ.ACT.3SG no-one.M.NOM.SG before
akoũsai
 hear.AOR.INF.ACT
 ‘No one speak before hearing me out!’
 (Ancient Greek. Demosthenes, *On the Peace* 5.15)

For (91b), we may claim that the imperative unit *mē psauṣēi* works as a host for *tis* not because it is the first word or constituent, but because it occupies the syntactic position in the left periphery that clitics in Ancient Greek also target. The mechanics of this are covered in §2.3.5. If we adopt this analysis, however, examples like (91c) lead us to conclude

2.3. SYNTACTIC MOVEMENT TO CP

that the movement of the imperative to this position is not obligatory, since enclitic *moi* instead appears after initial *kaí*.

As with *nûn* in (90b) and *pròs taûta* in (91b), the constituent that the early imperative verb follows often has an extra-clausal, topical function. In (92), we find the initial subject *humeîs*, which refers back to the soldiers addressed by the speaker and acts as a contrastive topic with *egò* in the subsequent clause, as well as the topic *taûta*, which refers to promises just mentioned.

- (92) a. *kaí humeîs* = *mèn mè ekdôte* = *me, egò*
 and you.NOM.PL and not give-up.AOR.SBJ.ACT.2PL I.ACC.SG I.NOM.SG
 = *dè emautón...*
 and myself.ACC.SG
 ‘And don’t you give me up, but I will myself...’

(Ancient Greek. Xenophon, *Anabasis* 6.6.18)

- b. *taûta* = *moi prâxon, téknon*
 this.N.ACC.PL I.DAT.SG do.AOR.IMP.ACT.2SG child.N.VOC.SG
 ‘Fulfil these things for me, child’

(Ancient Greek. Sophocles, *Philoctetes* 1399)

Ancient Greek therefore patterns with Latin, in that something attracts the imperative verb into the left periphery, yet not into the functional projection reserved for topical constituents (see Chapter 3).

With Sanskrit imperative clauses, there does not appear to be one distinct position or distribution in the clause for the imperative verb. It can be clause-medial; example (93a) contains two imperative verbs, the third words of their respective clauses. It can also be

2.3. SYNTACTIC MOVEMENT TO CP

clause-initial (as in (93c)) and clause-penultimate (93b).

- (93) a. [*evá anéna havíṣā yakṣi*
 so this.N.INS.SG offering.N.INS.SG sacrifice.PRS.IMP.ACT.2SG
devān] [*manusvād yajñám prá tira*
 god.M.ACC.PL Manu-like worship.N.ACC.SG PV perform.PRS.IMP.ACT.2SG
imám adyá]
 this.N.ACC.SG today
 ‘So, with this offering sacrifice to the gods. Like Manu, carry out this worship
 today’

(Sanskrit. RV 3.17.2, from Schäufele (1991b, 5))

- b. *prá hót্রে pūrvyám váco*
 PV high-priest.M.DAT.SG excellent.N.ACC.SG speech.N.ACC.SG
Agnáye bharatā bṛhát
 Agni.M.DAT.SG bear.PRS.IMP.ACT.2PL great.N.ACC.SG
 ‘To Agni, the high priest, offer up your best, greatest speech’

(Sanskrit. RV 3.10.5)

- c. [*stuhí śrutám gartasádam*
 praise.PRS.IMP.ACT.2SG listen.PSTP.M.ACC.SG chariot-sit.M.ACC.SG
yúvānam...] [*mṛlā jaritré*
 young.M.ACC.SG pardon.PRS.IMP.ACT.2SG singer.M.DAT.SG
Rudrá stāvāno] [*anyám =te*
 Rudra.M.VOC.SG praise.PRSP.MED.M.VOC.SG other.M.ACC.SG you.GEN.SG
asmán ní vapantu sēnāḥ]
 we.ABL.PL PV scatter.PRS.IMP.ACT.3PL army.F.NOM.PL
 ‘Praise the famous, the chariot-borne, the young one ... Oh Rudra, pardon
 the singer, let your armies smite down another, not us’

(Sanskrit. RV 2.33.11)

Numerous constituents of an imperative clause are therefore able to precede the main

2.3. SYNTACTIC MOVEMENT TO CP

verb. The tentative conclusion may be that imperative clauses operate no differently from declarative clauses and according to the same discourse-information structure. However, although imperative verbs are variable in their relative position, their distribution is not completely without tendencies. Imperative verbs frequently make use of the clause-initial position, presumably to take on the salience that it offers. They also seem to avoid the final position in the clause, contrary to the typical verb-finality in declarative clauses. This tendency may not be evidence of any significant syntactic processes; it may simply be Sanskrit utilising its syntactic flexibility to avoid placing the verb in absolute final position, which, considering the intonation of the clause as a whole, is so prosodically weak that it would work against the intended force of an imperative verb.

Notably, there are clauses like those in (94) that display a non-initial imperative that hosts an enclitic pronoun. The pronouns apparently contravene Wackernagel's law (1892), coming third in the clause (although in fact, this is not a contravention, see §2.3.5). In each, the host of the pronouns *naḥ* and *tvā* is the third-person imperative verb *vārdhantu* 'increase'.

- (94) a. *tvāṃ vardhantu =no gírah*
you.ACC.SG increase.PRS.IMP.ACT.3PL we.GEN.PL hymn.M.NOM.PL
'May our hymns increase you'
(Sanskrit. RV 1.5.8)
- b. *Agnīm vardhantu =no gírah*
Agni.M.ACC.SG increase.PRS.IMP.ACT.3PL we.GEN.PL hymn.M.NOM.PL
'May our hymns increase Agni'
(Sanskrit. RV 3.10.6)

2.3. SYNTACTIC MOVEMENT TO CP

- c. *divédive* *sadṛśīr* *Indra* *túbhyaṃ*
 day-day.M.LOC.SG similar.F.NOM.PL Indra.M.VOC.SG you.DAT.SG
várdhantu =*tvā* *somapéyāya*
 increase.PRS.IMP.ACT.3PL you.ACC.SG Soma-drink.M.DAT.SG
dhṛṣṇo
 bold.M.VOC.SG
 ‘Day by day may the same for you, Indra, increase you, bold one, for the soma
 drink’
 (Sanskrit. RV 3.52.8)

Various analyses are possible for the arrangements in (94a) and (94b). One is that *naḥ* is enclitic on the *várdhantu* because it remains within its noun phrase, which so happens to follow the verb. Alternatively, both the imperative and clitic have moved into somewhere in the higher clausal syntax. This movement analysis of pronouns is integral to the arguments in §2.3.5. Specifically, *várdhantu* and the enclitic pronoun may have both moved to C, while *tvām* and *Agnīm* occupy an even higher topical position. This multiple-movement derivation produces the attested orders in (94a) and (94b), although the orders could equally be derived through the movement of the pronoun of C, followed by its prosodic inversion with the lower verb *várdhantu*. However, this would not account for the unusual post-verbal subject *gírah* in (94a) and (94b). If both *várdhantu* and the enclitic pronoun move to C, this order is accounted for.

The multiple-movement analysis can also derive the much later position of the enclitic pronoun *tvā* in (94c); here *várdhantu* may have again moved to C, where it can host the pronoun post-syntax, but stands lower than the positions of the constituents *divédive*

2.3. SYNTACTIC MOVEMENT TO CP

sadṛśīr ... túbhyaṃ. A complicating factor though is that (94c) comprises two *padas* (lines of a stanza), with a line break between *túbhyaṃ* and *várdhantu*. It may be the metrical template of Vedic language instead that puts *tvā* after *várdhantu*, and not after the preceding adverb *divédive* or the subject phrase *sadṛśīr túbhyaṃ*.

In sum, Latin and Ancient Greek evidence (with weak evidence from Sanskrit) for the fronting of imperative verbs within their clause. The position that they each target appears to be the same; they may move to somewhere high in the syntax, which produces a clause-early surface position, but are consistently preceded by discourse-informationally marked constituents. Ancient Greek and Sanskrit offer some evidence that in that position, imperative verbs may host Wackernagelian clitic pronouns. Given all of this behaviour, imperative verbs resemble other subordinating, relative and interrogative markers of clause type. If we posit a C head that is responsible for clause type, we gain a motivation for the behaviour of imperative verbs in three languages; with the feature [Imperative], C attracts the imperative verb. Verbal movement does not seem to have been obligatory though, given the instances of its later position in all three ancient languages, beyond what the proposal of an additional position above C can account for. This is especially true of weaker commands that used optative or subjunctive verbal morphology instead. We might therefore distinguish between C with a [Strong Imperative] feature, which triggered verbal movement, and C with [Weak Imperative], which did not.

2.3.3 MOVEMENT TO CP IN DECLARATIVE MAIN CLAUSES

There are a handful of examples across the language set in which declarative verbs can be said to move up to a C position. There is the case of the pre-Old Irish Cowgill particle **eti*, mentioned in §2.2.3. Eska (2012) has proposed that this particle triggered the fronting of simple finite verbs to bear it as an affix; their fusion over time produced Old Irish absolute inflection.²⁷ However, it could also trigger the fronting of a compound verb's first preverb instead, in time producing the prosodic gap between deuterotonic compound verbs. So, while Pre-Old Irish may offer an example of verbal movement to C in declarative main clauses, expressing clause type through a distinct position for the verb does not seem to have been a factor. Rather, its purpose was to provide **eti* with a host.

The two Germanic languages though do provide ample evidence of verbal movement to C within declarative main clauses. They offer this through the well-documented Germanic phenomenon of Verb Second. This term is somewhat erroneously applied to Old English and Old Norse word order, since Verb First and Verb Third orders are copiously attested and show systematic alternation with Verb Second. These word orders and their connection to clause typing are discussed in detail in §3.2.2. For the moment, it suffices to say that the analysis for early Germanic V1, V2 and V3 relies wholly on the obligatory movement of the verb to a C position in the syntax of the declarative main clause, following den Besten

²⁷Quite why it had to be the simple finite verb, and not any other constituent, that hosted enclitic **eti* is not clear. We can speculate about a close association between the declarative-typing particle **eti* and declarative verbs, but this does not explain why preverbs, if present, took priority in hosting **eti*. This behaviour of Pre-Old Irish clauses with **eti* demands further work.

2.3. SYNTACTIC MOVEMENT TO CP

(1977).

2.3.4 MOVEMENT TO CP IN SUBORDINATE CLAUSES

As a final section on verbal movement to C, it remains to be said that Old English and Old Norse display the movement of the finite verb to C in subordinate (specifically conditional) clauses. Alongside conditional clauses with the overt subordinators *gif* and *ef*, the two Germanic languages have the option of marking the protasis with a verb-initial order, so long as the verb is subjunctive.

- (95) [*hæfde ic ælteowe þēnas*]
have.PST.SBJ.1SG I.NOM.SG all-good.ST.M.ACC.PL servant.M.ACC.PL
nāre ic þus ēaðelīce oferswīðed
not-be.PST.SBJ.1SG I.NOM.SG thus easily conquer.PSTP.M.NOM.SG
'Had I had honest servants, I would not be so easily conquered'

(Old English. *Ælfric's Forty Holy Soldiers* 226, from Links (2019, 24))

For Old Norse, Faarlund (2004, 252) reports that such clauses serve two functions: as the protasis of conditional constructions and as the second of two coordinated subordinate clauses.

- (96) [*hefði þá verit þetta boðit*] þá
have.PST.SBJ.3SG then be.PSTP. this.N.NOM.SG offer.PSTP.N.NOM.SG then
væri margr maðr...
be.PST.SBJ.3SG many.M.NOM.SG man.M.NOM.SG
'Had this then been offered, then many a man would have been...'

(Old Norse. *Heimskringla* 9.91, from Faarlund, 2007, 252)

2.3. SYNTACTIC MOVEMENT TO CP

In the absence of an overt subordinator to realise C, we may treat these as instances of verbal movement to C. Rather than being movement for the purpose of signalling subordination, this is an instance of generic Germanic V-to-C movement, which is discussed in §3.2.2. Yet, in tandem with subjunctive morphology, this generic movement helps to express the conditional status of such clauses.

2.3.5 PRONOMINAL MOVEMENT TO C: WACKERNAGEL'S LAW RECONSIDERED

The final example of movement to C does not involve verbs, but rather pronouns in early Indo-European languages. It specifically concerns oblique pronouns in early Indo-European that consistently ‘gravitate’ towards the start of their clause. This thesis understands this pattern to be the product of syntactic movement up to C. On the basis of its considerably wide distribution across early Indo-European, both the C target and the operation of pronominal movement should be reconstructed for the proto-syntax.

The word-order behaviour in question is referred to for short as ‘Wackernagel’s law’, after Wackernagel (1892). Yet the history of this idea begins earlier with Delbrück (1878) and the following observation for Sanskrit:

“Enklitische Wörter rücken möglichst an den Anfang des Satzes. Die Richtigkeit dieser Beobachtung lässt sich besonders an den enklitischen Casus von Pronominibus erweisen, weil wir bei ihnen wissen, an welcher Stelle des Satzes sie stehen müssten, wenn sie nicht enklitisch wären...”

2.3. SYNTACTIC MOVEMENT TO CP

(Enclitic words preferably move to the start of the sentence. The correctness of this observation can be proven particularly with the enclitic cases of pronouns, because we know where they would have to be in the sentence if they were not enclitic...)

(Delbrück, 1878, 47-8)

Delbrück has in mind examples like the two in (97), in which pronouns and (coordinating and subordinating) conjunctions *vā* and *ca* cluster after the initial word in their clause.

- (97) a. *átha =enam =me púnar dadat*
then he.ACC.SG I.DAT.SG again give.PRS.SBJ.ACT.3SG
'Then may he give him back to me'
(Sanskrit. RV 4.24.10)
- b. *ító =vā =tám Indra pāhi*
hence or that.M.ACC.SG Indra.M.VOC.SG protect.PRS.IMP.ACT.2SG
riṣáh, amá =ca =enam áraṇye
harm.F.ABL.SG at-home and he.M.ACC.SG wilderness.N.LOC.SG
pāhi riṣó
protect.PRS.IMP.ACT.2SG harm.F.ABL.SG
'Or protect him from harm here, Indra, and protect him from harm at home
and in the wilderness!'
(Sanskrit. RV 6.24.10)

What unites these words of different categories is thought to be a certain prosodic weakness and a lack of an accent.

Wackernagel (1892) recognised this behaviour with Indo-European pronouns and con-

2.3. SYNTACTIC MOVEMENT TO CP

junctions more widely, specifically in Latin, Ancient Greek, Sanskrit and Avestan, with brief mentions of similar phenomena in Germanic and Celtic. Wackernagel conducted his comparative review prior to the decipherment of Hittite in the 1910s; the same stacking after the first word of the clause is observed in Hittite too (Melchert and Hoffner, 2008, 410-2)

- (98) a. *dī* = ***illum*** *omnēs* *perdant,* *ita*
 god.M.NOM.PL that.M.ACC.SG all.M.NOM.PL destroy.PRS.SBJ.ACT.3PL so
 = ***mihi*** *hunc* *hodiē* *corrūpit* *diem*
 I.DAT.SG this.M.ACC.SG today spoil.PRF.IND.ACT.3SG day.M.ACC.SG
 ‘May all the gods ruin him, as he has ruined this day for me’

(Latin. Plautus, *Menaechmi* 4.2, from Wackernagel (1892))

- b. *ékhonti* = ***dé*** = ***hoi*** *en khersi* *toû*
 have.PRS.P.M.GEN.SG and he.DAT.SG in hand.F.DAT.PL the.M.GEN.SG
paidòs *tòn* *gámon* ... *Kroîsos*
 child.M.GEN.SG the.M.ACC.SG wedding.M.ACC.SG Croesus.M.NOM.SG
 = ***dé*** = ***min*** *ekáthēre*
 and he.ACC.SG purify.AOR.IND.ACT.3SG
 ‘Having to deal with the wedding of his son ... Croesus purified him’

(Ancient Greek. Herodotus, *Histories* 1.35, from Goldstein (2016, 2))

- c. *vṛdhé* = ***ca*** = ***no*** *bhavataṃ*
 support.F.DAT.SG and we.ACC.PL be.PRS.IMP.ACT.2DU
vájasātau
 prize-winning.F.LOC.SG
 ‘And be you two our support in winning the prize’

(Sanskrit. RV 1.34.12)

- d. *ašte* = ***se*** = ***bi*** *ne rodil̥*
 if self.ACC.SG be.COND.3SG not be-born.RESP.M.NOM.SG

2.3. SYNTACTIC MOVEMENT TO CP

vā in Sanskrit or *mén* and *dé* in Ancient Greek, and to the emphatic and causal particle *hí* in Sanskrit and its Avestan cognate *zī* (cf. West, 2011, 111). This common behaviour across the language family has been explained as an inheritance from Proto-Indo-European itself; for Watkins (1964, 1036), the position of unstressed elements at the start of the clause is “one of the few generally accepted syntactic statements” about the word order of the proto-language.

Yet Wackernagel’s law conceals a lot of variety across language, era and words affected. ‘Second position’ is ambiguous, because the Wackernagelian element may follow the first word or the first phrasal constituent in the clause. In examples like (98a), the pronoun (*illum*) interrupts the initial noun phrase to be enclitic on the very first word (*dī*) in ‘strict’ second position. Yet elsewhere we find the enclitic pronoun appearing later, following one or two words or constituents of the clause. This we can see in Sanskrit, as already noted for (94), and also in Latin, Ancient Greek and Old Irish.

- (99) a. *kaì mē boúlou* = *me* *anankásai*
 and not wish.PRS.IMP.MED.2SG I.ACC.SG force.AOR.INF.ACT
légein...
 say.PRS.INF.ACT
 ‘And do not wish to force me to say...’
 (Ancient Greek. Plato, *Phaedrus* 236)
- b. *autàr egò theós eimi* [*diamperès*
 but I.NOM.SG god.M.NOM.SG be.PRS.IND.3SG continuously
hē =se phulássō en pántessi
 who.F.NOM.SG you.ACC.SG guard.PRS.IND.ACT.1SG in all.M.DAT.PL
pónois]
 toil.M.DAT.PL

2.3. SYNTACTIC MOVEMENT TO CP

‘But I am a god, who guards you unfailingly in all your struggles’

(Ancient Greek. Homer, *Odyssey* 20.47-8)

- c. *idhmám yás =te jabhárac*
 fuel.M.ACC.SG who.NOM.SG you.DAT.SG bring.PRF.SBJ.ACT.3SG
chaśramāṇáh
 exert.PRSP.M.NOM.SG
 ‘Who, exerting himself, will have brought you fuel’

(Sanskrit. RV 4.12.2, from Hale (2018, 11))

- d. *múkham kím =asya?*
 mouth.N.NOM.SG what.N.NOM.SG he.GEN.SG
 ‘His mouth, what [was it called]?’

(Sanskrit. RV 10.90.11, from Ram-Prasad (2022, 82))

- e. *dogailse fo=d·ruar*
 grief.F.NOM.SG PV-it-cause.PRF.3SG
 ‘Grief has caused it’

(Old Irish. Ml. 20b17, from DiGirolamo (2018, 151))

Moreover, one group of Wackernagelian elements more consistently occupy the strict second position than the other, namely clausal conjunctions. Examples abound in which the Sanskrit or Ancient Greek clause contains both a Wackernagelian conjunction and pronoun, yet the two stand at a distance, with the only the conjunction in strict second position.

- (100) a. *tò =dè bêma tí =soi*
 the.N.NOM.SG and platform.N.NOM.SG what.N.NOM.SG you.DAT.SG
khrésimon éstai?
 use.N.NOM.SG be.FUT.IND.3SG
 ‘And the platform, what use will it be to you?’

(Ancient Greek. Aristophanes, *Ecclesiazusae* 677, from Goldstein (2015,

2.3. SYNTACTIC MOVEMENT TO CP

121))

- b. *stútaś* = *ca yās* = *te*
 hymn.M.NOM.PL and which.M.NOM.PL you.DAT.SG
cakánanta
 please.PRF.SBJ.MED.3PL
 ‘And may hymns which please you’

(Sanskrit. RV 1.169.4)

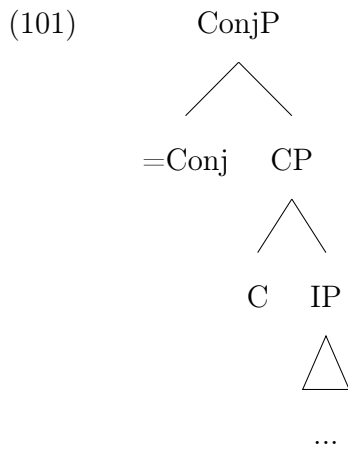
Wackernagelian conjunctions and pronouns thus seem subject to different rules of position. Furthermore, there is evidence in Ancient Greek for a prosodic difference between the two; Wackernagelian pronouns appear to trigger a secondary accent on their host, while conjunctions do not (Goldstein, 2015, 55-6). This difference suggests that the two groups form different prosodic units with their host.²⁸ All things considered, it is better therefore to understand Wackernagel’s law as a composite word-order phenomenon, and that the common clitic cluster is in fact a byproduct of the coincidence of different operations. As Ram-Prasad (2022, 88) puts it, it “cannot be described with a singular generalisation (“enclitics move to second position”), but rather requires a detailed grammatical analysis”.

On this basis, Wackernagelian elements should be derived as two distinct phenomena, namely WL1 (pronouns) and WL2 (conjunctions) (cf. Keydana, 2011; Ram-Prasad, 2022). The latter are “less complicated”, since they always follow the first word of the clause, whatever it be (Keydana, 2011, 3). Their placement can be straightforwardly accounted

²⁸A related issue is the traditional orthographic accent given in Ancient Greek to the second-position conjunctions *dé* and *mén*, but not to pronouns like *soi* or *tis*. This again indicates a different prosodic status for the two, although Goldstein (2015, 60) states that conjunctions are nonetheless “prosodically dependent” and are not equivalent in accent to lexical words.

2.3. SYNTACTIC MOVEMENT TO CP

for with Halpern’s idea of prosodic inversion (1992; 1995). Using this, the analysis of Hale (1996), Agbayani and Golston (2010) and Ram-Prasad (2022) is that WL2 conjunctions stand outside the clause in terms of their syntax, yet their prosodic weakness requires them to be hosted by the next proper prosodic word in the linear order. This post-syntactic operation ‘flip’ produces the strict second position of *dé*, *vā* and *ca* in (97), (98) and (100).



Prosodic inversion may at least have been the original cause of this position; we should note that, in Classical Greek, *mén* and *dé* appear after initial words generally considered to be clitic themselves, such as prepositions (e.g. *eis* ‘to’) or certain forms of the definite article. For example, in (102), the initial preposition *en* does not lend itself to an analysis with prosodic inversion, since it is typically proclitic on what follows, and so does not offer a suitable prosodic host for *dé*. We might suppose that initial words like *en* are no longer clitics and have been upgraded to full prosodic words in the presence of *mén* and *dé*; that the enclitic and proclitic ‘lean’ on each other to satisfy their prosodic requirements; or that

2.3. SYNTACTIC MOVEMENT TO CP

the two together are both proclitic on the third word in the clause.

- (102) *en=* (?=) *dè* (?=) *tê_i* *lím_{nē}_i* *taútē_i* *tà*
in and the.F.DAT.SG lake.F.DAT.SG this.F.DAT.SG the.N.ACC.PL
deíkē_{la} *tôn* *pathéō_n* *autoû* *nuktòs*
image.N.ACC.PL the.N.GEN.PL suffering.N.GEN.PL he.GEN.SG night.F.GEN.SG
poieû_{si}
do.PRS.IND.ACT.3PL
‘And on this lake they tell by night the stories of his sufferings’
(Ancient Greek. Herodotus, *Histories* 2.171)

This possible proclisis of WL2 elements in Ancient Greek may be a later development, in conjunction with the further grammaticalisation and prosodic weakening of prepositions and the definite article.

WL1 pronouns are more of a challenge to derive. These are less rigid in their position within the overall word order, but more constrained by what type of word they follow. They consistently follow interrogative or relative words and phrases, as the examples for Greek (100a), Sanskrit (100b) and Avestan (104b) illustrate (Hale, 1987a, 42). Examples in which the WL1 pronoun precedes are not attested. They seem unable to move any earlier in the clause than interrogative or relative material, a limitation that WL2 conjunctions like *dé* do not share.

Some scholars have understood the position of WL1 elements also to be primarily phonologically determined; such ‘prosody-dominant’ accounts include Goldstein (2010) and Keydana (2011). For Keydana, the difference between the two groups in Sanskrit is due to the prosodic unit on which they are enclitic; for WL2, it is a prosodic word, whereas

2.3. SYNTACTIC MOVEMENT TO CP

for WL1, it is instead the first prosodic ‘phrase’. This initial phrase Keydana understands as arising from the entire syntactic left periphery; this is to say, the topmost structure of the Sanskrit clause, which comprise a CP and a higher “discourse-functional slot”, is treated as one unit by the operations of prosody. Because of this unity, it offers a suitable host for WL1 pronouns further down within the clausal structure. Keydana adopts such an approach to avoid deriving their position through syntactic movement, an analysis he considers “unclear” and “problematic”.

However, prosody-dominant explanations for WL1 pronouns fall short. Taking Keydana’s account (2011) as representative, while it does make the correct predictions as to the word-order data in Sanskrit, it lacks a coherent definition of the initial prosodic unit. Keydana accepts that the left periphery of the Sanskrit clause may host multiple complex elements fronted from elsewhere in the clause, such as a *wh*-word and a topicalised noun (phrase) in the discourse-functional slot. Even in the absence of recorded contemporary Sanskrit phonology, it is difficult to envision how all this material in the left periphery might be interpreted by prosody as one unit, and therefore as one prosodic host. Keydana must consider such complex initial strings as *tvám vardhantu...* in (94a), *idhmám yás...* in (99c), *múkham kím...* in (99d) and the long participle phrase *amṛtatvám rákṣamāṇāsa...* in (103) as somehow prosodically unitary, if they are to host enclitic pronouns like *enam*.

- (103) [*amṛtatvám* *rákṣamāṇāsa*] = *enam* *devā*
immortality.N.ACC.SG protect.PRSP.MED.M.NOM.PL he.NOM.SG god.M.NOM.PL
Agním *dhārayan* *draviṇodám*
Agni.M.ACC.SG maintain.IPRF.IND.ACT.3PL wealth-giver.M.ACC.SG
‘Protecting their immortality, the gods maintained him as Agni, the giver of

2.3. SYNTACTIC MOVEMENT TO CP

wealth'

(Sanskrit. RV 1.96.6, from Hale (2018, 1935))

The Ancient Greek evidence likewise shows the placement of WL1 pronouns after strings that can hardly be considered one prosodic unit, such as *diampèrès hě...* in (69). Inversely, the Old Irish infixed pronoun *b* in (98e) follows the pretonic preverb *do*. The prosodic status of these preverbs in deuterotonic verbs is unclear, but it is agreed at least that they do not bear the primary stress of the verbal complex, and they may be proclitic (McCone, 1997, 4-5). In either case, they cannot be considered prosodic phrases, on which pronouns like *b* are enclitic. Moreover, Keydana does not satisfactorily explain for Sanskrit how the enclitic pronoun comes to stand at the right edge of the initial prosodic phrase. He derives it simply “in a PF operation”. Yet WL1 pronouns have various functions within the clause; they may be direct objects (103), indirect objects (100) and dependent genitives within noun phrases (99d). Regardless of their function, they seem to shift from whatever basic word-order position we propose for them, skipping over many suitable prosodic hosts along the way, to arrive at the same position in the linear order. Something seems to attract them all there, but it is difficult to consider that common centre of gravity as something phonological; the prosodic peak of the clause as a whole comes towards the middle in a typically verb-final language like Sanskrit (Hock, 1991, 476-7).

This thesis therefore instead belongs to the ‘syntax-dominant’ camp, which understands the word order of WL1 pronouns to be the product of syntax first, phonology second, and not only for Sanskrit, but for early Indo-European in general. This view is well established,

2.3. SYNTACTIC MOVEMENT TO CP

going back to Hale (1987b) (for Sanskrit), Garrett (1990) (for Hittite) and Taylor (1990) (for Ancient Greek). In brief, these analyses share the view that WL1 pronouns in early Indo-European move within the syntax to a high position. For Hale (1987b) and Sanskrit, this was originally “COMP”, later labelled simply ‘C’. Granted, this is movement of a phrasal unit (a DP) to adjoin to a head position (C), which is theoretically irregular. Such movement may be permissible considering the unusual nature of clitics (cf. Uriagereka, 1995a; Roberts, 2010); the view of Roberts is that clitic pronouns have phrasal and non-phrasal properties, being “simultaneously maximal and minimal elements ... able to move both as XPs and as heads” in the syntax (2010, 41).

A key claim of this chapter is that movement to an equivalent C derives this behaviour in general across the language set. The best evidence for it comes from Ancient Greek, Sanskrit, Old Church Slavonic and Old Irish, and additionally from Avestan. In Ancient Greek, lexical subordinators, such as *ei*, consistently precede the WL1 pronoun and host the enclitic, as *éite* does in (104a). Considering such subordinators as heads of category C, this typical immediately post-C position for WL1 pronouns can be treated as movement to C. It must be acknowledged that there is a small minority of orders in which the WL1 pronoun does not immediately follow, as noted for Homeric Greek by Taylor (1990, 37-52). These exceptional orders are addressed in §3.3.3.

- (104) a. *tí* = **dé** = **soi** ... *toûto*
 what.N.ACC.SG and you.DAT.SG this.N.NOM.SG
diaphérei [*éite* = **moi** *dokéi*]
 be-different.PRS.IND.ACT.3SG whether I.DAT.SG seem.PRS.IND.ACT.3SG

2.3. SYNTACTIC MOVEMENT TO CP

eíte mē...?

whether not

‘But what difference does it make to you, whether it seems so to me or not...?’

(Ancient Greek. Plato, *Republic* 1.349)

- b. *yōi =īt aθa vərəziiqn [yaθa =īt
 who.M.NOM.PL it.ACC.SG so make.FUT.IND.ACT.3PL as it.NOM.SG
 asti]
 be.PRS.IND.3SG*

‘Who will so make it, as it is’

(Old Avestan. *Yasna Haptanghaiti* 35.6, from West (2011, 112))

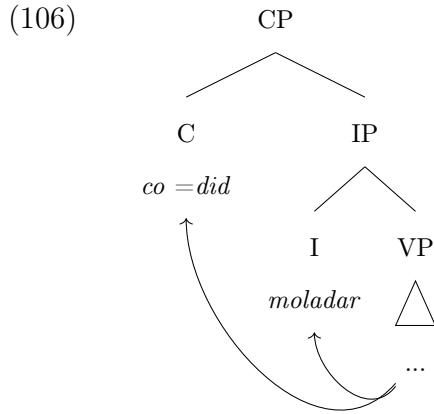
For Old Irish, WL1 movement to C likewise derives the position of infixed pronouns after an initial subordinating conjunct particle (Bate, 2024); the syntax in (106) produces the position of the pronoun *did* after the subordinator *co* in (105).

- (105) *oc precept soscéli [co=**ndid**·moladar
 at preaching.F.NOM.SG. gospel.M.GEN.SG so-that-him·praise.PRS.IND.ACT.3SG
 cách]*
 each.NOM.SG

‘Preaching the gospel so that everyone praises him’

(Old Irish. Wb. 16d1)

2.3. SYNTACTIC MOVEMENT TO CP



The subordinator may have its own prosodic weakness. In such cases, it attracts the pronouns in the syntax, yet the two will together flip positions with the next available host. This offers a derivation for Sanskrit *hí*, as in (107), and perhaps also Ancient Greek *gár*.

- (107) *Índro* *vidvāñ* *ánu = hí = tvā*
 Indra.M.NOM.SG know.PRSP.M.NOM.SG PV for you.ACC.SG
cacákṣa...
 look.PRF.IND.ACT.3SG
 ‘Because knowing Indra has looked out for you...’
 (Sanskrit. RV 5.2.8, from Hale (2017, 17))

This movement to an enclitic C head also applies to declarative and relative clauses in Old Irish. As with overt subordinators, null relative = $\emptyset^{L/N}$ and declarative = \emptyset (proposed in §2.2.1 and §2.2.3) are C heads that likewise trigger the movement of object pronouns from their original position. This is a product of movement to the formerly overt pre-Old Irish particles **eti* and **yo(m)*. It is this movement that is responsible for the long-recognised

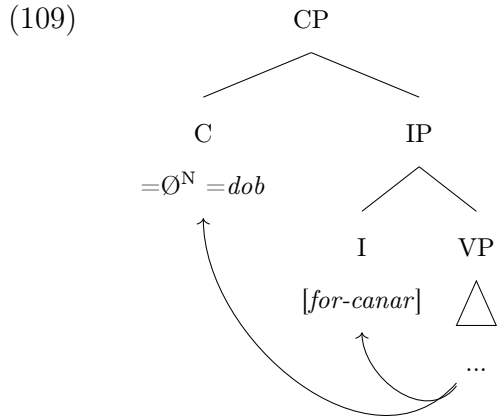
2.3. SYNTACTIC MOVEMENT TO CP

‘infixed’ position of object pronouns after the initial [X] slot of the verbal complex. In clauses with null $=\emptyset^{L/N}$ and $=\emptyset$, the clitic chain of ‘C + pronoun’ will trigger prosodic inversion with the next available host, be it a negator or a preverb. Since it follows the relative C head, the pronoun, not the verb, will be affected by the mutation of $=\emptyset^{L/N}$. We observe this in (108a), in which the preverb *for* hosts the chain $=\emptyset^{L/N}=dob$, and *dob* shows the nasal mutation.²⁹

- (108) a. *a forcital*
 the.N.NOM.SG teaching.VN.N.NOM.SG
 [*for=ndob·canar*]
 PV-you(PL)·teach.PRS.IND.PASS.3SG.REL
 ‘The teaching by which you are taught’
 (Wb. 3b23)
- b. *sechi chruth [do=nd·rón]*
 whatever.3SG manner.M.NOM.SG PV-it·do.PRS.SBJ.ACT.1SG
 ‘In whatever manner that I may do it’
 (Wb. 5b18)

²⁹The pronoun was most likely affected by the leniting mutation too, but this is not reflected in spelling, since the Class C pronouns of relative clauses begin with *d-* and Old Irish orthography uses ⟨d⟩ for both unlenited /d/ and lenited /ð/.

2.3. SYNTACTIC MOVEMENT TO CP



This ‘syntacto-prosodic’ approach also derives the Gothic order *ga-u-hva-sēhwi* in (98h).

- (110) *frah* *ina* [*ga =u =hva* *sēhwi*]
 ask.PST.IND.3SG he.ACC.SG PRT Q something see.PST.SBJ.3SG
 ‘He asked him whether he saw anything’

(Gothic. Mk. 8.23, from Wackernagel (1892))

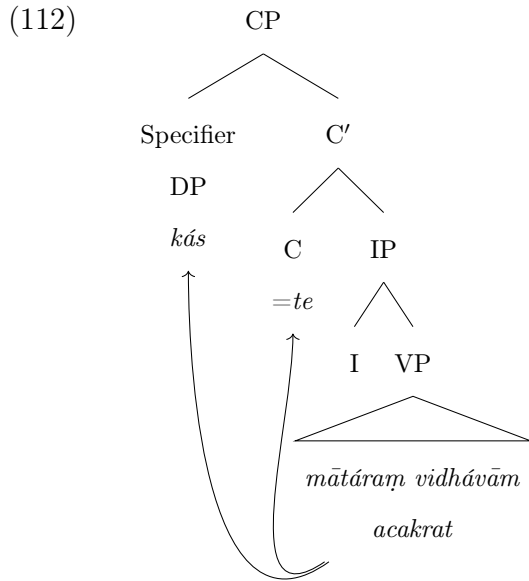
If we treat the element *u* as an enclitic interrogative C head and the indefinite *hva* as adjoined to C through movement, we can bring in prosodic inversion to move this string around around the perfective particle *ga*. This analysis for Gothic *u* therefore matches the analysis of Sanskrit *hi* and the Pre-Old Irish particles **eti* and **yo(m)* (see §2.2.3 and §2.2.1).

Furthermore, we can consider the C that triggers WL1 movement as that same functional component that is responsible for *wh*-movement too, as put forward in §2.3.1. This accounts for the ‘blocking’ of WL1 pronouns’ movement past interrogative and relative

2.3. SYNTACTIC MOVEMENT TO CP

elements; if one is present, it moves to the Spec,CP, and will precede the enclitic pronoun and serve as its post-syntactic host, as seen in the Ancient Greek examples (69) and (100a) and the Sanskrit examples (99c), (99d) and (100b). A typical Sanskrit interrogative clause like (111) has the following syntactic derivation.

- (111) *kás* =*te* *mātáram* *vidhávām*
 who.MF.NOM.SG you.GEN.SG mother.F.ACC.SG widow.F.ACC.SG
acakrat?
 do.PPRF.IND.ACT.3SG
 ‘Who made your mother a widow?’
 (Sanskrit. RV 4.18.12)



Consequently, as we see in (99c), (99d) and (100), the WL1 pronoun will immediately

2.3. SYNTACTIC MOVEMENT TO CP

follow the *wh*-moved word. *Wh*-movement in Sanskrit moves elements of the *ká*- and *yá*-stems up into Spec,CP, while a separate operation of movement attracts WL1 pronouns to C. Initial nouns like *idhmám* in (99c) are instances of topicalisation, with the noun moved or base-generated in TopicP, higher than the CP.

This account furthermore predicts that WL1 pronouns should follow whole *wh*-phrases, in which both *wh*-word and a head noun have moved up to Spec,CP, akin to English *which thing...?*. However, this is difficult to verify or refute; there are very few unambiguous examples of moved *wh*-phrases with head nouns. As noted in the conclusions to §2.3.1, interrogative and relative elements rarely form continuous constituents with a head noun. Consequently, there are few unambiguous examples of such phrases interacting with Wackernagel's law. We do find examples of WL1 pronouns immediately following an interrogative or relative element, and preceding a noun that agrees with that element, but the examples are too short to demonstrate that the pronoun is interrupting a moved syntactic phrase. (113) demonstrates this issue; these Sanskrit examples put the *ká*-/*yá*- elements first and WL1 pronouns second, as usual, but we cannot further diagnose that the agreeing nouns and adjectives are part of the same phrase as *ká*-/*yá*-, since we have no further material with which to compare their relative position and diagnose their *wh*-movement.

- (113) a. ***kéna*** =*vā* =*te* ***mánasā***
what.N.INS.SG or you.DAT.SG mind.N.INS.SG
dāśema?
pay-homage.PRS.OPT.ACT.1PL
'Or with what mind shall we pay homage to you?'
(Sanskrit. RV 1.76.1)

2.3. SYNTACTIC MOVEMENT TO CP

- b. *yás* = *te* *stánaḥ* *śaśayáḥ*
 which.M.NOM.SG you.GEN.SG breast.M.NOM.SG inexhaustible.M.NOM.SG
 ‘Which inexhaustible breast of yours...’

(Sanskrit. RV 1.164.49)

There is tentative evidence though that phrasal *wh*-movement can happen with prepositions. Ancient Greek has prepositions, which must precede the interrogative or relative word that they govern. In these very rare cases, the WL1 pronoun follows the phrase.

- (114) a. *eis hó* = *ké = s'* *é álokhon*
 into which.N.ACC.SG PRT you.ACC.SG or wife.F.ACC.SG
poiésetai
 make.FUT.IND.MED.3SG
 ‘Until he will make you his wife’

(Homeric Greek. *Odyssey* 3.409)

- b. *pròs hó* = *tis* *péphuken*
 to which.N.ACC.SG someone.NOM.SG produce.PRF.IND.ACT.3SG
 ‘For which someone has made (them)’

(Classical Greek. Plato, *Republic* 4.423)

These examples might support the account, since the ‘preposition + relative pronoun’ phrase can be taken as additional evidence that the target of *wh*-movement is higher than the target of WL1 pronouns. If so, the phrase’s underlying location in Spec,CP then produces its expected pre-pronoun position in (114). However, this analysis runs into the question of the prosodic status of Ancient Greek prepositions; if they are proclitic on the noun or pronoun that they govern, then they form a single prosodic unit that hosts the

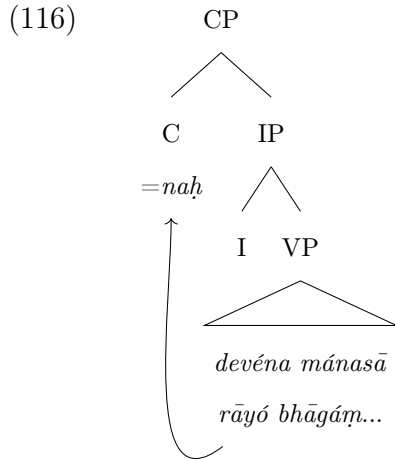
2.3. SYNTACTIC MOVEMENT TO CP

enclitic pronoun. This is to say, there are both syntactic and prosodic ways to derive the position of the WL1 pronoun in (114), and explain why it is not enclitic on the initial preposition alone. This question of how pronominal movement to C interacts with moved phrases must therefore remain unanswered.

Moving on now to other types of clause, since we have expanded the function of C to clause typing in general, we can identify WL1 movement in main declarative and imperative clauses too. As seen in §2.2, declarative clauses in the seven Indo-European languages are not typically marked by a declarative particle. Yet we may still posit a C head that is lexically null, because it still attracts a WL1 pronoun. In the case of a null C, the pronoun can trigger prosodic inversion with the nearest host from the syntax below the level of C. This may interrupt a phrasal constituent, as Latin *illum* does to the subject *dī omnēs* in (98a) and *naḥ* does to *devéna mánasā* in (115).

- (115) *devéna* =**no** *mánasā* *deva* *soma*
divine.N.INS.SG we.DAT.PL mind.N.INS.SG god.M.VOC.SG soma.M.VOC.SG
rāyó *bhāgáṃ* ... *abhí yudhya*
wealth.M.GEN.SG part.M.ACC.SG PV fight.PRS.IMP.ACT.2SG
'With your divine mind, oh god, oh soma, fight for a share of the wealth for us'
(Sanskrit. RV 1.91.23)

2.3. SYNTACTIC MOVEMENT TO CP



Moreover, if we assume that there is a C in early Indo-European imperative clauses too that attracts the imperative verb, we gain a reason for why the imperative verb in Ancient Greek and Sanskrit may host the enclitic pronouns, as in (99a) and (94). It does so, because both constituents have moved to C. Likewise, movement of object pronouns to C can account for the distinct behaviour of imperative clauses in Old Irish. While they maintain the typical VSO order, a notable difference is that the imperative verb nearly always appears in its dependent form, with conjunct inflection for simple verbs and with stress on the first syllable of simple and compound verbs; it does so in the absence of a preceding conjunct particle. The division of the verbal complex into [X] and [Y Z W] does not seem to apply to imperative clauses.

- (117) a. *gaib* *do chuil* *i =sin* *charcair*
 take.IMP.ACT.2SG your corner.F.ACC.SG in the.F.DAT.SG prison.F.DAT.SG
 ‘Take your corner in this prison’

2.3. SYNTACTIC MOVEMENT TO CP

(Old Irish. Sg. 229)

- b. *léic* *uáit* *inna* *biada* *mílsi*
 leave.IMP.ACT.2SG from.2SG the.N.ACC.PL food.N.ACC.PL sweet.N.ACC.PL
et tomil *inna* *hí* *siu*
 and consume.IMP.ACT.2SG the.N.ACC.SG PRT this
do·mmeil *do* *chenél*
 consume.PRS.IND.ACT.3SG.REL your people.N.NOM.SG
 ‘Cast away the sweet foods and consume this that your people consume’

(Old Irish. Wb. 6c7)

- c. *taibred* *cách* *airmitin* *dí*
 give.IMP.ACT.3SG everyone.NOM.SG honour.F.ACC.SG from
alailiu *et nách·taibred* *do feisin*
 other.M.DAT.SG and not-it-give.IMP.ACT.3SG to himself
 ‘May everyone give honour to one another and not give it to himself’

(Old Irish. Wb. 5d16)

The compound verb *tomil* in (117b) is prototonic, despite the lack of a preceding conjunct particle. The imperative compound verb *taibred* in (117c) appears so both with and without the negator *nách*. Imperative clauses display the deuterotonic forms of declarative main clauses *only* when the clause contains an object pronoun. This must be enclitic on the initial negator *ná* or the first preverb, in which case an imperative compound verb will be deuterotonic. Suffixed pronouns are not an option for simple imperative verbs (García-Castillero, 2020, 217).

- (118) a. *imm=a·n·imcab*
 PV-him·avoid.IMP.ACT.2SG
 ‘Avoid him’

2.3. SYNTACTIC MOVEMENT TO CP

(Old Irish. Wb. 30d20)

- b. *nach=a.telcid* *fwirib*
 not-them-pass.IMP.ACT.2PL over.2PL
 ‘Do not let them pass you’

(Old Irish. Wb. 15d4)

Considering this in light of the movement of WL1 in Indo-European more widely, the reason for the behaviour becomes clear. The Old Irish imperative clause again contains a C head, which attracts object pronouns within the syntax. Yet, unlike declarative =Ø and relative =Ø^{L/N}, this imperative C was never lexically overt; it has no post-syntactic, prosodic effect on the verb in IP. It is therefore only object pronouns, in need of a prosodic host, that trigger prosodic inversion with the first syllable of the verb and produce the prosodic gap within a compound imperative verb.³⁰

Lastly, adopting this analysis for WL1 pronouns also offers a diagnostic for the underlying structure of non-finite clauses. In Ancient Greek at least, enclitic pronouns stand to the left in non-finite clauses too, especially with the Accusative with Infinitive construction.

- (119) a. *epikataspháxai* =*min* *keleúōn* *tô_i*
 slay-on.AOR.INF.ACT he.ACC.SG urge.PRSP.M.NOM.SG the.M.DAT.SG
nekrô_i
 corpse.M.DAT.SG
 ‘... Urging him to kill him over the corpse’

³⁰By adopting a derivation with WL1 to C for the Old Irish finite clause in general, we gain a potential synchronic motivation for the distribution of the three classes of infixed pronouns. These classes show sensitivity to clause type, with Class C pronouns for example (e.g. *dom*, *dot*, etc.) appearing in relative clauses and after the interrogative particle *in* (Stifter, 2009, 79). We can derive this morphological variation through a form of agreement between the clause-type feature of the C head and the adjoined pronoun.

2.3. SYNTACTIC MOVEMENT TO CP

though it concerns the position of clitic pronouns in the Romance languages, Uriagereka likewise adopts a syntacto-prosodic derivation for them, and posits that these pronouns move up to a high functional position in the syntax (“F”). He defines the role of F as follows:

“My view is that F syntactically encodes a speaker’s or an embedded subject’s point of view. This is what allows attribution of reference, loaded descriptions (e.g., epithets), emphasis, and so on. We may think of F as a point of interface at LF between the competence levels of syntax and the performance levels of pragmatics.”

(Uriagereka, 1995a, 15)

Certain pronouns target F because of an “intrinsic defectiveness” on their part; they are deictic and lack certain features, such as a set grammatical person. Unlike nouns, they gain their reference according to current context and point of view. For Uriagereka, they gain such features via their movement up to the locus of speaker reference in the syntax, “F”. As the syntactic component “associated with specificity and point of view”, there in F the pronouns receive their reference. What Uriagereka offers here is an account of the word order of clitic pronouns that is motivated by their syntactic and semantic minimality; their prosodic weakness is a coincidence, not a factor. Uriagereka is not alone in proposing this interface component between speaker and syntax; for Bianchi (2003), high functional heads of the clause encode the details of the “speech event”, namely the circumstances (i.e. spatial/temporal coordinates) and participants (i.e. addresser/addressee) in the production of the clause. Specifically, Bianchi (2003, 7) identifies the Fin head of the Rizgian (1997)

2.3. SYNTACTIC MOVEMENT TO CP

Split CP is the interface between the speech event and the syntax; Fin is the “centre of deixis”, which supplies person features (first, second, third) for any pronouns present.

In this may lie the origins of Wackernagel’s law, with the movement of pronouns in Proto-Indo-European to C, where they might gain their reference and person features, according to the speech event and perspective of the speaker. Granted, the chapter so far has identified C as the syntactic locus of clause type and sub-/superordinate status, not of deixis and speaker perspective. However, these features are not so separate. Clause type is linked to speaker attitudes towards the truth of the clause; the use of a declarative clause presupposes belief in a particular truth value; interrogative clauses presuppose that a particular truth value is unknown and also sought; imperative clauses presuppose that the speaker wants a particular event to be true, and obliges the addressee to make it happen. The function of unitary C may therefore extend beyond clause typing to encoding the speaker’s general relationship to the overall clause, including their attitudes towards its truth, its function as a speech act, and its context as an event. Speculating further, this arrangement of ‘weak’ elements towards the start of the clause may have been the first step towards the clustering of Wackernagelian elements (both WL1 and WL2) in the clausal left periphery that we observe in the descendant languages, despite the varying synchronic causes for that behaviour.

2.4 CONCLUSIONS: THE CASE FOR PIE C

The case for the existence of the syntactic category C across early Indo-European has, like this chapter, two main parts: evidence from lexical items and evidence from movement. Both sources of evidence are important. The category C is a common thread through the plethora of words that mark their clause as subordinate to another, and as a concept can unite the etymologically diverse set of markers of interrogativity. Moreover, the proposal of C ties up many cases of movement to the left periphery of the clause, especially if we first assume a general function of C, beyond simply subordination.³¹ Through different syntactic features of clause type and status, C attracts elements from lower in the syntactic structure to help express those features in the visible surface order. Thus it offers us an explanation for the distinct position of *wh*-material, relative words and phrases, and the verb in interrogative and imperative clauses. It likewise helps to derive the well-known Indo-European phenomena known collectively as Wackernagel's law.

With such a strong case for the presence of C as a common feature of Indo-European syntax, the reconstruction of C as a functional category and a component of the Proto-Indo-European (finite) clause becomes a natural step to take. Given its presence and similar behaviour across the languages examined, from Old Irish to Vedic Sanskrit, we have gone beyond what parallel development or inter-linguistic contact can explain. The sundry sources of evidence for C across the language set satisfies the principle of the 'majority

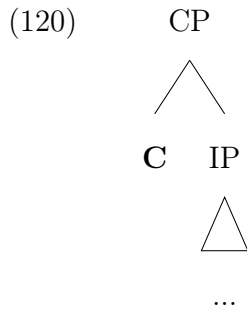
³¹With a terminological sleight of hand, consider the label 'C' in this thesis to short not for '(C)omplementizer', but rather '(C)lause'.

2.4. CONCLUSIONS: THE CASE FOR PIE C

rule' in §1.2.1 that guides traditional reconstruction, as well as the 'elder rule' on the basis of the Vedic, Ancient Greek and Latin evidence. C is too consistently present in early Indo-European clauses to be something other than an inheritance from the proto-syntax.

Furthermore, deriving the data discussed in this chapter also do not require a more complex syntactic conclusion. The five general phenomena outlined in the introduction (clause-typing elements, the syntactic distinction between main and subordinate clauses, *wh*-movement, certain instances of verbal movement, Wackernagelian second-position pronouns) each find an explanation in a Proto-Indo-European CP, which contains both the head C and the position Spec,CP. The relatively stable word-order position and complementary distribution of markers of clause type best supports a unitary C, while Spec,CP offers a single landing site for *wh*-movement. A more articulated syntax for these five phenomena, in the manner of cartography and Rizzi (1997), would find itself at the sharp end of Occam's razor. That is to say, a unitary CP is a sufficient analysis. Moreover, the theoretical benefits of adopting a CP-syntax speak for themselves; without C, we would lose a simple syntactic motivation for the different lexical and word-order phenomena discussed.

All in all, we may conclude that Proto-Indo-European had an underlying system of syntax that built finite clauses to include a C component; this C was part of a left-headed CP and took the remainder of the clause as its complement, thereby producing its usual clause-initial position. At least one of its functions was convey the type of the whole clause.



All of this applies to finite clauses at least; the evidence for C in Indo-European non-finite clauses is considerably weaker and varies from language to language. There is tentative support for a similar functional structure in some non-finite constructions, but the general lack of lexical markers of subordination make diagnosing that structure difficult.

Aside from this syntactic evidence, it is reasonable to ask whether and how this functional category was lexically expressed in Proto-Indo-European. This is a fair question, because the ‘visibility’ of C in the linguistic input would prompt C to be acquired by learners of the language, if we do not assume that C is somehow an innate feature of their syntax³². How then was C detectable and therefore learnable? The awkward fact remains that traditional lexico-phonological reconstruction has uncovered no common lexical items that we might say were exponents of C specifically in Proto-Indo-European (Comrie, 1998, 81). In other words, we have no reconstructible *ur*-subordinators.³³ Movement, such as

³²To be clear: this is not an assumption of this thesis. To propose a specific functional category like C as innate may seem like a simple explanation, but it demands a very syntactically specific model of whatever in our linguistic capacities may be genetically encoded. Aside from the fact that this would need to be neurologically substantiated, it is also not in keeping with current Minimalist views of what in syntax is innate, such as the general operation of Merge (cf. Chomsky, 1995).

³³With the possible exception of the element **h₁e* ‘if’.

2.4. CONCLUSIONS: THE CASE FOR PIE C

wh-movement and the fronting of imperative verbs, might therefore be reconstructed as the only cue in the linguistic input for acquiring the functional category of C. In clauses both with and without movement to the CP, we may have to conclude that the C head was always lexically null.

However, the syntactic case for C and subordinators is so strong that it prompts us to re-evaluate the lexical lack. As mentioned in §2.2.4, there is a considerable time gap between Proto-Indo-European and our earliest documents for the later Indo-European languages, time enough for C-elements to be born and to die, as part of the natural cycle of grammaticalisation. Specific lexical items may go through this process and disappear into morphophonological oblivion much more quickly than the syntactic category to which they belong. With this in mind, an absence of common and reconstructible C-elements becomes less problematic, although we should also return to the lexical evidence to see if any traces of such ancient elements might still survive; Sanskrit *hí* and Ancient Greek *gár* may be two examples of an older stage of subordinators, reduced to enclitics by the time of our sources. All this is to say that Proto-Indo-European may well have had clause-typing, lexically overt C-elements, even if it is only syntax that points us to their existence.

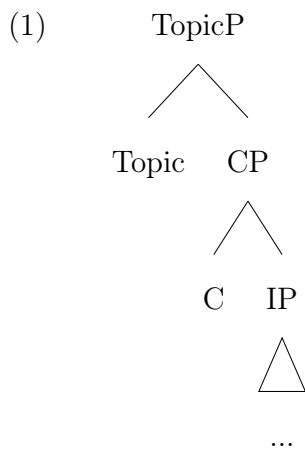
BEYOND C: TOPIC, FOCUS AND
DISCOURSE-CONFIGURATIONALITY IN
PROTO-INDO-EUROPEAN

3.1 INTRODUCTIONS AND CHAPTER AIMS

3.1.1 CHAPTER AIMS

This task of this chapter is to expand on the syntactic structure reconstructed in Chapter 2, in order to account for a variety of other left-peripheral behaviour in the early Indo-

European finite clause. In brief, this behaviour cannot be successfully accounted for with only CP, but rather requires an additional syntactic component characterised by a role in discourse information. On the basis of the comparative evidence, the ultimate reconstruction for Proto-Indo-European also includes such a component, a lexically null head labelled ‘Topic’ within a TopicP.



This syntax is specifically intended to derive finite clauses; whether it accounts for non-finite clauses too remains to be seen. It may be that non-finite clauses in Indo-European also display discourse-configurationality, but, following the hesitant syntactic conclusions for them in Chapter 2, identifying and deriving that configurationality across the language set would be a serious challenge. Non-finite clauses are not further addressed.

A concurrent goal is to demonstrate that to reconstruct a more articulated structure than (1), in particular all the components of the Split CP of Rizzi (1997), is empirically unnecessary for the language set, while adopting it *a priori* for early Indo-European is

3.1. INTRODUCTIONS AND CHAPTER AIMS

theoretically unsound. This is to say, we should not automatically adopt these influential syntactic configurations, but rather follow the lexical and word-order evidence; in the language set of this thesis, the evidence for the full Rizgian syntax of Force, Topic, Focus and Fin is found to be lacking. The chapter begins by introducing the idea of discourse configurationality and its relevance for the word order of the early Indo-European finite clause.

3.1.2 INTRODUCING INDO-EUROPEAN DISCOURSE CONFIGURATIONALITY

A common observation about early Indo-European languages is the flexibility of their attested word order. The widespread word order variation leads us to suppose that such flexibility was permitted by their grammar, that is, somehow encoded into it. Of the seven in the language set, this is especially true of Latin, Ancient Greek and Sanskrit. Any arrangement of the lexical components of the clause (e.g. verbs, nouns, adjectives) is in principle possible for these three languages. They furthermore display considerable use of discontinuous constituents (hyperbaton). There are six possible relative orders of the key constituents of subject, verb and object, and (2) illustrates all six in the same Latin text by the same author.

- (2) a. *Nēmō* *enim caelum* *caelum*
no-one.M.NOM.SG for heaven.N.ACC.SG heaven.N.ACC.SG
putat ... *sed omnēs* *opertīs*
believe.PRS.IND.ACT.3SG but all.M.NOM.PL close.PRFP.M.ABL.PL

3.1. INTRODUCTIONS AND CHAPTER AIMS

oculīs bona sua computant (SOV)
 eye.M.ABL.PL good.N.ACC.PL own.N.ACC.PL count.PRS.IND.ACT.3PL
 ‘For no one considers heaven to be heaven ... But with closed eyes they all
 count up their good things’

(Latin. Petronius, *Satyricon* 44)

- b. *numquid alius scit hanc*
 Q other.M.NOM.SG know.PRS.IND.ACT.3SG this.F.ACC.SG
conditūram vitreōrum? (SVO)
 making.F.ACC.SG glass.N.GEN.PL
 ‘Does anyone else know how to make such glass?’

(Latin. Petronius, *Satyricon* 51)

- c. *Sibyllam quidem Cūmās ego ipse*
 Sibyl.F.ACC.SG indeed Cumae.F.GEN.SG I.NOM.SG self.NOM.SG
oculīs meīs vidī in ampullā
 eye.M.ABL.PL my.M.ABL.PL see.PRF.IND.ACT.1SG in cage.F.ABL.SG
pendēre (OSV)
 hang.PRS.INF.ACT
 ‘The Cumaean Sybil I saw with my own eyes, hanging in a cage’

(Latin. Petronius, *Satyricon* 48)

- d. *bene mē admonet domina mea*
 well I.ACC.SG remind.PRS.IND.ACT.3SG mistress.F.NOM.SG my.F.NOM.SG
 (OVS)

‘Yes, my lady reminds me’

(Latin. Petronius, *Satyricon* 66)

- e. *habet haec rēs pānem*
 have.PRS.IND.ACT.3SG this.F.NOM.SG thing.F.NOM.SG bread.M.ACC.SG
 (VSO)

‘This thing has bread’

3.1. INTRODUCTIONS AND CHAPTER AIMS

(Latin. Petronius, *Satyricon* 46)

- f. *postquam negāvit,* *iussit* *illum*
after deny.PRF.IND.ACT.3SG order.PRF.IND.ACT.3SG that.M.ACC.SG
Caesar *dēcollārī* (VOS)
Caesar.M.NOM.SG behead.PRS.INF.PASS
'After he said no, Caesar ordered him to be beheaded'

(Latin. Petronius, *Satyricon* 51)

Some orders are certainly more common than others; in the *Satyricon* specifically, Polo (2004) reports that SOV is the dominant order for Petronius (66%), with SVO as the second most common (22.5%). The imbalance in frequency has typically been taken as evidence for a distinction between 'unmarked' (that is, neutral or ordinary) and 'marked' (somehow notable) word orders, which are therefore derivations away from the unmarked base. Across early Indo-European, SOV has generally held the status of the default order (cf. Lehmann, 1974), but this claim is not without challenges; Pinkster (1991, 2021) rejects claims that Latin is "basically SOV" (cf. *inter alia* Bailey, 2010).

"The important role of finite verbs as final constituents of clauses and sentences was already recognized in Antiquity ... and has been stressed by Latinists ever since. Indeed, no constituent can be found as often in that position as the (finite) verb. But there is much variation from one text or author to another and even within a single text of the same author." (Pinkster, 2021, 1025)

Regardless of whether a particular attested order is pragmatically unmarked and 'basic', the word order of Latin, Ancient Greek and Sanskrit is not random or chaotic. Different

3.1. INTRODUCTIONS AND CHAPTER AIMS

factors contribute to the attested arrangements. Some have been encountered already in Chapter 2, such as *wh*-movement and the prosodic operation of clisis, while metrical templates also play a great role in the composition of ancient verse. Yet a further factor that has perhaps the most weight in producing the final surface order in early Indo-European is discourse information – that is to say, the use of word order to situate the clause in its wider context and the ongoing discourse. Certain relative positions communicate how a given clause relates to what came before it, and what new information it is to provide. This relationship between word order and discourse information in Indo-European has long been acknowledged; the author of the Ancient Greek text *On Style* writes that “the natural arrangement” is to put first “the about which” then the “what it is”¹.

By the first, the author is referring to the **topic** of the clause. Topicality is multifaceted phenomenon; following Frascarelli and Hinterhölzl (2007, 88), three broad types of topic can be recognised. A topic may refer to entities or events communicated about prior to the current clause, and which its producer can assume to be given and part of the common ground of the interlocutors (cf. Stalnaker, 2002); this is the ‘familiar’ or anaphoric function. Alternatively, a topic may not continue an established referent, but rather purposefully offer something in opposition to it; this is a ‘contrastive’ topic. Another key use is the ‘aboutness’ quality of some topics. This conveys, simply put, the thing that the whole clause is ‘about’ in the mind of its producer; it is “what is a matter of standing or current interest or concern” (Strawson, 1964, 92).² The qualities of anaphora, contrast

¹“*Têi phusikêi táxei ... tò perì hoû ... hò toutó estin*”, 4.199, from de Jonge (2007, 6).

²For Paul (1886, 124), an aboutness topic is often the “psychological subject” of a clause, which may differ from its grammatical subject.

3.1. INTRODUCTIONS AND CHAPTER AIMS

and aboutness frequently combine in the same topical constituent of a clause, but also can remain separate; new aboutness topics are not given information, and therefore are non-anaphoric.

A methodological matter needs addressing: how we identify topics in our sources. The primary method is a triangulation of a clause's word order with its wider context. Our attested clauses do not exist in a vacuum; they usually form part of longer, multi-clausal passages, within which we can detect the flow of old and new information. The three types listed above are of increasing difficulty to identify on this basis. The referent of a familiar and continuing topic is usually recognisable in the preceding passages, while the topic itself may be pronominal, to avoid repeating a noun or noun phrase; to diagnose a contrastive topic, we have to find the different referent with which it is contrastive; and for an aboutness topic, we can only suppose what the producer of the text had as their 'current interest'.

Infrequent word orders, such as pre-subject objects in English, moreover offer evidence of topicality. For anaphoric topics, an early position for the clausal topic is expected; by coming first in the chronological sequence, it links its clause to what came before. Yet, in the absence of a clear unmarked order for the early Indo-European clause and stable elements (except subordinators), identifying topics on the basis of clause-initiality alone becomes circular; they are topics because they stand initially, and they stand initially because they are topics. Evidence from word order should therefore be combined with contextual reasons for a topical status, as well as with topicalising words; certain prepositions (e.g. Latin *dē* 'from, of, concerning') link their noun phrase to the ongoing

3.1. INTRODUCTIONS AND CHAPTER AIMS

discourse (cf. Danckaert, 2011, 55-59). Individual Latin words like *intereā* ‘meanwhile’ and *praetereā* ‘besides’ also serve a scene-setting function, situating the new clause in the general context; these ‘scene-setters’ therefore have a topical quality too (Frascarelli and Hinterhölzl, 2007).

After initial topics, what follows is often new information about the topic, the communication of which is the primary purpose of the clause. This divides the clause into two parts: topic/comment or old/new (or theme/rheme, after the Prague school). The second part contains the **focus** of the clause. This is a constituent entity typically assumed not to be known already, and so contributes key information about a given topic. Specifically, the focus is “the most important or salient parts of what we say about the topical things” (Dik, 1997, 310) and it is often accompanied with a distinct prosodic salience. This division of the utterance has been identified in a great many languages. Yet languages differ as to the extent to which they can modify their word order to reflect it (Kiss, 1995b). English represents a more inflexible type, though it is capable of object fronting for contrast (e.g. *that car, I like*). Latin, Ancient Greek and Sanskrit are towards the opposite extreme; they are prototypical ‘discourse-configurational’ languages, with word order that is consistently set to communicate the topical and focal status of certain constituents.

Because this information is expressed in word order, and also interacts with syntactic constraints (e.g. anaphoric binding, case marking), generative analysis of discourse-configurational languages has sought to derive this syntactic behaviour from the underlying structure (cf. Chomsky, 1977). This is to say, topics in the surface word order may arise from a stable ‘Topic’ head, and likewise a ‘Focus’ for foci. While discourse information

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

looks beyond the clause, under this view, it is nonetheless encoded within the clausal syntax itself of a language like Latin. The goal of the next section is to review the data and previous literature on discourse-configurationality in early Indo-European, with a view to proposing syntactic projections that can produce those data in the individual languages. These in turn are argued to be inherited from the *ur*-syntax of Proto-Indo-European, which contained a high Topic head and TopicP, hierarchically above CP, and a Focus position below the two.

3.2 TOPIC IN EARLY INDO-EUROPEAN AND PIE

3.2.1 TOPICS IN LATIN, ANCIENT GREEK AND VEDIC SANSKRIT

The case for (1) begins with Latin, Ancient Greek and Sanskrit, the three languages that best continue the discourse-driven variation of the ancestral grammar. We can identify old and new information in the following passage of Livy's *History of Rome*. It begins with the subject, who in this case is the previously mentioned consul Gnaeus Manlius Vulso. In (3), the cities mentioned in (a) form part of the initial *civitatium cognitīs causīs* phrase, an Ablative Absolute, which is a common Latin device for providing a setting for a clause. Then in the third part comes a list of the agreements made, each starting with a dative recipient. While these have not been specifically named before, they are contrastive examples of the aforementioned cities.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

- (3) (a) *Cn. Manlius ... causās deinde*
 Gnaeus.M.NOM.SG Manlius.M.NOM.SG case.F.ACC.PL then
civitatium, multīs inter novās rēs
 city.F.GEN.PL much.F.ABL.PL between new.F.ACC.PL thing.F.ACC.PL
turbātis, cognōvit ... (b) civitatium autem
 disturb.PRFP.F.ABL.PL learn.PRF.IND.ACT.3SG city.F.GEN.PL but
cognitīs causīs, decem lēgātī aliam
 learn.PRFP.F.ABL.PL case.F.ABL.PL ten legate.M.NOM.PL other.F.ACC.SG
aliārum fēcērunt condiōnem ... (c) nōminātīm
 other.F.GEN.PL make.PRF.IND.ACT.3PL agreement.F.ACC.SG by-name
praetereā Colophōniīs, quī in Notiō
 besides Colophonian.M.DAT.PL who.M.NOM.PL in Notium.M.ABL.SG
habitant, et Cymaeīs et Mylasēnīs
 live.PRS.IND.ACT.3PL and Cymaeian.M.ABL.PL and Mylasenian.M.DAT.PL
immunitātem concessērunt; Clāzomeniīs super
 immunity.F.ACC.SG grant.PRF.IND.ACT.3PL Clazomenian.M.DAT.PL beyond
immunitātem et Drymussam īnsulam dōnō
 immunity.F.ACC.SG and Drymusa.F.ACC.SG island.F.ACC.SG gift.N.ABL.SG
dedērunt; et Mīlesiīs quem sacrum
 give.PRF.IND.ACT.3PL and Milesian.M.DAT.PL which.M.ACC.SG sacred.M.ACC.SG
appellant restituērunt; et Iliēnsibus
 call.PRF.IND.ACT.3PL restore.PRF.IND.ACT.3PL and Trojan.M.DAT.PL
Rhoetēum et Gergithum addidērunt...
 Rhoeteum.M.ACC.SG and Gergithus.M.ACC.SG add.PRF.IND.ACT.3PL
 ‘(1) Gnaeus Manlius ... then got to know the cases of the cities that had been much
 disturbed amidst the new affairs ... (2) Having got to know the cities’ cases, the
 ten legates made different agreements for each ... (3) By name furthermore, to the
 Colophians, who live in Notium, and to the Cymaeans and the Mylasenians, they
 granted immunity; to the Clazomenians, beyond just their immunity, they also gave
 the island of Drymusa as a gift; to the Milesians, they restored that which they call
 sacred land; to the Trojans, they also added Rhoeteum and Gergithus’

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

(Livy, *History* 38.39, from Pinkster (2021, 1010))

Examples like (3) illustrate how the Latin clause tends to begin with its topical, known material. Additional indicators of topicality support this generalisation; initial phrases may be ‘resumed’ with a co-referential pronoun later in the clause (4a), and may begin with a topicalising word, such as the preposition *dē*.

- (4) a. *sed urbāna* *plēbēs*, *ea* *vērō praeceps*
 but urban.F.NOM.SG people.F.NOM.SG that.F.NOM.SG truly headlong
erat *dē multīs* *causīs*
 be.IPRF.IND.ACT.3SG from many.F.ABL.PL causes.F.ABL.PL
 ‘But the city people, they were declining for many reasons’

(Latin. Sallust, *Catiline’s War* 37, from Devine and Stephens (2019, 143))

- b. *dē quibus*, *trīs* *videō* *sententiās*
 of which.N.ABL.PL three.F.ACC.PL see.PRS.IND.ACT.1SG opinion.F.ACC.PL
ferrī
 bear.PRS.INF.PASS
 ‘Concerning such things, I see three opinions supported’

(Latin. Cicero, *On Friendship* 56)

- c. *praetereā, iuventūs*, *quae* *in agrīs* ... *praetereā*,
 besides youth.NOM.SG who.F.NOM.SG in field.M.ABL.PL besides
quōrum *victōriā* *Sullae* *parentēs*
 which.M.GEN.PL victory.F.ABL.SG Sulla.M.GEN.SG parent.M.NOM.PL
prōscriptī...
 proscribe.PRFP.M.NOM.PL
 ‘Besides, the youth, who in the fields ... in addition, those whose parents had
 been proscribed, by the victory of Sulla...’

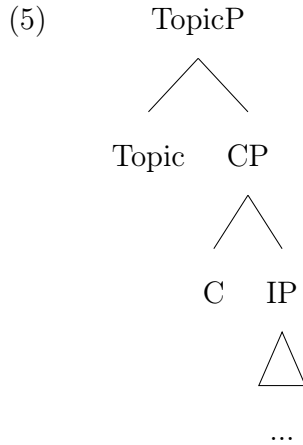
(Latin. Sallust, *Catiline’s War* 37)

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

All of these indicators can be found within our Latin sources, and they confirm the association between the clause-initial position and topicality in Latin.

Two theoretical options present themselves to account for this behaviour (cf. Cinque, 1977). One is to consider initial topics to be paratactic, that is, preceding the clause proper but fundamentally outside it. This extra-clausal material includes ‘hanging topics’ and the *nominativus pendens* construction. Under this view, initial topics are not integrated into the clausal syntax, and so do not occupy their own topicality-characterised functional projection. The alternative is to derive topics through such a component of the syntax, a ‘TopicP’, to which intra-clausal topical material can move, or in which it can be base-generated. The claim of this chapter is that both analyses should be adopted for Latin and for early Indo-European more widely. For some initial topics, the paratactic view is the stronger; for others, movement to TopicP is the better analysis; and for many, either view would be theoretically cogent. Crucially, a very high TopicP and Topic head should be proposed for the syntax of Latin.³ TopicP is even higher than the CP established in Chapter 2, since its topics precede both subordinators (in C) and interrogative and relative material (in Spec,CP).

³Also proposed for Latin by Salvi (2005), Danckaert (2011), Oniga (2014), *inter alios*, in the manner of Rizzi (1997).



To consolidate both the data and the previous literature, this thesis adopts ‘Topic’ as the working label and key quality of this syntactic component for all three languages. What the material that it hosts has in common aligns with the typical views of topicality, with the qualities of aboutness, contrast and anaphora. It is an ‘outward-looking’ component of the syntax, linking its clause to the wider context and the mental states of the producer of the clause, with regards to its psychological subject and what information is assumed to be given.

The reason for adopting both analyses for initial topics in Latin is their morphological and word-order variety. Devine and Stephens (2019) distinguish between simple, contrastive and dislocated topics. Simple topics (like *Cn. Manlius* in (3)) and contrastive (like the different cities in (3)) can be understood to be fronted to their position from within the clause; this involves movement up to Spec,TopicP in (5). Yet the third kind, dislocated topics, Devine and Stephens (2019) define as “nonclausal”; they appear less in-

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

tegrated into the structure of the overall clause. The signs of this status are the use of resumptive pronouns and the *nominativus pendens* construction, in which the initial topical constituent appears in the nominative case, unlike the pronoun that later resumes it, which instead bears the case appropriate to its role as an argument. The material that follows a dislocated topic altogether constitutes a “grammatically complete” clause (Halla-Aho, 2018, 21).

- (6) *Mercātor Siculus, quoi erant*
 merchant.M.NOM.SG Sicilian.M.NOM.SG which.M.DAT.SG be.IPRF.IND.3PL
gemini filii, / ei surrupto
 twin.M.NOM.PL son.M.NOM.PL he.DAT.SG snatch.PRFP.M.ABL.SG
altero mors optigit
 other.M.ABL.SG death.F.NOM.SG strike.PRF.IND.ACT.3SG
 ‘A merchant of Sicily, who had two sons – when one was snatched away, him did death strike’

(Latin. Plautus (possibly), *Menaechmi argumentum*, from Devine and Stephens (2019, 145))

Here the initial phrase *Mercātor Siculus...* bears nominative case, yet so does *mors*, which is the actual grammatical subject of the clause. *Nominativus pendens* constructions like that in (6) therefore stand somehow outside the clausal syntax in (5), being “isolated” and “emphatic”. The pronoun *ei* that resumes the noun phrase is a part of the clause proper and its system of case assignment, while the *nominativus pendens* is not.

One way to analyse a *nominativus pendens* is that it belongs to a separate clause, which has been elided (cf. Devine and Stephens, 2019, 144-7). It is thus a fragment of a

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

clause, the remainder of which is deleted to avoid repetition of the subsequent comment that both clauses contain. This raises the question of why the full dislocated topic and the resumptive pronoun sometimes do not match in case, if their two clauses are in essence identical. In instances like (6), it is unclear what the elided material might be and why *Mercator Siculus...* bears nominative case. An alternative is the ‘speech act analysis’, in which a dislocated topic does not belong to any clause, but rather is a complete (albeit short) speech act; “it is a mere act of reference, a bit like a vocative” (Devine and Stephens, 2019, 144). The speech act is a signal to the audience to expect imminent information on that topic. Regardless of the analysis, topics like *Mercātor Siculus...* can be economically considered to be paratactic, not integrated into the clausal syntax. Their nominative case is a default, given to a noun (phrase) that has no case-receiving role (Havers, 1926). They therefore do not require a TopicP.

Then there are topics whose syntactic status is unclear. Scene-setting adverbial topics like *intereā* and prepositional ‘as for...’ phrases with *dē* can be considered paratactic or integrated into the following clause, in the latter case being base-generated in either the head or specifier position of the TopicP in (5). They have no morphological form (i.e. case) that they gain through participation in the argument structure of the clause, and they are ill-suited to a movement analysis; we could propose that adverbs like *intereā* are base-generated in a lower position, such as within the VP, and then move to TopicP, but this seems unfounded. The same ambiguity of analysis applies to whole clauses. The two clauses of the correlative construction might be only paratactically associated, or we could

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

integrate one clause into the syntax of the other.⁴ In (7a), for example, the initial clause *quōs ferrō...* in the Spec,TopicP of *eōs nōndum...*

- (7) a. [*quōs ferrō trucidārī*
 which.M.ACC.PL sword.N.ABL.SG slaughter.PRS.INF.PASS
oportēbat] [*eōs nōndum vōce*
 ought.IPRF.IND.ACT.3SG they.ACC.PL not-yet voice.F.ABL.SG
volnerō]
 wound.PRS.IND.ACT.1SG
 ‘Those who should have been put to the sword, them am I not yet wounding
 with my voice’

(Latin. Cicero, *Catilinarian Orations* 1.9, from Probert and Dickey (2016, 1))

- b. [*quoi hominī dī sunt*
 which.M.DAT.SG man.M.DAT.SG god.M.NOM.PL be.PRS.IND.3PL
propitiū], [*eī nōn esse irātōs*
 favorable.M.NOM.PL he.DAT.SG not be.PRS.INF angry.M.ACC.PL
putō]
 think.PRS.IND.ACT.1SG
 ‘The man to whom the gods are good, with him I do not think the gods are
 angry’

(Latin. Plautus, *Curculio* 4.4, from Halla-Aho (2018, 34))

Probert (2015, 311-14) notes for Ancient Greek that this bi-clausal construction comprises “(i) something the sentence is construed as being about, a ‘topic’ (the relative clause), and (ii) something being said about this topic, a ‘comment’ (the correlative clause)”. An integrated position in Spec,TopicP is therefore appropriate for an initial relative clause, but

⁴The traditional view of early Indo-European is a paratactic one, in which all finite clauses were once syntactically independent (cf. Clackson, 2007, 171-6).

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

without clear diagnostics for integrated status, it is difficult to claim so with confidence. A location in the Spec,TopicP of another clause would at least agree with the “half-way” view of Haudry (1973), in which the correlative construction in Latin exists somewhere between total parataxis and embedding.

Instead, the best support for TopicP comes from initial topics that are more clearly a part of a clause, as affirmed by their argument role and their case marking. Best among these are topics in oblique cases (non-nominative), since this avoids the possibility of default case assignment. Both simple and contrastive topics can appear initially and with a case appropriate for their role within the argument structure of the verb. For example, in (8), Cicero’s question begins with the accusative object and a contrastive topic, namely Catiline, who contrasts with other enemies previously mentioned.

- (8) [*Catīlīnam, orbem terrae caede atque*
Catiline.M.ACC.SG world.F.ACC.SG earth.F.GEN.SG cutting.F.ABL.SG and
incendīis vastāre cupientem], *nōs*
burning.N.ABL.PL devastate.PRS.INF.ACT desire.PRS.M.ACC.SG we.NOM.PL
cōsulēs perferēmus...?
consul.M.NOM.PL bear-through.FUT.IND.ACT.1PL
‘But Catiline, who desires to lay waste to the whole world through slaughter and
burning, we consuls will tolerate?’

(Latin. Cicero, *Catilinarian Orations* 1.3, from Pinkster (1991, 75))

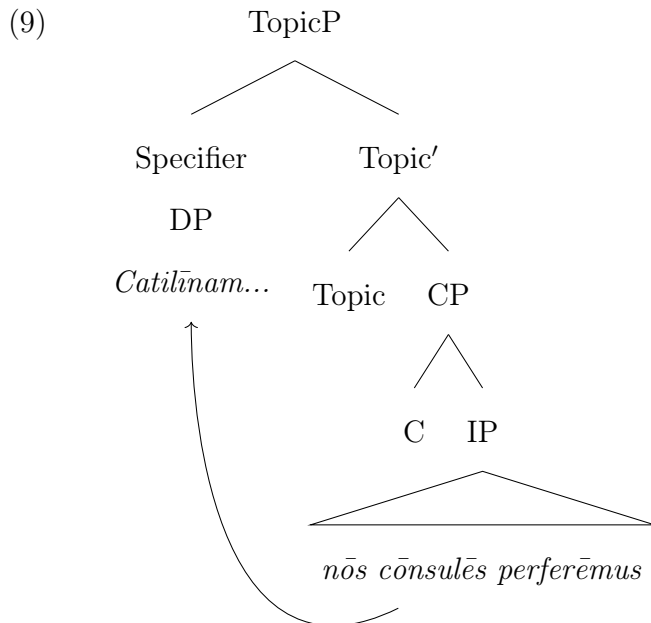
Clauses like (8) offer clear examples of initial integrated topics, since the accusative object atypically precedes the overt subject *nōs cōsulēs*.⁵ Further examples of such topics include

⁵In general, the possibility of null subjects (and objects) in the three ancient Indo-European languages

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

the dative cities in (3), which each precede their accusative objects, the resumptive dative pronoun *eī* in (6) and (7b), which precedes the nominative subject *mors* in (6), and the accusative correlate *eōs* in (7a), which stands first in the correlative clause.

In these cases, we have good grounds to propose the TopicP component to the syntax in (5), to which the topical constituents move, and which produces their clause-initial position.⁶



(9) is intentionally ambiguous about the position of the subject *nōs cōsulēs*; it may be

compounds the difficulty in identifying topics through word order.

⁶Good grounds, but not certainty; initial topics like (8) could alternatively involve a parataxis and a null object (cf. Luraghi, 2004), but accounting for the accusative case of the topic and positing the null object is a costlier derivation.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

in Spec,IP, as the nominative subject of a finite clause, or within the lower VP. The usual syntactic position of subjects in Latin, Ancient Greek and Sanskrit may in fact be in Spec,TopicP. By virtue of their common topical quality, movement to Spec,TopicP would produce their typical pre-verbal and pre-object position, except in cases like (8), when another argument is topical (cf. for Sanskrit, Lowe, 2015, 38).

Turning now to the other two ancient languages in the set, what pertains to Latin pertains also to Ancient Greek and Vedic Sanskrit. Concerning the surface order, the association between topicality and clause-initial position is again well recognised. In the Greek of Herodotus, for instance, Dik (1995) identifies sections of the clausal associated with a discourse-informational function. She proposes the following abstract word order for the Ancient Greek clause:

- (10) $P1 \quad P\emptyset \quad V \quad X$
Topic Focus Verb Remaining unmarked constituents
(Dik, 1995, 12)

Again, topics occur at the left edge, before the focus, the verb and any remaining material of the clause, occupying P(osition)1, and presumably separated in prosody by an intonation break (Fraser, 1999, 14). As with Latin, we find the same kinds of topical material, such as aboutness topics, contrastive topics and prepositional phrases with *perí* ‘around, about’.

- (11) a. *strouthòn* = *dè oudeìs* *élaben*
ostrich.M.ACC.SG and no-one.M.NOM.SG take.AOR.IND.ACT.3SG
‘But an ostrich, no one caught one’
(Ancient Greek. Xenophon, *Anabasis* 1.5.3, from Matic’ (2003, 2))

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

- b. *tò* =dè *stráteuma* *ho* *sítos*
 the.N.ACC.SG and army.N.ACC.SG the.M.NOM.SG grain.M.NOM.SG
epélipe
 fail.AOR.IND.ACT.3SG
 ‘And as for the army, the grain ran out’
 (Ancient Greek. Xenophon, *Anabasis* 1.5.6)
- c. *kai* [*perì toû* *ekhthroû*] =dè *hē*
 and about the.M.GEN.SG enemy.M.GEN.SG and the.F.NOM.SG
autē *thésis*
 same.F.NOM.SG position.F.NOM.SG
 ‘And as for the enemy too, the same opinion’
 (Ancient Greek. Plato, *Republic* 1.335)

Subsequent scholars have modified the schema in (10). Among the fronted constituents in his corpus, Matic (2003) distinguishes between the “exclusive contrastive topic”, which is clause-initial, and the “frame-setting topic”, which stands after the first type. Moreover, Allan adds ‘extra-clausal’ constituents to (10), which are “loosely associated” with the constituents within the clause (2014, 3); these include examples of the *nominativus pendens* construction.

- (12) *en toutōi* =dè *kai* [*hoi* *ek toû pedíou*]
 in this.N.DAT.SG and and the.M.NOM.PL from the.M.GEN.SG plain.M.GEN.SG
hoi *mèn peltastai* *tôn Hellénōn*
 the.M.NOM.PL and peltast.M.NOM.PL the.M.GEN.PL Greek.M.GEN.PL
drómōi *étheon*
 running.M.DAT.SG run.IPRF.IND.ACT.3PL
 ‘Meanwhile, as for those from the plain, the Greeks’ peltasts charged at a run...’
 (Ancient Greek. Xenophon, *Anabasis* 4.6, from Recht (2015, 24))

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

Example (12) includes the topic *hoi ek toû pedíou*, which appears in the nominative case.⁷ While the discourse-informational arrangement of an Ancient Greek clause has the potential to appear very complex, there is really one system of configuration for topics and another for focus, and each system may exhibit more than one ‘value’ simultaneously. Allan argues that the observed variety of topics and foci can all be reduced to “two separate systems of constructional templates” that “belong to different dimensions in the organization of information structure” (Allan, 2014, 17). Within the linear order, the system of topics precedes all else.

Ancient Greek therefore matches Latin, in that it consistently displays clause-initial topics, which can be more or less integrated into the clause proper. Examples of the less integrated type include the *nominativus pendens* construction (12); examples of the more integrated type include initial oblique nouns and noun phrases (11a). The latter, the “clause-integrated, valency-bound” type (Matić, 2003, 580), are best derived through fronting and syntactic movement within the clause. This is specifically to the highest specifier (Spec,TopicP) in the syntax (cf. Dal Lago, 2010; Vai, 2017).

As for Sanskrit, its discourse-informational structure is “essentially parallel” (Lowe, 2015, 38) to what Dik (1995) has proposed for Ancient Greek in (10), with the same key divisions within the material of the clause. In the second part of (13), we find the typical arrangement, with the anaphoric topical subject *sá* referring back to Agni and preceding the object and the verb.

⁷No source for this nominative case presents itself within the clause, since *hoi peltastaí* is the subject of *étheon*. The nominative may again be a default case, borne in the absence of an assigned case.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

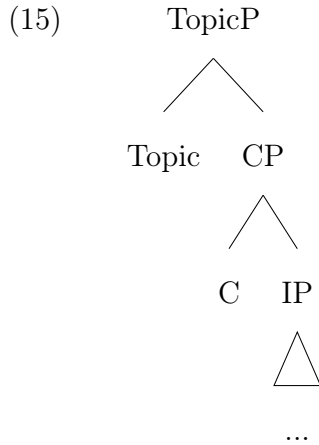
- (13) *Agnīḥ pūrvēbhir ṛṣibhir īdya ...*
 Agni.M.NOM.SG ancient.M.INS.PL seer.M.INS.PL praiseworthy.M.NOM.SG
sá devāṃ ā ihā vakṣati
 he.NOM.SG god.M.ACC.PL to here bring.AOR.SBJ.ACT.3SG
 ‘Agni, worthy of praise by the ancient seers ... may he bring the Gods here’
 (Sanskrit. RV 1.1.2)

Hale (1991, 2018) agrees with Lowe (and therefore Dik) about a typical Topic-Focus order for the Sanskrit clause; he even reconstructs it for the ancestral Proto-Indo-Iranian language and proposes lexical items closely associated with the different roles: the topicalising determiner and pronoun *tá-*, and the more emphatic focalising *etá-*. The latter typically follows the former when both are present in a clause; in (14), pronominal *tá-* is topical, referring back to the *asura* Svarbhanu, while the object *etá malhá* is new information. In (14), dative *tásmā* begins the clause.

- (14) *tásmā etá malhá álabhanta*
 he.DAT.SG this.N.ACC.PL dewlapped.N.ACC.PL offer.IPRF.IND.MED.3PL
 ‘For him they offered these dewlapped beasts’
 (*Taittiriya Samhita* 2.1.2.3, from Hale (2018, 5))

As with Latin and Ancient Greek, these integrated initial topics in Sanskrit can be, and have been⁸, straightforwardly derived by positing at least one topic-associated component within the underlying syntax of each language. Being very high in the hierarchical structure, the topical material that it hosts will surface in initial position. This configuration is repeated in (15).

⁸See, for instance, Hale (1987a), Ram-Prasad (2022).



There are differing views though as to the defining quality of the discourse-functional component for Sanskrit. Keydana (2011) and Lowe (2014) do not commit themselves to any characterisation beyond a “discourse functional slot”. Hale (2018) labels his highest discourse-informational component of Sanskrit syntax “Emphasis”, a change from his earlier “Top” (1987a). Hale does not elaborate on the features of Emphasis, beyond stating that material moved there is “clearly emphatic in some sense” (2018, 8). This may be true for many or even most clauses, but it is not necessarily always so.⁹ To avoid committing the initial material to an emphatic reading, and to try to consolidate the previous characterisations of the position, this thesis sticks with the label ‘TopicP’ for the topmost projection of the Sanskrit clause.

⁹Sanskrit, for one, makes use of initial introductory particles like *átha* ‘then’ and *evá* ‘so’ (Hale, 1987a, 5), and while it is possible that these are sometimes emphatic, to grab the audience’s attention, it seems too restrictive to claim that they were always so. For comparison, equivalent elements in English (like *so*) are often no more prosodically prominent than what follows, serving simply to set the scene and situate the new clause in the discourse up to that point.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

As with Latin, the TopicP in (15) offers a means to integrate the relative clause of the correlative construction, found also in Ancient Greek and especially common in Sanskrit.

- (16) a. [hósoi =gàr nân mēdetérois ksummakhoûsi] [pôs ou
 as-many for now neither.M.DAT.PL be-ally.PRS.IND.ACT.3PL how not
 polemōsesthe autoús?]
 make-enemy.FUT.IND.MED.2PL they.ACC.PL
 ‘For those who are currently allies of neither, how will you not make enemies
 of them?’

(Ancient Greek. Thucydides, *The Peloponnesian War* 5.98, from Fraser (1999, 176))

- b. Ágne [yám yajñám adhvarám]
 Agni.M.VOC.SG which.M.ACC.SG sacrifice.M.ACC.SG worship.M.ACC.SG
 viśvātaḥ paribhūr ási] [sá íd
 all-sides encompassing.M.NOM.SG be.PRS.IND.2SG he.NOM.SG indeed
 devēṣu gacchati]
 god.M.LOC.PL go.PRS.IND.3SG
 ‘Agni, what sacrifice and worship you encompass on all sides, that certainly
 goes among the gods’

(Sanskrit. RV 1.1.4, from Viti (2007, 64))

This location in the TopicP of the correlative clause provides a means and a discourse-informational motivation for the ‘adjunction’ of the two clauses noted for early Indo-European by Hock (1989), Ram-Prasad (2022, 143-9) and others.

The common syntax in (15) prompts the question of the position of TopicP relative to the CP established for early Indo-European in Chapter 2. C-elements have been shown

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

to occupy a consistently early word-order position, yet topics typically precede even the various lexical exponents of C and instances of movement to the CP. As (17) shows, we occasionally find words or a phrase preceding a subordinator in Latin and Sanskrit, and also in Ancient Greek, albeit very rarely.

- (17) a. *cōnsul* *postquam dētractārī* *certāmen*
 consul.M.NOM.SG after refuse.PRS.INF.PASS battle.N.ACC.SG
vīdit...
 see.PRF.IND.ACT.3SG
 ‘After the consul saw that battle was refused...’
 (Latin. Livy, *History* 37.39, from Pinkster (2021, 979))
- b. *equidem sciō* *iam [filius* *quod*
 indeed know.PRS.IND.ACT.1SG now son.M.NOM.SG that
amet *meus* *istanc* *meretrīcem* *ē*
 love.PRS.SBJ.ACT.3SG my.M.NOM.SG that.F.ACC.SG courtesan.ACC.SG from
proximō *Philaenium]*
 nearest.N.ABL.SG Philaenium.F.ACC.SG
 ‘Indeed I now know that my son may love that courtesan from nearby, Philaenium’
 (Latin. Plautus, *Asinaria* 1.1)
- c. *Caesarī* *cum id* *nūntiātum*
 Caesar.M.DAT.SG when that.N.NOM.SG account.PRFP.N.NOM.SG
esset...
 be.IPRF.SBJ.3SG
 ‘When it was announced to Caesar...’
 (Latin. Caesar, *Gallic War* 1.7)
- d. ... *lēgōn* *hōti [egō* *ei mē perì* *polloû*
 say.PRSP.M.NOM.SG that I.NOM.SG if not about many.N.GEN.SG

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

hēgeúmēn *mēdéna* *xeínōn*
 rule.IPRF.IND.MED.3PL no-one.M.ACC.SG foreign.M.GEN.PL
kteínein]...
 kill.PRS.INF.ACT
 ‘... saying ‘if I had not decreed most earnestly to kill no one of foreign birth...’

(Ancient Greek. Herodotus, *Histories* 2.115.4)

- e. ***mahī*** *yádi* *Dhiṣáṇā* *śiśnáthe*
 great.F.NOM.SG when Dhiṣáṇā.F.NOM.SG piercing.M.LOC.SG
dhāt
 put.AOR.INJ.ACT.3SG
 ‘When the great Dhiṣáṇā had determined to pierce (him)’

(Sanskrit. RV 3.31.13, from from Viti (2007, 117))

- f. ***Índro*** ***vidvāñ*** *ánu = hí = tvā*
 Indra.M.NOM.SG know.PRSP.M.NOM.SG PV for you.ACC.SG
cacákṣa...
 look.PRF.IND.ACT.3SG
 ‘Because knowing Indra has looked out for you...’

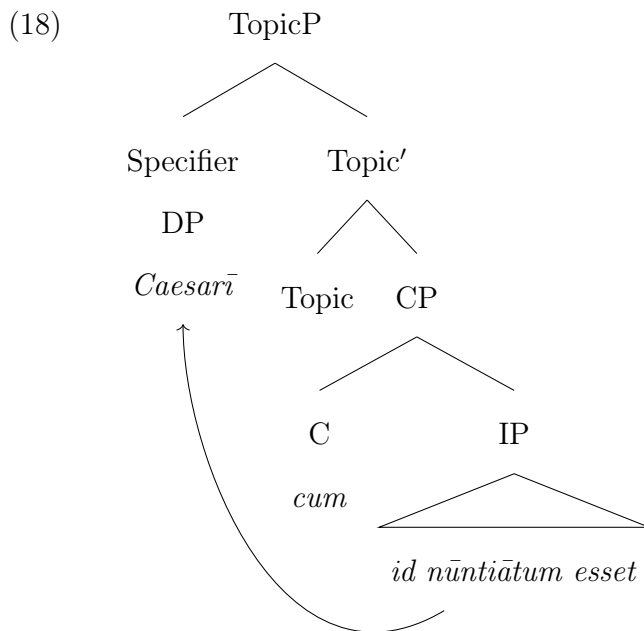
(Sanskrit. RV 5.2.8, from Hale (2017, 17))¹⁰

The words in bold are plausibly analysed as topics; for example, the consul in (17a) is the primary Roman protagonist of the passage, while *egō* in (17d), the Egyptian king Proteus, is the speaker and a subject fronted to contrast with the behaviour of the addressee, Alexandrus. Granted, different analyses are always possible for examples like those in (17). Some of the pre-subordinator words, like *cōnsul* in (17a), could instead belong to another clause or to no clause, while the clause containing the subordinator instead has a null subject. Others though are more secure cases of fronting; nominative *filius* can

¹⁰As proposed in §2.3.5, *hí* is considered to belong to the category of C, although its word order is partly determined by enclisis, here with the preverb *ánu*.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

only belong to the subordinate clause in (17b), while *mahī* in (17e) is the first word of its stanza and agrees with the post-subordinator subject *Dhiṣáṇā*. We have good grounds therefore that such topics occupy the high TopicP projection above the CP, some moved there, others base-generated.



Rare examples also exist in which initial topical material precedes an interrogative particle. In (19b), the anaphoric phrase *toûto poioñti* refers back to the hypothetical situation in the previous question.

- (19) a. *hāc* *lēge* *sublāta* *videntur*
 this.F.ABL.SG law.F.ABL.SG abolish.PRFP.F.ABL.SG see.PRS.IND.PASS.3PL

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

=*ne vōbīs* *posse* *Caesaris* *ācta*
 Q you.DAT.PL be-possible.PRS.INF Caesar.M.GEN.SG act.N.ACC.PL
servārī?
 save.PRS.INF.PASS
 ‘With this law abolished, do Caesar’s acts seem to you able to be maintained?’

(Latin. Cicero, *Philippics* 1.9, from Pinkster (2021, 1000))

b. *kaī toūto* *poioūnti* *âra áxion* =*soi*
 and this.N.ACC.SG do.PRSP.M.DAT.SG Q worthy.N.NOM.SG you.DAT.SG
zên *éstai?*
 live.PRS.INF.ACT be.FUT.IND.3SG
 ‘And by doing this, will your life be worth living?’

(Ancient Greek. Plato, *Crito* 53c)

Like subordinators, the interrogative particles *ne* (which is enclitic on the following verb, *videntur*) and *âra* are claimed in Chapter 2 to belong to the syntactic category of C. If topics can precede the exponent of the interrogative C head, we can locate them in a higher position in the syntax.

Furthermore, topical words and phrases precede material that has undergone *wh*-movement, such as in interrogative clauses. Since *wh*-movement is considered to involve movement to Spec,CP (see §2.3.1), it follows that standing in Spec,CP is not an option for the preceding topic.

(20) a. *sed Stālagmus* *quoius* *erat* *tunc*
 but Stalagmus.M.NOM.SG which.F.GEN.SG be.IPRF.IND.3SG then
nātiōnis?
 nation.F.GEN.SG
 ‘But Stalagmus, of what nation was he then?’

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

(Latin. Plautus, *Captivi* 887)

- b. *dē cohorte meā reiciēs,*
 from cohort.F.ABL.SG my.F.ABL.SG reject.FUT.IND.ACT.2SG
inquit, quid? [ista cohors
 say.PRF.IND.ACT.3SG what.N.NOM.SG this.F.NOM.SG cohort.F.NOM.SG
quōrum hominum est?]
 which.M.GEN.PL man.M.GEN.PL be.PRS.IND.3SG
 “‘You will reject them from my retinue”, he said. “What? This retinue, who’s
 it composed of?””

(Latin. Cicero, *Against Verres* 2.3.11, from Pinkster (2021, 954))

- c. *puretòs =dè tí poiêi? ... keramìs*
 fever.M.NOM.SG and what.N.ACC.SG do.PRS.IND.ACT.3SG tile.F.NOM.SG
 =dè tí poiêi?
 and what.N.ACC.SG do.PRS.IND.ACT.3SG
 ‘And a fever, what does it do? ... And a roof tile, what does it do?’

(Ancient Greek. Epictetus, *Discourses* 4)

- d. *tò =dè bêma tí =soi*
 the.N.NOM.SG and platform.N.NOM.SG what.N.NOM.SG you.DAT.SG
khresimon éstai?
 use.N.NOM.SG be.FUT.IND.3SG
 ‘And the platform, what use will it be to you?’

(Ancient Greek. Aristophanes, *Ecclesiazusae* 677, from Goldstein (2015, 121))

- e. *múkham kím =asya?*
 mouth.M.NOM.SG what.N.NOM.SG he.GEN.SG
 ‘His mouth, what [was it called]?’

(Sanskrit. RV 10.90.11, from Ram-Prasad (2022, 82))

- f. *kó =asya vīráh sadhamādam*
 who.M.NOM.SG he.GEN.SG hero.M.NOM.SG company.M.ACC.SG

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

āpa? *ām ānaṃśa* *sumatībhiḥ*
 reach.PRF.IND.ACT.3SG PV reach.PRF.IND.ACT.3SG favour.F.INS.PL
kó = *asya?*
 who.M.NOM.SG he.GEN.SG
 ‘Which hero has taken up his company? Who has shared in his favours?’
 (Sanskrit. RV 4.23.2, partly from Halpern (1995, 5))

Likewise, topics can precede markers of relative clauses. It is at least well attested in Sanskrit. The stanza of (21a) concerns the beneficial consequences of bringing *idhmám* ‘fuel, firewood’ to the worship of Agni. Similarly, *rayiḥ* ‘goods, riches’ has been mentioned twice in the hymn prior to the stanza of (21b), so it is an established entity and benefit of the hymn’s addressee, the ritual Soma drink. However, no sure examples of ‘Topic-Relative’ orders have been found so far in Latin, and even the rare example in the Homeric Greek of (21d) may be inadmissible if we ascribe its order instead to the metrical patterns of epic verse and not to the features of ordinary syntax.

- (21) a. *idhmám* *yás* = *te* *jabhárac*
 fuel.M.ACC.SG which.M.NOM.SG you.DAT.SG bring.PRF.SBJ.ACT.3SG
chaśramāṇáh
 exert.PRSP.M.NOM.SG
 ‘He who, exerting himself, will have brought you firewood’
 (Sanskrit. RV 4.12.2, from Hale (2018, 11))
- b. *rayím* *yéna* *vānāmahai*
 wealth.M.ACC.SG which.M.INS.SG win.AOR.SBJ.MED.1PL
 ‘Through which we will win wealth’
 (Sanskrit. RV 9.101.9)

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

- c. *āvīr yásmāi cārutamo babhūtha*
 openly which.M.DAT.SG most-lovely.M.NOM.SG be.PRF.IND.2SG
 ‘To whom you have appeared most lovely’
 (Sanskrit. RV 5.1.9)
- d. *autàr egò theós eimi [diamperès*
 but I.NOM.SG god.M.NOM.SG be.PRS.IND.3SG continuously
hế =se phulássō en pántessi
 who.F.NOM.SG you.ACC.SG guard.PRS.IND.ACT.1SG in all.M.DAT.PL
pónois]
 toil.M.DAT.PL
 ‘But I am a god, who guards you unfailingly in all your struggles’
 (Ancient Greek. Homer, *Odyssey* 20.47-8)

Compared with the greater occurrence of the order in Sanskrit, the absence we can explain by taking into account the productivity of clausal embedding in Latin and Ancient Greek. Relative clauses in the two languages are consistently placed within the main clause, after their antecedent, whereas Sanskrit *yá*-clauses appear at either its left or right edge. Such embedding may have blocked fronting beyond the relative marker to avoid ambiguity; it would be unclear to which clause the fronted material belongs.

Examples like (20d)-(21a) also show that initial topics precede the position of Wackernagelian clitic pronouns. For instance, the object pronoun *se* in (21d) follows the relative pronoun *hế*, coming third in the order of the clause. Such orders are further evidence for a syntactic factor in the derivation of Wackernagelian pronouns, which demonstrably do not indiscriminately stand enclitic on the first word of their clause. As proposed in Chapter 2, we can locate *hế* and *se* in the CP of that clause, while the adverb *diamperès* stands in the

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

higher TopicP. Furthermore, it was also claimed in Chapter 2 that Latin and Ancient Greek optionally move imperative verbs to C, producing a verb-early word order in imperative clauses.

- (22) a. *pròs taûta mē psaiúsēi =tis*
 towards this.N.ACC.PL not touch.AOR.SBJ.ACT.3SG someone.M.NOM.SG
Argeiōn emoû
 Argive.M.GEN.PL I.GEN.SG
 ‘Therefore, let no one of the Argives touch me’

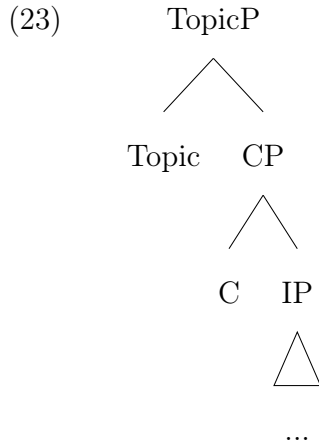
(Ancient Greek. Euripides, *Iphigenia in Aulis* 1559, from Hale (2017))

- b. *Agnīm vardhantu =no gírah*
 Agni.M.ACC.SG increase.PRS.IMP.ACT.3PL we.GEN.PL hymn.M.NOM.PL
 ‘May our hymns increase Agni’

(Sanskrit. RV 3.10.6)

Such fronted verbs still frequently follow initial topical material, such as *pròs taûta* (‘in view of these things, therefore’) in (22a) and *Agni*, the addressee of the hymn in (22b). Since the imperative verbs in (22) also precede Wackernagelian clitic pronouns, we may locate verb and pronoun in C, while in the initial material stands higher in a topical projection.

All in all, the evidence considered leads us to posit that the three early Indo-European languages of Latin, Ancient Greek and Sanskrit shared (at least) the underlying syntactic structure in (23). Within the clear clause-initial behaviour of their topical material, there is a subset of topics that are best derived through a dedicated Topic head and TopicP projection, which can motivate the topicalisation and fronting of clausal constituents to the left periphery.



This is very similar to the analysis for Sanskrit of Hale (1987a), though it differs slightly in the details; C here is a more general locus of clause type and status. (23) holds in particular for declarative, imperative and interrogative main clauses across the three languages, but less so for types of subordinate clause. Compared with Sanskrit (see the examples in (17) and (21)), Latin and Ancient Greek clauses in which topical material precedes a subordinator or a relative element seem to be very rare.

An outstanding issue is the derivation of initial sequences of more than one topic, and whether the syntactic structure in (23) needs to be able to generate clauses that begin with a string of topics. These may be of different kinds, such as scene-setting phrases, contrastive topics and aboutness topics. Such multi-topic clauses are certainly attested across Latin and Ancient Greek.

- (24) a. *[intereā] [Rōmae] multa simul*
 meanwhile Rome.F.LOC.SG many.N.NOM.PL simultaneously

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

mōlīrī

strive.PRS.INF.DEP

‘Meanwhile, in Rome, many things did he simultaneously set in motion’

(Latin. Sallust, *Catiline’s War* 27)

- b. [*toû* =*d’ epigignoménou* *kheimōnos*]
the.M.GEN.SG and come-after.PRSP.MED.M.GEN.SG winter.M.GEN.SG
[*hē* *nósos*] *tò* *deúteron*
the.F.NOM.SG plague.F.NOM.SG the.N.ACC.SG second.N.ACC.SG
epépe *toîs* *Athēnaíois*
fall-on.AOR.IND.ACT.3SG the.M.DAT.PL Athenian.M.DAT.PL
‘And the following winter the plague fell on the Athenians for a second time’
(Ancient Greek. Thucydides, *The Peloponnesian War* 3.87, from Matić (2003, 18))

The initial adverb *intereā* in (24a) serves a topical function, situating its clause in a known time frame, yet the locative *Rōmae* also works as a contrastive topic, since it contrasts the location of the events with other regions (Etruria, Picenum, Apulia) mentioned previously. Likewise, the initial phrase of (24b) provides a time, but *hē nósos* can also be understood to be an aboutness topic.

Multiple topics in Sanskrit seem to be comparatively rarer than in its two sister languages. However, for the very common correlative construction, we may wish to locate both the initial relative *yá*-clause and its anaphoric correlate (*tásya* and *tám idám* in (25)) in the high TopicP of the same main clause.

- (25) a. [*yám* *Agním* *Médhiyātithiḥ* *Kāṇva*
which.M.ACC.SG Agni.M.ACC.SG Medhyātithi.M.NOM.SG Kaṇva.M.NOM.SG

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

lexical diagnostics to claim whether initial *intereā* in (24a) is syntactically integrated with or separate from what follows. This is all to say that there is little in the way of evidence to guide us either towards deriving multiple initial topics from the clausal syntax in (23), or only one. It may be that both analyses are on the right lines, and that we should posit both a recursive TopicP and syntactically extra-clausal topics.

The different analyses do make different predictions, however. If TopicP can be recursive, we should expect to see free ordering in the sequence, since the multiple specifiers impose no restrictions of what they can host. Alternatively, if only the last of the sequence is generated from an intra-clausal TopicP, here we should expect to find clearly integrated topics, such as oblique nouns. The impression of all three languages is that the latter case is true, and that there are some ordering tendencies within initial sequences of topics; topical arguments (subjects, direct and indirect objects) that serve as aboutness or contrastive topics tend to come later in the sequence, following themes and scene-setting material (cf. for Ancient Greek, Allan, 2012, 9). This suggests that deriving multiple topics from only one intra-clausal TopicP and otherwise from extra-clausal parataxis is the stronger analysis, but it seems too strong without further dedicated research to claim that all topic sequences must be thus derived. The conclusion of this section therefore is that initial topics in Latin, Ancient Greek and Sanskrit should be accounted for with both extra- and intra-clausal processes, and with *at least one* TopicP in the underlying syntax of the clause.

As a final point, it should be noted that the end of the Latin, Ancient Greek and Sanskrit clause also displays distinct word-order patterns that may be part of the discourse configuration. This of course does not concern the clausal left periphery (quite the oppo-

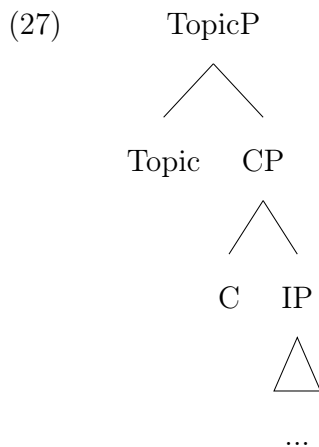
3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

These nominative nouns act as the ‘tail’ of their clause. In some cases, tails like those in (26) could be the result of the leftward movement of other constituents, stranding the subject in its basic position. In others, tails may be evidence of rightward movement, shifting the constituent beyond the end of the clause. While, agreeing with Pinkster (1991, 2021), identifying a clear syntactic end to the clause is difficult, the limited evidence for its prosodic end suggests that the tail is paratactic and stands outside the clause proper. *Sárasvatī* in (26c) follows the finite verb *dadhe*, whose lack of accent indicates the coda of the falling pitch of the clause as a whole (Hock, 1991, 189). However we derive them, the observation remains that tails may have had a certain salience, but they are not part of the same discourse-informational system as topics and foci. For Devine and Stephens (2019, 173), tails are “postponed anchored parentheticals ... afterthoughts”. Adding such clarificational material is often felt to be necessary by the producer of the clause; all three languages allow null subjects and objects, so a tail is desirable “when the speaker realizes that a zero or pronominal topic might be insufficient for the hearer to identify the topic referent” (Allan, 2014, 4). In examples like (26b) and (26c), the subject is not null, but its exact reference still may be unknown or unclear. Tails therefore serve primarily to clarify the content of their clause, rather than to indicate its topic, and so are not relevant to the TopicP in (23).

Now the task is to see how the structure in (23), proposed as the common syntax of three early Indo-European languages, fares with the rest of the language set.

3.2.2 TOPICS AND EARLY GERMANIC WORD ORDER

Early Germanic languages have much evidence to contribute in support of the syntactic structure in (23). Their clausal word order is in many respects less flexible than that of Latin, Ancient Greek and Sanskrit; this lesser flexibility is the product of word-order rules that are innovations within Germanic, and that restrict the position of the finite verb. However, we can derive the key patterns of the clause in early West and North Germanic through reconstructing the same syntax in (23) for their common ancestor, Proto-Northwest-Germanic.



The presence of C and Topic in the inherited syntax are the origins of the clause-early position of the finite verb and its word-order alternation with topical material in early Germanic.

Both Old English and Old Norse exhibit the well-known phenomenon of Verb Second

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

(V2), “the most celebrated feature of Germanic syntax” (Kiparsky, 1995, 22). It takes its name from the consistent second position of the finite verb in certain types of clause. The verb is preceded by a single syntactic constituent of unspecified category, marked by square brackets in the Old English and Old Norse examples in (28).

- (28) a. *[on þone seofþana dæg] bið*
on the.M.ACC.SG seventh.WK.M.ACC.SG day.M.ACC.SG be.PRS.IND.3SG
Sancte Siluestres tād ... [sē]
Saint.M.GEN.SG Sylvester.M.GEN.SG time.F.NOM.SG that.M.NOM.SG
dide fela wundra ār hē papa
do.PRS.IND.3SG many wonder.N.GEN.PL before he.NOM.SG pope.M.NOM.SG
wære. [Tarquinius] hātte
be.PST.IND.3SG Tarquinius.M.NOM.SG be-called.PST.IND.3SG
Rōmeburge ġerēua. [hē] brēatode
Rome.F.GEN.SG reeve.M.NOM.SG he.NOM.SG threaten.PST.IND.3SG
hine tō hāþenscipe [þā] nolde
he.ACC.SG to heathenness.M.DAT.SG then not-want.PST.IND.3SG
hē þæt. [þā] hēt hine man
he.NOM.SG that.N.ACC.SG then order.PST.IND.3SG he.ACC.SG one
ġebindan and lādan tō carcerne. [þā] cwæt
bind.INF and lead.INF to prison.N.DAT.SG then say.PST.IND.3SG
Sanctus Silluester ...
Saint.M.NOM.SG Sylvester.M.NOM.SG
‘On the seventh day is the feast of Saint Sylvester ... he performed many
wonders before he was pope. Tarquinius the prefect of Rome was called. He
threatened him into paganism. But Sylvester did not want this. Then he
ordered that he be bound and led to prison. Then said Saint Sylvester...’
(Old English. *Martyrology* 7)
- b. *anno dccc xciii [hēr] [on ðissum ġēare] wæs*
here in this.N.DAT.SG year.N.DAT.SG be.PST.IND.3SG

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

Bæbbanburh *tōbrocon* *and mycel*
 Bamburgh.F.NOM.SG destroy.PSTP.F.NOM.SG and great.ST.F.ACC.SG
herehuðe *þær genumen* *and [æfter þām]*
 plunder.F.ACC.SG there take.PSTP.F.NOM.SG and after that.N.DAT.SG
cōm *tō Humbran* *mūðe* *se*
 come.PST.IND.3SG to Humber.F.GEN.SG mouth.M.DAT.SG the.M.NOM.SG
here ... [*on þysum* *ilcan* *gēare*]
 army.M.NOM.SG in this.N.DAT.SG same.N.DAT.SG year.N.DAT.SG
hēt *se* *cyng* *āblendan*
 order.PST.IND.3SG the.M.NOM.SG king.M.NOM.SG blind.INF
Ælfgār
 Ælfgār.M.ACC.SG

‘In the year 993: here in this year was Bamburgh destroyed and much plunder there taken, and after that came the army to the mouth of the Humber ... in the same year the king ordered that Ælgar be blinded’

(Old English. *Anglo-Saxon Chronicle* E 993)

- c. *en [Suttungamjōð]* ***gaf*** *Óðinn*
 but Suttungr-mead.M.ACC.SG give.PST.IND.3SG Odin.M.NOM.SG
ásunum *ok þeim* *mōnnum,* *er yrkja*
 Æsir.M.DAT.PL and that.M.DAT.PL man.M.DAT.PL REL compose.INF
kunnu. [*því*] ***kollum*** *vér*
 can.PRS.IND.3PL therefore call.PRS.IND.1PL we.NOM.PL
skáldskapinn *feng* *Óðins*
 poetry.DEF.M.ACC.SG loot.M.ACC.SG Odin.M.GEN.SG

‘But Suttungr’s mead Odin gave to the Æsir and to those men who can compose poetry. Thus we call poetry *Odin’s loot*’

(Old Norse. *Prose Edda* 3.6)

- d. [*síðan*] ***gengu*** *þau* *til svefns* ... *en [um*
 afterwards go.PST.IND.3PL they.NOM.PL to sleep.M.GEN.SG and in
morguninn] ***fóru*** *þau* *til drykkju,*
 morning.DEF.M.ACC.SG go.PST.IND.3PL they.NOM.PL to drinking.F.GEN.SG

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

ok [*allan* *hálfan* *mánuð*] *lágu*
 and all.M.ACC.SG half.M.ACC.SG month.M.ACC.SG lie.PST.IND.3PL
þau *þar* *tvau* *ein* *í loptinu*
 they.NOM.PL there two.N.NOM.PL alone.N.NOM.PL in attic.DEF.N.DAT.SG
 ‘Afterwards they went to sleep ... and in the morning they set to drinking
 and the whole half month they lay there, two alone in the attic’
 (Old Norse. *Njál’s Saga* 3)

In contrast with Modern English, the Old English declarative main clauses in (28a) and (28b) display the conjugated finite verb (e.g. *bið*, *wæs*) as their second constituent, excluding coordinating conjunctions. It follows a variety of initial constituents of varying length; the position of the verb respects syntactic constituency.

Following den Besten (1977) for modern-day German and van Kemenade (1987) for Old English, we can derive the V2 behaviour through the movement of the finite verb from the VP and up to the C head in (27). Specifically, this is only possible when the C head is null. In Old English subordinate clauses with a lexically overt C-element, the movement is blocked and the verb appears later or last in the clause; Old English V-to-C movement can therefore be derived with a C_{Main} head that attracts the verb. Meanwhile, the initial constituent of V2 could occupy Spec,CP in both Old English and Old Norse, a standard analysis in present-day Germanic languages.

However, while modern-day V2 can be derived through a syntactic constraint to ‘fill’ Spec,CP, early Germanic languages display greater variation in the position of the finite verb, such as Verb First (V1) and Verb Third (V3). To explain the former at least,

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

any obligatory ‘Fill Spec,CP’ rule cannot hold.¹¹ This thesis advances instead that the early Germanic variation is better understood through topicality, conditioned specifically through non-obligatory movement up to a TopicP inherited from the proto-syntax.

Verb Second, while a handy label for the Germanic phenomenon, is a deceptive term. Early Germanic languages do not display V2 order uniformly; it differs as to its obligatoriness and the clauses in which it appears. It is certainly extremely common in Old English main clauses; Walkden (2014, 68) states that V2 is “the majority pattern in main clauses in general”, finding that the finite verb is placed second in 65.4% of declarative main clauses in *Ælfric’s Lives of the Saints* and in 41.9% in the Old English translation of Bede’s *Historia*. It is obligatory when the adverbial elements *þonne* ‘then’ and *þā* ‘then’ are initial. It is also the norm in *wh*-questions, in which the *wh*-material immediately precedes the verb.

However, in declarative main clauses, a frequent alternative in Old English is Verb Third. This order is especially common with personal pronouns, and the second constituent, immediately before the verb, is always a familiar topic (Bech, 1998, 2001).

- (29) a. *ond* [*sēo* *clāsnung*] [*him*] *wæs* *swā*
 and the.F.NOM.SG cleansing.F.NOM.SG they.DAT.PL be.PST.IND.3SG so
hālīg *swā ūs* *is* *fullwiht*
 holy.ST.NOM.SG so we.DAT.PL be.PST.IND.3SG baptism.N.NOM.SG
 ‘And this purification was to them as holy as baptism is to us’

(Old English. *Martyrology* 9)

- b. [*þær*] [*hire*] *brohte* *Godes* *engel*
 there she.DAT.SG bring.PST.IND.3SG God.M.GEN.SG angel.M.NOM.SG

¹¹It certainly cannot hold in subordinate clauses with subordinators; the subordinator invariably stands initially, while the underlying CP still has an empty specifier position.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

swylcne *gerelan*
 such.M.ACC.SG robe.M.ACC.SG
 ‘There God’s angel brought her such a robe’

(Old English. *Martyrology* 9)

- c. [*iċ* *Apollonius* *se* *Tirisca*
 I.NOM.SG Apollonius.M.NOM.SG the.M.NOM.SG Tyrian.ST.M.NOM.SG
ealdorman] [*ēow*] *cȳðe* *þæt iċ*
 nobleman.M.NOM.SG you.DAT.PL proclaim.PRS.IND.1SG that I.NOM.SG
gelife
 believe.PRS.IND.1SG
 ‘I, Apollonius the Tyrian prince, declare to you that I believe...’

(Old English. *Apollonius of Tyre*, from Bech (2001, 4))

Pronominal subjects and objects frequently precede the finite verb, and may therefore produce a V3 order. It would be reasonable on this basis to propose that this pre-verbal position is a product of some feature specific to pronouns; we might claim that it is the result of clisis, and therefore that V3 can be reduced at the syntactic level to V2. Yet the wider evidence does not support this; there is little independent support for the proclitic status of personal pronouns, and Bech (2001) finds that such a ‘clitic analysis’ resolves many but not all of the exceptions to the V2 rule. It leaves the rare full nouns and noun phrases in a V3 order unaccounted for.

- (30) [*þære* *tīde*] [*Dinoð*] *wæs*
 the.F.DAT.SG time.F.DAT.SG Dinoth.M.NOM.SG be.PST.IND.3SG
hāten *þæs* *mynstres* *abbod*
 call.PSTP.M.NOM.SG the.N.GEN.SG monastery.N.GEN.SG abbot.M.NOM.SG
 ‘At that time, the abbot of the monastery was called Dinoth’

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

(Old English. Bede's *Historia* 2, from Bech (2001, 64))

Even if Old English did have clitic pronouns, it cannot be considered a consistent V2 language (Bech, 2001, 84). Both V2 and V3 therefore appear to be productive orders in Old English declarative main clauses. V2 is “relatively common, but by no means obligatory” in Old English (Swan, 1994, 68), while being considerably more robust in Old Norse; Rögnvaldsson (1995, 14) identifies a V3 order in one single clause in his corpus of Old Norse texts. Walkden (2014, 89-92) reconstructs V3 back to Proto-Northwest-Germanic, although recognises that this is only for the sake of diachronic parsimony (Walkden, 2017, 73-4); it is more plausible that Old English (and apparently Old High German writers too) retained V3, rather than independently innovating it.

A further word-order option that is available to both Old English and Old Norse is Verb First. This order is found in imperative clauses (31) and polar questions (32) across all early Germanic languages.

- (31) a. *nim* *buteran*, *wyl* *on pannan.* *āflēot*
 take.IMP.2SG butter.F.ACC.SG boil.IMP.2SG in pan.F.DAT.SG skim.IMP.2SG
þæt *fām* *of and āhlýttre* *þā*
 the.ACC.SG foam.N.ACC.SG off and clarify.IMP.2SG the.F.ACC.SG
buteran *on blede.* *dō* *eft þæt*
 butter.F.ACC.SG in bowl.F.DAT.SG put.IMP.2SG after the.N.ACC.SG
hlýttre *on pannan.* *ģecnua* *celeponian*,
 clarified.N.ACC.SG in pan.F.DAT.SG pound.IMP.2SG celandine.F.ACC.SG
bisċeopwyr̄t, *wudumerċe.* *wyl* *swīþe.*
 marsh-mallow.F.ACC.SG wild-celery.M.ACC.SG boil.IMP.2SG very
āseoh *þurh clāð.* *nytta* *swā þe*
 strain.IMP.2SG through cloth.M.ACC.SG use.IMP.2SG so that

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

þearf *sīe*.
 need.F.NOM.SG be.PRS.SBJ.3SG
 ‘Take butter, and boil in a pan. Skim the foam off and clarify the butter in a bowl. Put the clarified butter afterwards in a pan. Pound celandine, marsh mallow and wild celery. Boil thoroughly. Strain through a cloth. Use as need be’

(Old English. *Bald’s Leechbook* 2)

- b. *fylg* *þú* *þeim* *til húsa* *minna*
 follow.IMP.2SG you.NOM.SG they.DAT.PL to house.N.GEN.PL my.N.GEN.PL
ok ger *þeim* *þar góða* *veizlu* ...
 and do.IMP.2SG they.DAT.PL there good.ST.F.ACC.SG feast.F.ACC.SG
sit *þú*
 sit.IMP.2SG you.NOM.SG
 ‘Follow them to my hall and make them a great feast there ... sit yourself!’

(Old Norse. *Njál’s Saga* 3)

- (32) a. *wāst* *þū* *hwæt* *mon*
 know.PRS.IND.2SG you.NOM.SG what.N.NOM.SG man.M.NOM.SG
sīe? ... *wēnst* *þū* *hwæðer āniġ*
 be.PRS.SBJ.3SG think.PRS.IND.2SG you.NOM.SG whether any.NOM.SG
ġesceaft *sīe...?*
 creature.F.NOM.SG be.PRS.SBJ.3SG
 ‘Do you know what a man is? ... Do you suppose there to be any creature...?’

(Old English. OE *Boethius* 34)

- b. *ert* *þú* *hugsjúkr,* *Hrútr?* ...
 be.PRS.IND.2SG you.NOM.SG mindsick.ST.M.NOM.SG Hrut.M.VOC.SG
villt *þú* *til Íslands?* ... *mun*
 want.PRS.IND.2SG you.NOM.SG to Island.N.GEN.SG will.PRS.IND.3SG
þinn *sómi* *þar meiri* *en hér?*
 your.NOM.SG honour.M.NOM.SG there greater.WK.M.NOM.SG than here

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

“Are you unhappy, Hrut? ... Do you wish to go to Iceland? ... Will your honour be greater there than here?”

(Old Norse. *Njál’s Saga* 6)

While it has been seen in Chapter 2 that imperative-verb-initial clauses are common across early Indo-European, this clause-early early position of the verb in both polar and *wh*-questions appears to be a distinctly Germanic phenomenon.¹²

Most notably, V1 orders constitute a significant statistical minority in declarative main clauses too. Walkden (2014, 92) reports for Old English that 11.1% of relevant clauses in Ælfric’s *Lives of Saints* and 28.0% in Bede’s *Historia* contain an initial verb.

- (33) a. *cōmon* *hī* *of þrīm* *folcum,*
 come.PST.IND.3PL they.NOM.PL from three.N.DAT.PL people.N.DAT.PL
þām *strangestan* *Germanie*
 the.N.DAT.SG strongestWK.N.DAT.SG Germany.F.GEN.SG
 ‘They came from three peoples, the strongest of Germany’
 (Old English. Bede’s *Historia* 1)
- b. *wæs* *mīnra* *gefērana* *mid mē*
 be.PST.IND.3SG my.M.GEN.PL companion.M.GEN.PL with I.DAT.SG
þrīo *hund* *monna.* *þā hēt*
 three.N.NOM.PL hundred.N.NOM.PL man.M.GEN.PL then order.PST.IND.3SG
se *bisceop* *mīne* *gefēran þæt ...*
 the.M.NOM.SG bishop.M.NOM.SG myCompanion.M.ACC.PL that and
ond *hēt* *ic* *ēghwæt swā* *dōn swā*
 order.PST.IND.1SG I.NOM.SG whatever so do.INF so he.NOM.SG

¹²The global survey of modern-day polar questions on WALS shows that the use of word order to mark their interrogativity is nearly exclusive to Germanic (Dryer, 2013b), and languages in prolonged contact with Germanic, such as Czech.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

hē ās bebead. wæs hit þā
 we.ACC.PL ask.PST.IND.3SG be.PST.IND.3SG it.NOM.SG then the.F.NOM.SG
sīo endlefte tīd dægēs
 eleventh.WK.F.NOM.SG time.F.NOM.SG day.M.GEN.SG
 ‘There were of my companions with me three hundred men. Then the bishop
 ordered my companions that ... And I ordered them to do whatever he asked
 of us. It was then the eleventh hour of the day’
 (Old English. *Letter of Alexander to Aristotle*)

- c. *fór Bólverkr þar til sem Gunnlöð*
 go.PST.IND.3SG Bölverk.M.NOM.SG there to as Gunnlöð.M.NOM.SG
var
 be.PST.IND.3SG
 ‘Bölverk went there to where Gunnlöð was’

(Old Norse. *Prose Edda* 3.6)

- d. *mátti þá enn sem optarr meira hamingja*
 can.PST.IND.3SG then yet as oftener greater.F.NOM.SG luck.F.NOM.SG
konungs en fjólkyngi Finna. fengu
 king.M.GEN.SG than witchcraft.F.NOM.SG Finn.M.GEN.PL get.PST.IND.3PL
þeir beitt um nóttina fyrir
 they.NOM.PL steer.SUP during night.DEF.F.ACC.SG before
Bálagarðs-síðu

Balegard-coast.F.ACC.SG

‘The king’s luck was then yet again, as is more usual, greater than the
 witchcraft of the Finns. They sailed throughout the night along the coast
 of Balegard’

(Old Norse. *Heimskringla* 7.9)

Granted, there may be various motivations for this order, including non-native influences¹³,

¹³The initial position of *cōmon* in (33a) does match that of the corresponding verb *advēnerant* in the

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

but V1 is too widespread and apparently productive to be excluded from native Germanic syntax. Walkden (2014, 94) concludes that, given both the wide distribution and the antiquity of its appearances, a V1-producing syntax can even be reconstructed for the common ancestor of North and West Germanic. The reasons for the selection of a V1 clause are discussed shortly.

Furthermore, we should acknowledge that Old English attests clauses in which the finite verb stands even later (V4+). This order is especially common in subordinate clauses, as noted in Chapter 2, in which the presence of a subordinator blocks an early position. Yet there are main clauses too that display V4+, though they are uncommon; Cichosz (2010) finds that 15.6% of her sample of 122 Old English prose clauses have a V4+ order.

- (34) *him* *þær* *se* *gionga* *cyning*
he.DAT.SG there the.M.NOM.SG young.WK.M.NOM.SG king.M.NOM.SG
þæs *oferfærelde* *forwiernan* *mehte*
the.N.GEN.SG passage.N.GEN.SG prevent.INF can.PST.IND.3SG
‘There the young king could have prevented his crossing’

(Old English. *Orosius* 44, from Pintzuk (1993, 18))

In (34), the finite verb comes last, being preceded by five constituents. It is difficult therefore to claim that such an order fundamentally includes the same clause-early position of the finite verb that produces V1/V2/V3. V4+ is better understood as an older (but still productive) grammar. There is also a diachronic trend in the order’s distribution; head-final orders at least appear to decrease over the course of the Old English period (Pintzuk

Latin original.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

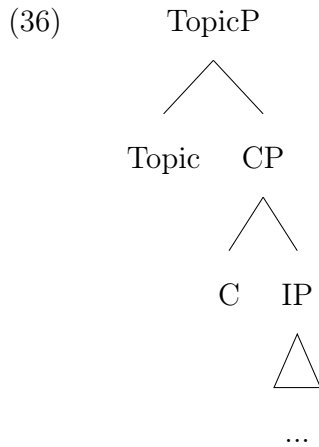
and Haeberli, 2008, 28). Furthermore, Þorgeirsson (2012) reports that V4+ occurs in the *fornýðislag* genre of Old Norse poetry, but not in other genres or in prose, and argues that this is an archaism inherited from early stages of North Germanic and preserved in the conservative grammar of poetry. As evidence of the antiquity of verb-finality, he cites the inscription on one of the two Golden Horns of Gallehus, dating to the fifth century AD and written in Proto-Norse.

- (35) *ek Hlewagastiz Holtijaz horna*
I.NOM.SG Hlewagastiz.M.NOM.SG Holtijaz.M.NOM.SG horn.N.ACC.SG
tawido
make.PST.IND.1SG
'I Hlewagastiz Holtijaz made the horn'

(Proto-Norse. Runic inscription, from Þorgeirsson (2012, 29))

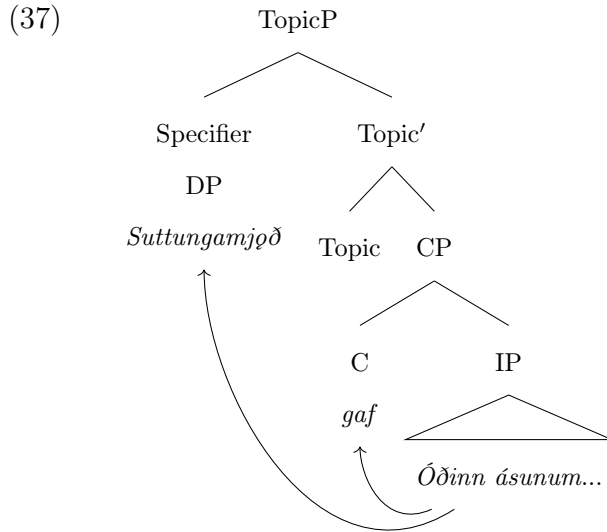
Other scholars agree that verb-final orders in Old Norse are relics of an older preference for a default Subject-Object-Verb order attested in many runic inscriptions (Faarlund, 1990, 51-2).

Despite all the word-order variation between and within early Germanic languages, this behaviour of the left periphery of declarative main clauses can be successfully derived through the common syntax for early Germanic in (27), repeated in (36) below.

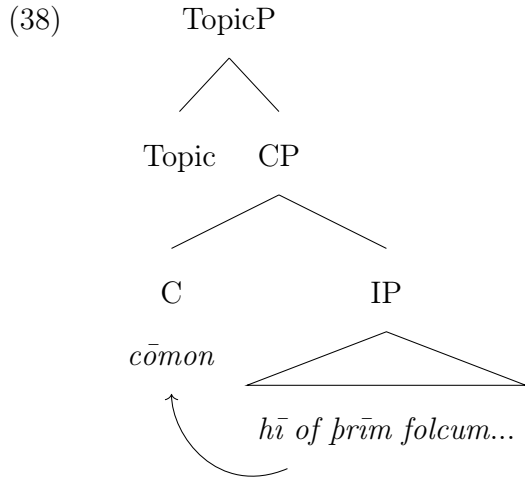


Firstly, as noted above, we can straightforwardly derive the V2 order in Old English and Old Norse declarative main clauses through the V-to-C analysis of den Besten (1977). The verb thus moves past the subject, object and any adjuncts, which will only precede the verb in the surface order if one of them is in turn moved even higher. The difference between Modern German and early Germanic is the obligatoriness of the secondary movement operation. In Modern German, it is obligatory, and thus only a V2 order is possible in declarative main clauses. Yet in early Germanic, we should instead posit that it is not obligatory, and is still conditioned by discourse information. If it happens, the syntax produces V2 (37).

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE



If not, there is only V-to-C movement, and the syntax produces V1 (38)



This naturally prompts the question of why this V2/V1 variation occurs; adopting the

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

TopicP in (36) provides a ready answer. The preverbal material in early Germanic may be categorially varied, but it is unified by discourse-informational status; a common thread across the previous literature is that the material is consistently defined by given-ness and topicality. Optional topicalisation and movement to TopicP therefore produce the systematic variation between V1 and V2. V1 clauses lack an initial topic that connect them to an established referent; they therefore serve to “introduce a new discourse referent”. They typically occur at the start of texts or of narrative episodes within a text (Hinterhölzl and Petrova, 2009, 2). This fits with the examples in (33), in which the V1 clauses do not contribute information about an existing topic. They instead recount a new event with a new subject, which may be the clausal focus.

V2 clauses meanwhile do have an initial topic in Spec,TopicP, and thus they provide additional information about an “already established discourse referent” (Hinterhölzl and Petrova, 2009, 2). They are therefore in a sense subordinate to V1 clauses within this framework; in principle, each episode within a narrative begins with a V1 clause, which may have one, many or no dependent V2 clauses to provide additional information, before the narrative then moves on to a new episode, a new topic and a new V1 clause (Hinterhölzl and Petrova, 2009, 4). Cichosz (2017) demonstrates that V1 in the Old English *Historia* of Bede is mostly used to the same effect, namely to “mark transition in the narrative structure of the text” (Cichosz, 2017, 2).

As Hinterhölzl and van Kemenade (2012, 16) note for Old English and Faarlund (2004, 100) for Old Norse, V-to-C has the overall effect of dividing the early Germanic clause into two zones: the material that precedes the finite verb and the rest of the clause. The

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

first zone (the ‘*Vorfeld*’) contains topical material, including anaphora and scene-setting adverbs. The remainder of the clause then provides information about that topic, if one is present. Early Germanic languages (excluding Gothic) therefore use the verb to express in word order the common dichotomy of topic and comment. Through V-to-C movement, the syntax produces a clause-early verb that serves to signal the end of the first zone (the topic) and the beginning of the second (the comment).

“This pattern in OE serves to separate ALL discourse-anaphoric elements from the focus domain of the clause, rather than only the aboutness topic; the pre-verbal domain in OE and ME thus consists of more than one anaphoric linking position: one for the clause-initial constituent, and one for the subject.”

(Hinterhölzl and van Kemenade, 2012, 16)

In TopicP we can also locate the common Old English elements *þonne* ‘then’ and *þā*; among other functions, they are used for “discourse-sequencing”, linking their clause to its wider context (van Kemenade, 1987; van Kemenade et al., 2023). In sum, we can derive and motivate both V1 and V2 through the syntax in (36), the same proposed for the three ancient languages of the set. Both orders involve the movement of the finite verb up to C. What differs is whether a topic is moved up from among the other constituents or base-generated in TopicP too.

However, the matter of Old English V3 remains. Its availability in different Germanic languages is subject to restrictions on the quantity and kind of clause-initial topics. In Old Norse, only one topic is permitted, ruling out V3. In Old English on the other hand, two

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

(or more) topics are possible, which may be of various syntactic categories. Third-person pronouns are by their nature often topical, referring back to a noun or noun phrase already given. Hence, it is not surprising that they frequently occur within the pre-verbal topic ‘zone’ in Old English, as in examples (29a) and (29b). Indeed, Bech (1998) argues that pre-verbal subjects in V3 typically have low ‘communicative dynamism’, meaning that they are given or context-dependent entities. She concludes that the Old English V3 order with subjects in second position is “governed by pragmatic factors in early OE” (1998, 20). Hinterhölzl and Petrova (2009) additionally note the kind of topics that can appear pre-verbally; while Old High German and Old Norse robustly only allow aboutness topics in a V2 order, Old English permits both the single aboutness topic of the clause and “other background elements” to be pre-verbal (2009, 5). These may be scene-setting adverbial phrases.

- (39) *millesimo ii* [*hēr*] [*on þissum gēare*] [*se*
 here in this.N.DAT.SG year.N.DAT.SG the.M.NOM.SG
cyng] *gerēdde* and his witan þæt ...
 king.M.NOM.SG arrange.PST.IND.3SG and he.GEN.SG witan.M.NOM.PL that
 [*on ðām gēare*] [*se cyng*]
 in that.N.DAT.SG year.N.DAT.SG the.M.NOM.SG king.M.NOM.SG
hēt *ofslēan ealle* *ðā* *Deniscan*
 order.PST.IND.3SG kill.INF all.M.ACC.PL the.M.ACC.PL Danish.WK.M.ACC.PL
men
 man.M.ACC.PL
 ‘1002: here in this year the king decreed, with his witan too, that ... in that year,
 the king ordered that all the Danish men be killed’
 (Old English. *Anglo-Saxon Chronicle* E 1002)

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

Here in (39), the second main clause has a V3 order (and the first also, depending on how we analyse *hēr on þissum gēare*). What precedes the verb includes a scene-setting adverbial, which situates the event that follows in the given year, followed by a known and familiar entity, the king. Old English V3 therefore, like V2 and V1, is a product of the division of the clause and its constituents into topic and comment.

To derive V3 syntactically, scholars have previously proposed articulated structures with different positions for the verb and its subjects (cf. Haeberli, 2002). van Kemenade et al. (2023), following Walkden (2017) for present-day V3, adopt a Rizgian syntax, in which the finite verb moves to either high Force (hence V2) or lower Fin (hence V3). However, this analysis dissolves V-to-C movement as a unified operation. Moreover, the particular syntactic heads involved do not account for the primarily topical and discourse-given characteristic of the preverbal *Vorfeld*, as mentioned above. It seems better therefore to keep V-to-C the same in V1, V2 and V3 orders, and to account for V3 instead through the same TopicP behind V2. The presence of two topical constituents in its *Vorfeld* is reminiscent of the issue of initial multi-topic sequences in §3.2.1. The same two options present themselves: to analyse the first constituent of V3 as extra-clausal, or to posit that TopicP in Old English is capable of recursion. The same problem also appears: diagnosing extra-clausal status in sequences of topics.

If, following previous literature, we want to keep the first two constituents of V3 within the underlying syntax, a recursive TopicP (‘TopicP*’) is one way to proceed. Perhaps this recursion was an option for early Indo-European more generally, not only for ancient languages like Latin, but also retained by Old English. The Old Norse V1/V2 variation

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

meanwhile arises from an innovative syntax without the possibility of recursive TopicP. Old Norse therefore represents a later step in the development of Germanic V2; this long-term change is “the syntacticization of an originally information-structurally conditioned pattern” (Walkden, 2014, 91). Over time, the flexibility and optionality of the pre-verbal topic zone was lost, and the common V2 order, which was at one point used only for the topic-comment structure that it produced, became syntactically obligatory.

Given its explanatory successes across early Germanic, we may reconstruct the syntax in (36) for the common ancestor of the languages in question, namely Proto-Northwest-Germanic.¹⁴ At this juncture, we can consider TopicP and clause-initial topicalisation in Proto-Northwest-Germanic to be inheritances, cognate with the equivalent features of Latin, Ancient Greek and Sanskrit. Under this view, initial topics of attested early Germanic are conservative; what is new is the V-to-C that accompanies them.

This verbal movement may have started in declarative main clauses, but spread to imperative and interrogative main clauses. V-to-C may have been a feature of Proto-Northwest-Germanic already, or diffused across the subsequent dialectal continuum. From this generalisation of V-to-C across clause types comes the aforementioned unusual verb position in Germanic questions, which presumably has since endured as a useful indicator of interrogativity. We can motivate the rise of Germanic V-to-C through assuming the pre-existence of C as a functional category and feature of the clausal syntax.

As mentioned, we have possible remnants of Germanic grammar prior to generalised

¹⁴To reconstruct it for Proto-Germanic itself, evidence from Gothic is necessary, yet, as mentioned in Chapter 1, Gothic poses philological challenges.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

V-to-C. These include Old English subordinate clauses and V4+ main clauses, and Old Norse *fornýðislag* poetry. Yet clauses without V-to-C are not lifeless fossils; the verb-late order must have had its own interpretative qualities, contrasting with clauses with V-to-C. Walkden (2014, 97-106) proposes that movement or the absence thereof could create an ‘asserted vs. presupposed’ distinction in early Germanic. This relates to the common ground of a particular discourse; a proposition is presupposed “if the speaker believes that its truth belongs to the common ground”, and asserted if it adds new information (Walkden, 2014, 100). Walkden suggests that the latter interpretation could be expressed syntactically by V-to-C movement, and the former by its absence. The early position of the verb may have started out as a marked order, but lost that status over time, transforming into the default division between topic and comment. The view of C in Chapter 2 defines it as the locus of clause type and status, and speaker attitudes towards the truth of the clause. If we assume this view of C, we can explain the origins of V-to-C as a means of signalling the declarative type and assertive quality of the overall clause, communicating to the audience that this clause is new information that its producer believes to be true. This V-to-C movement would naturally not occur in subordinate clauses, in which a lexical subordinator already served to express their type and status.

An unresolved problem for this account is that, if we assume (36) for all early Germanic clause types, it predicts that we should find at least some topic-initial questions and imperative clauses. In the latter, the imperative verb can be understood to move to C as usual. This usually leaves TopicP empty, but examples exist in which topic-like material precedes the verb.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

- (40) a. *ac* [*tō Fasiacen* *and Porre* *þāem* *cyninge*]
 but to Fasiace.F.DAT.SG and Porus.M.DAT.SG the.M.DAT.SG king.M.DAT.SG
eft *gehworf* *þū*
 back turn.IMP.2SG you.NOM.SG
 ‘But return yourself to Fasiacen and to Porus the king’

(Old English. *Letter of Alexander to Aristotle*)

- b. *en* [*ef honum* *þykkja* *á því* *úhægendi*]
 but if he.DAT.SG seem.PRS.IND.3PL on that.DAT.SG discomfort.N.NOM.SG
 [*þá*] *farið* *þér* *hingat*
 then go.IMP.2PL you.NOM.PL thither
 ‘But if he finds them a problem, then go there yourself’

(Old Norse. *Droplaugarsona Saga* 14, from Faarlund (2004, 229))

In (40a), one constituent precedes the imperative verb (two if we do not consider *eft* part of the verb), while *farið* in (40b) follows a whole protasis clause and *þá* in a rare instance of Old Norse V3. These constituents we may locate in TopicP as either known material or a background condition. Likewise, in polar and *wh*-questions, the movement of the verb and *wh*-material to CP leaves TopicP free to host topics. In subordinate clauses too, the markers of subordination likewise stand within the CP. Adopting the syntax in (36) for the latter types of clauses predicts that we should find Old English and Old Norse examples of topic-initial questions and finite subordinate clauses. However, at present, none have been found. The search has not been exhaustive, but an absence of such orders may suggest that these types of clauses have a simpler syntax than that in (36), lacking TopicP. This relates to the speculation in §3.2.1 that subordinate clauses may have lost TopicP in tandem with the development of finite-clausal embedding.

3.2. *TOPIC IN EARLY INDO-EUROPEAN AND PIE*

To conclude, the two early Germanic languages in the language set, Old English and Old Norse, work well with the abstract syntax of the clausal left periphery in (36) proposed for their three Indo-European cousins. Its two components of TopicP and CP can account for the typically early position of the verb and for the topical material that variably precedes it in main clauses. On the basis of shared word-order patterns, we may conceivably reconstruct it (for main clauses at least) as far back as Proto-Northwest-Germanic. In the loss of V1 orders and the evidence of an older grammar without V-to-C, we can also see how a primarily discourse-configurational language like Latin might undergo ‘syntacticization’ and become a language like Modern German. While Latin maintains the option to host material in TopicP and CP, early Germanic did this to such a high degree that through successive generations, the word order that this movement generated lost its marked status and eventually became obligatory in most descendent languages. So far, (36) accounts well for five of the seven languages in the set. However, before we can address the reason for this commonality and talk about reconstruction, there are two more languages to consider.

3.2.3 TOPICS IN OLD CHURCH SLAVONIC AND OLD IRISH

Old Church Slavonic offers little direct evidence either to confirm or reject the syntactic structure (1) in the introduction. All non-clitic (i.e. accented) material is hard to disentangle from the general effort to mimic the word-order patterns of original Greek texts. This challenges any identification of topics in our OCS texts. However, the comparative evidence does suggest that clause-initial topics are the norm in Slavic languages in general;

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

the Slavic clause typically places old and presupposed information first, and overall has a topic-comment order (Sussex and Cubberley, 2006, 417-20). It is unlikely that OCS differed in this regard, and so any initial topics in our OCS sources may well be cases of utilising a native high TopicP to copy the Greek. Furthermore, there is the example mentioned in Chapter 2, in which the Slavic interrogative particle *li* follows a full phrase, rather than the first word.

- (41) [*ne i mytare*] = *li tožde tvoręť*
not and tax-collector.M.NOM.PL Q also do.PRS.3PL
'Do not even tax-collectors do likewise?'

(Old Church Slavonic. *C. Mar.* Mt. 5.46)

We might understand *ne i mytare* to be one prosodic unit, with *ne* and *i* as proclitic and *li* as enclitic on the noun. Alternatively, we may hesitantly treat it instead as a single syntactic constituent, located by the translator high in the syntax above the position of the C-head *li*. All of this though is speculative and offers only indirect evidence for a similar TopicP; the fact remains that OCS can tell us very little that is unambiguous.

Old Irish, however, is consistent with (1). §2.2.3 and Bate (2024) set out a new derivation of the Old Irish declarative main clause, in which the behaviour of the verb is understood to arise from the interactions between the null C-element = \emptyset and the verb. With compound verbs, this interaction involves the clisis of = \emptyset to the first preverb, producing deuterotonic verbal forms. If the CP is somehow lexically overt or = \emptyset is enclitic on the negator *ní*, compound verbs are left to be prototonic. On this basis, in the rare instances

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

that the verbal complex is not the first component of its clause, the form of the verb can tell us whether the initial material is in CP or not.

DiGirolamo (2018) notes that sometimes a single constituent does precede the verbal complex within the same clause, and that they appear not only in poetry, but also as a notable minority (7%) of applicable clauses in the prose language of the Würzburg and Milan glosses. The fronted constituent may be a prepositional phrase, a nominative noun phrase or, most notably, an accusative noun phrase. In the latter two cases, the constituent may be resumed in the remainder of the clause with a pronoun.

- (42) a. [*ind* *foisitiu* *i* *ngiun*]
 the.F.NOM.SG confession.VN.F.NOM.SG in mouth.M.DAT.SG
 im·folngi *in* *duine* *slán*
 produce.PRS.IND.ACT.3SG the.M.ACC.SG man.M.ACC.SG whole.M.ACC.SG
 ‘The confession in the mouth makes the person whole’
 (Old Irish. Wb. 4d33, from DiGirolamo (2018, 4))
- b. [*i* *n-aimsir* *Mordochei*] *ro·comallad* *techt*
 in time.F.DAT.SG Mordechai.M.GEN.SG fulfil.PRF.ACT.3SG go.VN.F.NOM.SG
 doib *huili* *fo* *recht* *n-indibi*
 to.3PL all.M.NOM.PL under law.M.ACC.SG circumcision.VN.N.GEN.SG
 ‘In the time of Mordechai it was fulfilled that all went under the law of cir-
 cuncision’
 (Old Irish. Ml. 81d5, from DiGirolamo (2018, 2))

Unlike a conjunct particle, we find that the initial phrase does not affect the form of the verb. If it stood in CP, we would expect simple verbs to bear conjunct inflection and compound verbs to be prototonic. Yet forms like *im·folngi* and *ro·comallad* are deuterotonic,

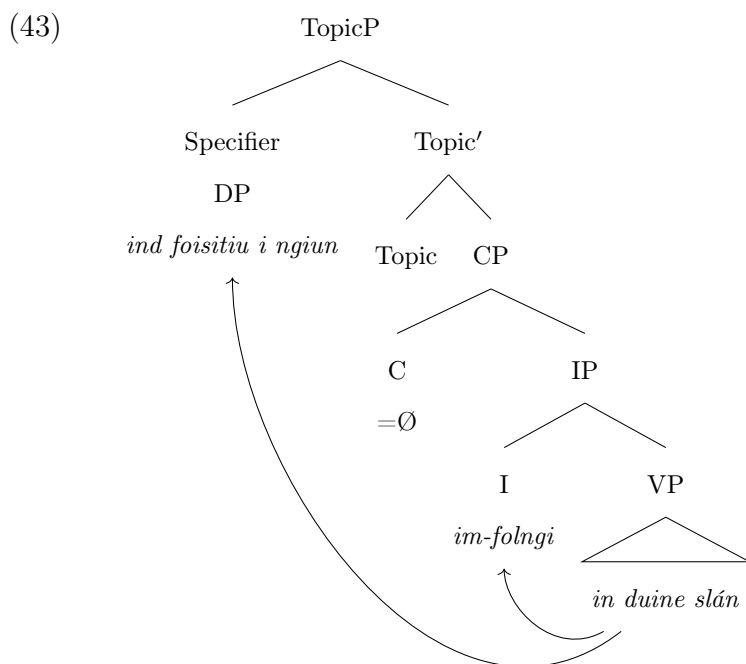
3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

showing the clisis of null = \emptyset . Since the verb is not relative, we have no grounds to claim that these are instances of clefting, containing two clauses; these topics stand within the same clause as the verb. Following DiGirolamo, and on the basis of the source context and marked word order, we can consider these clauses to include fronted topics. Specifically, clauses like the two in (42) involve the movement of the initial phrase to Spec,TopicP, higher than the verb and the CP.

DiGirolamo states that both fronted aboutness and contrastive topics were a productive construction in Old Irish, and suggests that these two types of topic are hosted in dedicated syntactic projections. DiGirolamo adopts the multi-layered left periphery of Frascarelli and Hinterhölzl (2007), yet this predicts multiple topics in our Old Irish sources. It does not appear that this prediction is borne out; while DiGirolamo does demonstrate that initial topics are of different types and functions, she does not provide evidence of multi-topic strings that would necessitate several projections in the syntax. To avoid the unnecessary multiplication of entities, one all-purpose (non-recursive) TopicP is sufficient for the synchronic syntax of Old Irish.

The analysis of Bate (2024) for a topic-initial clause like (42a) looks so:

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE



Placing topical material in the high TopicP was presumably accompanied by a distinct prosodic gap between the fronted material and the rest of the clause. This bars the topic from offering a suitable prosodic host for the enclitic =Ø. This C-element instead continues to trigger prosodic inversion with the first preverb of a verb, itself in the IP, as proposed in §2.2.3.

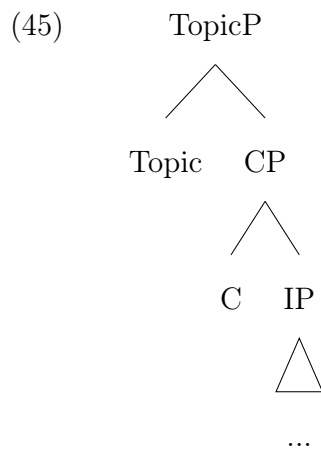
TopicP furthermore supplies a position for those elements in Old Irish mentioned in §2.2.1 that mark subordination, yet precede deuterotonic compound verbs. These include *má* ‘if’, (*h*)*óre*/ ‘because’ and *cía*/ ‘although’. As lexical items, these function as subordinators, but the form of the following verb indicates that they are not in the CP.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

TopicP incorporates those occasional phrases that stand even further in the left edge than the verbal complex.

3.2.4 RECONSTRUCTING TOPIC FOR PROTO-INDO-EUROPEAN

Out of the seven representatives for early Indo-European, six can be derived through the following syntactic structure:



The seventh language, Old Church Slavonic, is at least not inconsistent with this structure. Both the lexical evidence from the language and the comparative evidence from Slavic support the existence of C in OCS, while we have both comparative and potential indirect evidence for a discourse-functional component like TopicP. This follows from the principle of translation in §2.2.1: since the translators (and subsequent copyists) must have wanted to avoid ungrammatical gibberish, we can propose that the native Slavic syntax could produce the patterns and variation that we observe in OCS sources, even when the determining

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

factors came from Greek.

The structure in (45) is very similar to the syntax for Sanskrit of Hale (1987a). It has two key components: with regards to the overall clause, one looks inward, the other outward. The role of C is to encode the type and status of the clause. Topic meanwhile serves to situate the clause in its wider context, hosting material that refers back to established information, ‘sets the scene’, provides a requisite condition, or conveys what is the psychological subject according to the clause’s producer. Multiple instances of such material lead us to posit a combination of both intra-clausal topics in TopicP and extra-clausal topics in parataxis, and additionally to suggest the possibility that TopicP is capable of recursion, to account for multiple integrated topics.

The claim of this thesis is that we should reconstruct at least the syntax in (45) back to Proto-Indo-European, or more specifically the common ancestor of the language set. Its compatibility across this spread of Indo-European languages takes us beyond what we can ascribe to mere coincidence. Having been proposed for six out of the seven in the set, (45) satisfies the majority rule of traditional reconstruction; since it works for Vedic Sanskrit, Ancient Greek and Latin, it satisfies the elder rule too (see §1.2.1). Proposing (45) for Proto-Indo-European also provides a starting point for the developments within its Germanic descendants; Topic and C in Proto-Indo-European motivate the changes that we observe in early Germanic, from the glimpses of earlier Latin-like discourse-configurationality into the ‘syntacticised’ word orders that characterise much of Germanic today.

This syntax in the proto-language would generate clauses in which the grammatical

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

order of constituents for the left periphery would be:

- Initial topic
- Initial marker of clause type
- Topic, then clause-type marker

If enclitic pronouns were present, these would follow in each case of these three cases. If neither a topic nor a clause-type marker were present, the enclitic pronoun would appear after the next accented word.

Following the lexico-phonological reconstructions adopted by Ringe (2006), we can illustrate these abstract orders with lexical elements reconstructed for Proto-Indo-European, such as with a *wh*-pronoun that types its clause as interrogative.¹⁷

- (46) a. *yugóm* *g^wén* *b^héreti*
 yoke.N.ACC.SG woman.F.NOM.SG carry.PRS.IND.ACT.3SG
 ‘A woman is carrying the yoke’
 (Proto-Indo-European. Artificial example.)
- b. *k^wíd* *g^wén* *b^héreti?*
 what.N.ACC.SG woman.F.NOM.SG carry.PRS.IND.ACT.3SG
 ‘What is a woman carrying?’
 (Proto-Indo-European. Artificial example.)
- c. *g^wén* *séh₂* *k^wíd* *b^héreti?*
 woman.F.NOM.SG that.F.NOM.SG what.N.ACC.SG carry.PRS.IND.ACT.3SG

¹⁷This vocabulary is reconstructed on the basis of the whole family, including the Anatolian branch. As admitted here and in §1.4, the syntactic reconstructions cannot claim the same antiquity.

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

‘That woman, what is she carrying?’

(Proto-Indo-European. Artificial example.)

- d. $g^wén$ $séh_2$ $k^wíd$ = moy
 woman.F.NOM.SG that.F.NOM.SG what.N.ACC.SG I.DAT.SG
 $b^h éreti?$
 carry.PRS.IND.ACT.3SG
 ‘That woman, what is she bringing me?’

(Proto-Indo-European. Artificial example.)

In (46b), we have a reconstructed question, with the interrogative object pronoun $k^wíd$ undergoing *wh*-movement to CP. In (46c), we have the same question, but with an initial topic preceding the pronoun, while in (46d), the unaccented personal pronoun moy undergoes movement to C and follows $k^wíd$ in the linear order.

An obvious criticism of this proposal is that it attempts reconstruction on the basis of a correspondence set that is too small. It is undeniable that (45) should be tested against more early Indo-European languages with sufficiently large corpora, most importantly Hittite and Tocharian; without comparative evidence from the Anatolian and Tocharian branches, we can only reconstruct Proto-*Core*-Indo-European. Yet already, there seems to be support from elsewhere; Garrett (1990, 35) for instance proposes a very similar syntax for the left periphery of the Hittite clause. Following Hale, he adopts a syntactic structure that comprises “Top” and “Comp”; through movement to these positions, he accounts for why “topicalized constituents and no other clause-internal elements may thus precede WH-moved elements”. A complication is that Garrett’s (1990) evidence for Comp seems to be only *wh*-movement, something that subsequent scholars (*inter alios* Huggard (2011)

3.2. TOPIC IN EARLY INDO-EUROPEAN AND PIE

and Sideltsev (2014)) have rejected; several of the examples of initial *wh*-elements that Garrett provides are nominative *kuiš*, and so can be taken instead as examples of Hittite's unmarked SOV order. However, relative clauses, also studied by Garrett (1994), offer a surer basis for a syntax similar to (45) in Hittite. These involve general movement of the relative element up to a C-like "Wh" position, followed by their further movement to "Front" in the case of indeterminate relative clauses. Garrett notes that his operation of movement up to Front is topicalisation from within the clause by another name. From the outset, the relevant Hittite behaviour and analysis seems to bolster the reconstruction of (45) even for PIE in its oldest state.

A final point to make is that the syntax in (45) is intended to pertain to all finite clauses, both main and subordinate, regardless of clause type. For one reason, the evidence for (45) appears across the broad types reviewed (not only declarative clauses, but also interrogative, imperative and subordinate ones). For another, the early Indo-European evidence, especially from Vedic Sanskrit, suggests that there was no syntactic distinction between clause types in the proto-language. This relates to long-established views that Proto-Indo-European did not embed finite clauses as complements of verbs or nouns, and that at the syntactic level all finite clauses were equal and main-clause-like (cf. Clackson, 2007, 171-6).

"It is thus highly plausible that Proto-Indo-European made relatively little use of finite subordinate clauses, preferring instead various non-finite constructions"
(Comrie, 1998, 94)

3.2. *TOPIC IN EARLY INDO-EUROPEAN AND PIE*

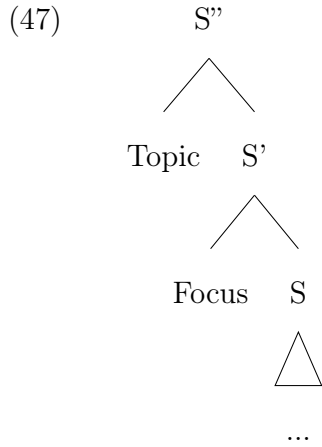
A definitive opinion on this issue is outside the scope of this thesis, but broadly the view is one of agreement. The only stipulation though is that, while Proto-Indo-European may not have had distinct syntactic subordination, it almost certainly had semantic subordination, conveyed by lexical material in the CP.

Other scholars may argue that (45) does not hold for the Indo-European languages that they study, which better support analyses with a more or less articulated syntax for the left periphery. The first response will be to suggest that these are subsequent modifications; (45) is intended as a common starting point for early Indo-European finite clauses, not the final word. For one thing, as clausal embedding developed in Indo-European, the underlying syntax surely developed in tandem; subordinate clauses in some descendants may have retained CP but gradually lost TopicP. This seems especially true of Ancient Greek and early Germanic, in which pre-subordinator topical material is very rare. To lose TopicP in the new subordinate contexts would be a reasonable change, since fronting material to the high TopicP of an embedded clause would run the risk of its being interpreted as part of the higher clause. One thing at least is clear: more Indo-European languages and further work to corroborate, improve, critique or refute (45) are necessary.

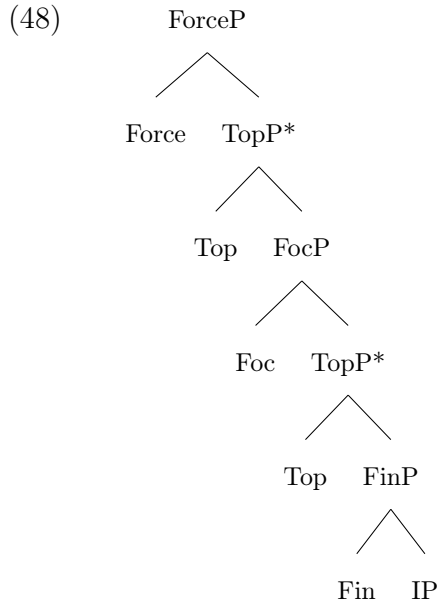
3.3 FOCUS IN EARLY INDO-EUROPEAN AND PIE

3.3.1 FOCUS AND ITS FUNCTIONS

We turn now to the matter of focus. This is a consistently present phenomenon in the literature of early Indo-European syntax, both in and out of the generative tradition. The main claim of this section is that a Focus position was a component of Proto-Indo-European syntax, but, contrary to previous reconstructions, it was not situated within the high clausal structure that produced the left periphery. This section aims to demonstrate instead that much of the behaviour ascribed to focus better suits a derivation from lower down in the underlying structure, and that the arguments for a focus position in the left periphery of PIE are consequently weakened. The section proceeds by first arguing for low Focus, then reviewing proposals for high Focus. The general conclusion is that we do not have the same strength of evidence to reconstruct high Focus; this is a deliberate departure from Kiparsky's (1995) syntactic proposal for Proto-Indo-European (47).



It also rejects the adoption *a priori* of the full Rizgian (1997) Split CP (48) for early Indo-European and the proto-language, on the grounds that the only heads directly supported by the comparative evidence is “Top” and either Force or Fin. The full structure is unnecessary.



3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

Focus is cross-linguistically a key component of clauses; while a clause can have multiple topics (or none), it can only have one focus (Kiss, 1995b, 17-8). Its singular status as “the most important or salient parts of what we say about the topical things” (Dik, 1997, 310) means that the focus is often distinguished from the rest of the clausal material with distinct prosody and/or word order. As mentioned in §3.1.2, it is often treated as the pragmatic ‘opposite’ to topics, with one providing old information, the other new. A qualitative difference between topics and foci though is that focus plays a role in the semantic composition and truth conditions of a clause; a focus implies a set of alternative propositions (Rooth, 1992; Krifka, 1993), and different foci imply different sets. We see this in (49) for English, which can achieve focal interpretations through emphatic prosody.

(49) a. *Mary saw [John]*

Implies: Mary did not see anyone else within a set of possible people, so *Mary saw Bill* is false.

b. *Mary [saw] John*

Implies: Mary did not take any other action towards John within a set of possible actions, so *Mary talked to John* is false.

Nonetheless, in the arrangement of the early Indo-European finite clause, topic and focus have long been considered jointly and as equal partners in discourse information, both able to determine word order. As with topics, we should first acknowledge the challenges in identifying foci within our sources as non-native speakers of these languages.¹⁸ One method

¹⁸Prosody is not available, nor is posing question tests to native speakers, in which minimal answers to

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

is identifying focalising words, akin to English *even* or *only*, which emphasise and create the inferences that are typical of foci. Another method is specific word-order constructions, such as clefts, akin to English *it is a car that I bought*. The main means though is textual context. If the focus of a clause is “the most important or salient parts” of a known topic, then we can try to identify what among the constituents of a clause is topical and known, and then what within the new information is the most important for the purposes of the text and the author.

3.3.2 LOW FOCUS IN PROTO-INDO-EUROPEAN

With all this in mind, the general observation is that foci come after topics in the early Indo-European clause. A Topic-Focus sequence is widely accepted as the norm, and is the natural order, according to the Greek author of *On Style*. Dik’s influential schema for Ancient Greek continues this ancient view:

- (50) $P1 \quad P\emptyset \quad V \quad X$
Topic Focus Verb Remaining unmarked constituents
(Dik, 1995, 12)

The trio of topic, focus and verb is the “pragmatic nucleus” (Torrego, 2017), after which comes material that has a minor discourse-informational role to play. (50) has been widely considered the typical surface configuration for early Indo-European in general (namely for Hittite, Sanskrit, Ancient Greek and Latin), although not without amendments; for *wh*-questions single out the key information of a clause.

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

Ancient Greek, Matić (2003) distinguishes between Narrow Focus and Broad Focus. The former is used when only an argument of the verb is the clausal focus, whereas the second configuration focuses a larger unit, comprising the verb itself and its arguments and perhaps adjuncts. Additionally, Lühr (2015, 3) separates “new-information focus” from “contrastive focus” in her analysis of Ancient Greek, Hittite and Sanskrit. Focus may not have been a unitary phenomenon, but it certainly had great weight in determining word order across early Indo-European languages.

Given the association between focus and distinct word-order positions in certain modern languages, some scholars in the generative tradition (cf. Antinucci and Cinque, 1977; Belletti and Shlonsky, 1995; Rizzi, 1997) account for focal word order with a dedicated Focus position or projection. For early Indo-European languages specifically, because foci frequently follow topics immediately, some have posited a Focus component for the left periphery, high in the syntax, including Oniga (2014) for Latin and Ram-Prasad (2022) as part of his Rizzian approach. This is notably the view of Kiparsky (1995, 14), who posits a high Focus for Proto-Indo-European itself in (47). His reconstruction has its successes; through movement into its positions, it will generate the Topic-Focus-Verb surface order that early Indo-European languages consistently display in declarative main clauses. This is typically, but not necessarily, a Subject-Object-Verb order too, as in (51).

- (51) a. *[Ad eās rēs cōnficiendās]*
to this.F.ACC.PL thing.F.ACC.PL accomplish.GER.F.ACC.PL
[Orgetorix] dēligitur. [Is sibi]
Orgetorix.M.NOM.SG choose.PRS.IND.PASS.3SG this.M.NOM.SG self.DAT.SG

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

[*lēgātiōnem* *ad cīvitātēs*] *suscipit.*
 embassy.F.ACC.SG to city.F.ACC.PL take-up.PRS.IND.ACT.3SG
 ‘To accomplish these things, Orgetorix was chosen. He took upon himself the
 office of ambassador to the cities’

(Latin. Caesar, *Gallic War* 1.3)

- b. [*pròs taúta*] [***Phalīnos***] *eîpe:*
 for this.N.ACC.PL Phalinos.M.NOM.SG say.AOR.IND.ACT.3SG
 [*basileùs*] [***nikân***] *hēgeítai,* *epei*
 king.M.NOM.SG victor.M.ACC.SG believe.PRS.IND.MED.3SG since
Kûron *apékteine*
 Cyrus.M.ACC.SG kill.AOR.IND.ACT.3SG
 ‘To these things Phalinos responded: “The king considers himself victor, be-
 cause he killed Cyrus”’

(Ancient Greek. Xenophon, *Anabasis*, 2.1.11)

- c. [*sahasrākṣó* *vīcarṣaṇir* *Agnī*]
 thousand-eyed.M.NOM.SG active.M.NOM.SG Agni.M.NOM.SG
 [***rākṣāṃsi***] *sedhati*
 Rakshasa.M.ACC.PL drive.PRS.IND.ACT.3SG
 ‘Thousand-eyed, vigorous Agni drives away the Rakshasas’

(Sanskrit. RV 1.79.12)

In (51), given the wider context, we may identify the constituents in bold as foci, and since they immediately follow the preceding topical material (except *Kûron* in (51b)), we can derive them straightforwardly through the high Focus position of (47).

However, an issue with (47) is that its TopP-FocP configuration predicts that that foci will always follow topics immediately. This is too restrictive; in those clauses with more constituents than only three, the immediately post-topical constituent often does not fit

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

an interpretation as the clausal focus.

- (52) a. *Caesar* [prīmō] [et propter multitudinem
 Caesar.M.NOM.SG at-first and because-of multitude.F.ACC.SG
hostium et propter eximiam op̄iniōnem
 enemy.M.GEN.PL and because-of exceptional.F.ACC.SG opinion.F.ACC.SG
virtūtis] [proeliō supersedēre] statuit.
 valour.F.GEN.SG battle.N.ABL.SG refrain.PRS.INF.ACT set.PRF.IND.ACT.3SG
cotīdiē tamen...
 daily however
 ‘Caesar at first, because of both the great number of the enemy and their
 exceptional reputation for bravery, decided to refuse battle. Daily, however...’
 (Latin. Caesar *Gallic War* 2.8)
- b. *Periandron* =dè [hupò apistiēs] [Ariona =mèn]
 Periander.M.ACC.SG and under mistrust.F.GEN.SG Arion.M.ACC.SG indeed
 [en phulakê_i] êkhein, oudamê_i metiénta
 in prison.F.DAT.SG have.PRS.INF.ACT nowhere let-go.PRS.PACT.ACC.SG
 ‘But Periander, out of mistrust, put Arion in prison, free to go nowhere’
 (Ancient Greek. Herodotus, *Histories* 1.24)
- c. *tā* =asya [nāmasā] [sāhaḥ]
 this.F.NOM.PL he.GEN.SG bowing.N.INS.SG power.N.ACC.SG
saparyānti prācetasah
 honour.PRS.IND.ACT.3PL wise.F.NOM.PL
 ‘They with reverence honour his might, those who are wise’
 (Sanskrit. RV 1.84.12)

In (52a), the initial subject is an aboutness and anaphoric topic, being a known individual, while the much later verbal phrase *proeliō supersedēre* is the most important information of the clause, and it functions as a contrastive focus that anticipates Caesar’s change of plans.

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

In (52b), the post-topical prepositional phrase *hupò apistiēs* is unlikely to be the focus; the object *Aríona* bears the emphasising particle *mén*, while the later new information of Arion’s unfortunate location, *en phulakêi*, is most important for the overall narrative. In (52c), *sáhas* is the more likely candidate for focus over *námasā*, since this clause and the clauses that precede and follow it each concern attributes of the overarching addressee of the hymns, Indra. Granted, these are not certain readings; clearer examples with explicit focal elements are needed for a more confident identification.

If we do accept that the aforementioned clause-late constituents are indeed focal, this is reminiscent of the generalisation that, both in and outside of Indo-European languages, focus is associated primarily not with the post-topical constituent, but rather the immediately preverbal position in the linear order. This pertains in particular to languages with a typical, unmarked SOV order (Dezső, 1978; Kim, 1988). Hock (1991, 476-7) states that the clause-final verb in SOV-type languages has the “lowest prosodic prominence” within the overall clause, while what precedes the verb is the often the most important and receives the highest prominence. This view of preverbal focus, developed on the basis of modern languages, has in turn been recognised for historical Indo-European languages, namely Latin (Torrego, 2017), Ancient Greek (Goldstein, 2015, 2016), Sanskrit (Hock, 1992; Lühr, 2015) and Hittite (Sideltsev, 2015a; Lühr, 2015).¹⁹

Having identified preverbal focus as the common feature of ancient Indo-European word

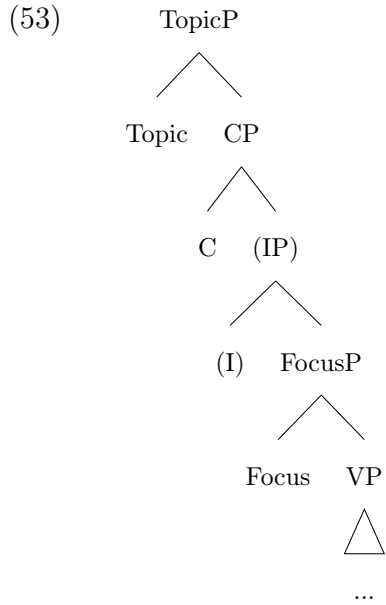
¹⁹However, a postverbal focal position has also been identified in early Indo-European. Pinkster (1991, 76-6) gives clear examples of foci in clause-final position, and his summary of the Latin evidence is that there is no single position for focus (Pinkster, 2021, 959). Likewise, our Ancient Greek sources display a shift from mostly pre- to mostly post-verbal foci over time (Lühr, 2015). The same diachrony in Ancient Greek may account for the variation found in Latin too.

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

order, we can derive this focal behaviour with a low Focus component in the underlying syntax. An alternative analysis is to derive it from the structural properties of the syntax, without positing a dedicated Focus; this would build on the Nuclear Stress Rule (Cinque, 1993), in which the main stress of an utterance is assigned to the constituent that is most deeply embedded in the underlying structure. This is typically the object, being the complement of the low VP. However, a specific Focus position gives more flexibility to what can be focal in early Indo-European – not only objects, but also subjects, adjuncts and various subordinate clauses. It also avoids committing the early Indo-European VP to a particular syntactic configuration for which there is insufficient evidence.

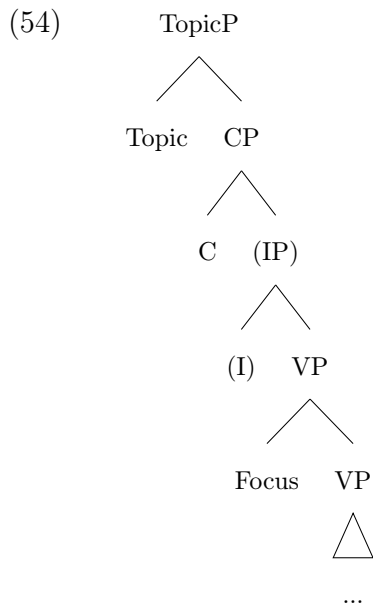
Hence, this thesis additionally suggests that the syntax of early Indo-European languages and of Proto-Indo-European itself included (at least) a low Focus component, which produced the word-order behaviour of the clausal foci in the examples given. This is an analysis with strong theoretical precedents, especially for the syntax of Italian; Belletti (2001, 2004), Bianchi (2013) and Cardinaletti (2018) have proposed and utilised a low Focus in the syntactic hierarchy, at or close to the edge of the VP. The syntactic structures in (53) and (54) illustrate two ways to model low Focus in Proto-Indo-European, in conjunction with the syntax proposed in §3.2.4.²⁰

²⁰“(IP)” and “(I)” are intended to leave room for the possibility of further functional categories within the clausal syntax of Proto-Indo-European, which would be responsible for verbal features like tense and aspect, as well as the word-order behaviour of auxiliaries and adverbs. Reconstruction of the IP, with a view on the relative order of its head and complement, is outside the scope of this thesis; for a thesis that does tackle this issue, see Windhearn (2020).



(53) specifically models Focus as constituting a separate phrase, with a null head. Alternatively, if we do not wish to include a null Focus head and a dedicated FocusP, we could instead derive preverbal foci from movement to a focus-associated adjunct or specifier position at the edge of the VP. This keeps the focal constituent within the VP, the domain of the propositional content of the clause.

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE



In both scenarios, the subject, object and verb are base-generated within the VP structure. From this origin, the arguments can move up to Spec,TopicP and the Focus position; it is usually the subject that moves up to the TopicP, while the object occupies Focus, producing the typical SOV order, but not necessarily so.²¹ What is important for this thesis is that the syntactic source of Focus is some way below the level of Topic and C in (45), outside the syntax that generates the left periphery. Low Focus will interact with the prosodic organisation of the overall clause, and receive the high prominence noted for preverbal constituents by Dezső (1978), Kim (1988) and Hock (1991).

²¹Since it is TopicP that produces the typical clause-early position of subjects in early Indo-European, an additional syntactic motivation, such as an equivalent to the Extended Projection Principle for English (Chomsky, 1982), seems redundant. If we assume an IP component for the syntax, IP can still provide verbal inflection and nominative case through downward feature checking with the verb and subject in the VP.

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

If we assume this view of low Focus, the common adjacency of topics and foci is largely an accident; at the syntactic level, they stand at a distance, but the brevity of many clauses leaves them adjacent in the surface order. This is especially true of declarative main clauses, in which C is usually lexically null in early Indo-European. (54) predicts that when C is lexically realised, it will precede the focus. This allows us to identify foci in Latin and Ancient Greek subordinate clauses with lexical subordinators, such as the object *Kûron* in (51b), which follows the subordinator *epeí*.

3.3.3 AGAINST A HIGH FOCUS FOR PROTO-INDO-EUROPEAN

A low Focus can derive a substantial amount of the word-order behaviour of foci in the early Indo-European languages considered so far, while not being bound to strict Topic-Focus initial sequences. All this prompts the question of whether Kiparsky's schema in (47), specifically its high Focus, is still theoretically necessary. This subsection presents a variety of evidence of possible focal behaviour in the early Indo-European left periphery, namely: *wh*-questions, negative imperatives, fronted preverbs and exceptional Homeric enclitic positions, as well as a notable absence of evidence in Old English, Old Norse and Old Irish.

One reason to continue to posit a high Focus is the derivation of *wh*-questions. There is a long-standing association between *wh*-words and the phenomenon of focus; as Rooth (1992, 10) puts it, "there is a correlation between questions and the position of focus in answers ... a question determines a set of potential answers". This is to say, the *wh*-word

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

of a question delimits a set of possible alternatives, from among which the focus of the answer then selects one.

(55) **Who** *did Mary meet?* / *Mary met* [**Bill**]

In (55), the *wh*-pronoun *who* defines a set of possible entities, namely humans that are the objects of meeting. A focus that does not match the features of the set defined by the *wh*-word produces an infelicitous answer.

(56) **Who** *did Mary meet?* / #*Mary* [**met**] *Bill*

Given this association between the two phenomena, it would be reasonable to propose that a *wh*-word or phrase is the focus of its clause and stands in the Focus component of its clausal syntax. Given the clause-early position of *wh*-material in early Indo-European, one might further posit that a high Focus motivates the upward movement of *wh*-material. Since we should assume *wh*-movement for the proto-language, this lends support to Kiparsky's reconstruction in (47). Yet this is at odds with the case made in Chapter 2, that C and its clause-typing function are the source of early Indo-European *wh*-movement.

There is a precedent analysis that incorporates both the role of focus and C in *wh*-questions and negates the need for high Focus. On the basis of very rare examples with clause-medial *wh*-material, Bertrand and Faure (2022) propose that *wh*-material in Ancient Greek moves from its base position up to Focus, but this is only a “necessary stopover” before it then continues even higher into a CP layer; “*wh*Ps must go through all three

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

positions, receiving their argument function *in situ*, checking their focus feature preverbally and verifying their *wh*-feature in the high position” (Bertrand and Faure, 2022, 1). Their analysis is therefore perfectly congruous with the syntax in (54), which contains both of the components where *wh*-material can move up from a VP-internal argument position, to achieve its two features of focus and interrogativity. There is no apparent word-order evidence that would prohibit us in applying this syntactic analysis of two-step movement for Ancient Greek to Latin and Sanskrit, and to a later stage of Proto-Indo-European itself.

Hittite may represent an archaic stage in this syntactic story. As mentioned in Chapter 2, Hittite does not display the same *wh*-movement as seen elsewhere in early Indo-European. The analysis of Huggard (2011) is that Hittite either leaves *wh*-material *in situ* or in the dedicated preverbal position for foci.

- (57) a. ^dU ^{URU}Nerik TUKU-an ZI-an [kuēzza] KASKAL-ahmi?
 Stormgod of-Nerik angry.ACC.SG soul.ACC.SG what.ABL appease.PRS.1SG
 ‘By what means can I appease the angry soul of the Stormgod of Nerik?’

(Hittite. KUB 5.1 i 92–3, oracular inquiry, from Huggard (2011, 7))

- b. KUR.KUR.ḪI.A [=m]a ḫūman [kuiš] ḫarzi?
 country-country-pl. but all.ACC.SG who.NOM.SG hold.PRS.3SG
 ‘Who holds all of the countries?’

(Hittite. OH/NS legend, #13, from Goedegebuure (2009, 3))

If Hittite *wh*-material undergoes *wh*-movement only as far as low Focus, this might attest to an archaic grammar, modified in the rest of Indo-European; in the later stage common to the Core Indo-European languages, the original *wh*-movement was extended beyond

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

Focus to target high C, where it would help to convey the interrogative type of its clause. This is speculation though; the main point here is that *wh*-questions do not require a high Focus in the syntax to keep their focal character.

Secondly, Ram-Prasad (2022) discusses another category of words that appears in the left periphery and suits a focal interpretation: prohibitive negators. Specifically, the Sanskrit negator *mā* negates imperative, subjunctive, optative and injunctive verbs, thereby creating prohibitions and negative wishes, and in this role it often stands initially and apart from the verb.

- (58) a. *mā* = *naḥ śáṃso áraruṣo dhūrtíḥ*
 not we.ACC.PL curse.M.NOM.SG resentful.M.GEN.SG damage.F.NOM.SG
práṇañ mártiyasya
 reach.AOR.INJ.ACT.3SG mortal.M.GEN.SG
 ‘May the curse of the resentful one, nor the affliction of the mortal reach us’
 (Sanskrit. RV 1.18.3)
- b. *mā* = *te rádhāṃsi, mā* = *ta ūtáyo*
 not you.GEN.SG gift.N.NOM.PL not you.GEN.SG support.F.NOM.PL
Vaso asmān kádā-canā dabhan
 Vasu.M.VOC.SG we.ACC.PL whenever deceive.AOR.INJ.ACT.3PL
 ‘May your gifts, may your help never let us down, oh Vasu’
 (Sanskrit. RV 1.84.20)

There is evidence also for a similar construction with the Ancient Greek cognate *mē*, at least in Homeric Greek.

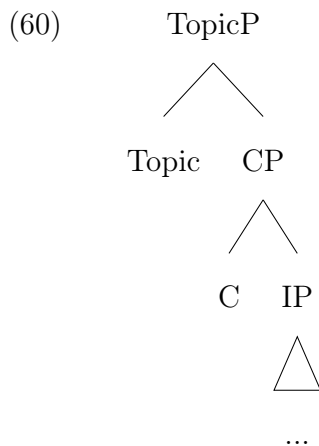
- (59) *mē* = *tis opíssō tetráphthō poti nêas*
 not someone.NOM.SG backwards turn.PRF.IMP.MED.3SG towards ship.F.ACC.PL

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

‘Let no man turn back towards the ships’

(Ancient Greek. Homer, *Iliad* 272-3)

In later Classical Greek, this construction seems no longer productive, and *mĕ* and its verb instead stand adjacent. This construction in the two languages raises the question of where in (45) (repeated in (60)) the negators in (58) and (59) might stand.



For Ram-Prasad (2022, 100), this is simple: they occupy his high FocP. This is somewhat appropriate, since this prohibitive element presumably had an emphatic intonation, typical of a clausal focus. Yet other constituents might be better suited to a focal interpretation; the adverb *kádā-canā* ‘sometime’ (or ‘never’ in this negative environment) occupies the preverbal position in (58b) and is plausibly focal. Regardless of what we identify as the focus in *mā/mĕ*-clauses, we can consider the negator instead to be an exponent of C. This fits its function as an indicator of clause type, signalling to the audience that its clause is a negative command. This alternative analysis does at least demonstrate that we can do

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

without a high Focus for deriving initial *má* and *mé*.

A third phenomenon with potential focal properties is the fronting of preverbs, mentioned in §3.2.4. The arrangement is found across early Indo-European in which preverbs (or ‘local particles’) are associated with a finite verb, yet stand apart from it. This word order, with a clause-initial preverb and typically clause-final verb, is referred to as *tmesis*, and has been recognised as a common word order of early Indo-European since Wackernagel (1892). It is clearly productive in Sanskrit and Avestan (cf. Hale, 1993), and to a lesser degree in Ancient Greek.

- (61) a. *en gâr Pátroklos phóbon hēken*
in for Patroclus.M.NOM.SG fear.M.ACC.SG send.AOR.IND.ACT.3SG
hápasin
all.M.DAT.PL
‘For Patroclus had sent fear into them all’
(Ancient Greek. Homer, *Iliad* 16.291, from Ram-Prasad (2022, 16))
- b. *katá =me phónios Aídas héloi...*
down I.ACC.SG bloody.M.NOM.SG Hades.M.NOM.SG take.AOR.OPT.ACT.3SG
‘If only murderous Hades might take me down...’
(Ancient Greek. Sophocles, *Oedipus at Colonus* 1688-9)
- c. *ví jánāñ chyāvāḥ śitipádo*
apart people.M.ACC.PL brown.M.NOM.PL white-footed.M.NOM.PL
akhyan
look.AOR.IND.ACT.3PL
‘The white-footed black horses looked at the people...’
(Sanskrit. RV 1.35.5)
- d. *prá yó =vām Mitrāvaruṇā*
forth which.M.NOM.SG you.ACC.DU Mitra-Varuna.M.VOC.DU

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

ajiró *dūtó* *ádravat*
 quick.M.NOM.SG messenger.NOM.SG run.IPRF.IND.ACT.3SG
 ‘The quick messenger who ran forth to you, O Mitra and Varuna...’
 (Sanskrit. RV 8.101.3)

- e. **ā** = *mōi* *rafəðrāi* *zauunəng* *jasatā*
 towards I.GEN.SG support.N.DAT.SG call.M.ACC.PL come.PRS.IMP.ACT.2PL
 ‘Come here to my calls for support’
 (Old Avestan. Yasna 28.3, from West (2011, 111))

The examples in (61) illustrate this fronting, with preverbs like *katá* and *prá* contributing a direction to the action of the verb, yet standing at a distance from it. A similar order also seems productive in Hittite (Melchert and Hoffner, 2008, 364-6, 410). In Sanskrit, a PV-*yá*- sequence can even be preceded itself by an initial topical phrase.

- (62) [*divyā* *ápo*] **abhí yád** = *enam*
 heavenly.F.NOM.PL water.F.NOM.PL PV when he.ACC.SG
áyan
 go.IPRF.IND.ACT.3PL
 ‘When the heavenly waters came upon him’
 (Sanskrit. RV 7.103.2, from Lowe (2014, 2))

Furthermore, traces of old tmesis can be found in other branches of Indo-European, namely in Latin, Gothic and Old Irish.

- (63) a. **sub** *vōs* *placo* ... **ob** *vōs*
 under you.ACC.PL fold.PRS.IND.ACT.1SG towards you.ACC.PL
sacro
 devote.PRS.IND.ACT.1SG

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

‘I entreat you ... I beseech you’

(Latin. Festus, *De Verborum Significatione* 190.8-9, 309.30-2, from Wackernagel (1892))

- b. ... **ga** =*u* =*hwa* *sēhwi*
 PRT Q something see.PST.SBJ.ACT.3SG
 ‘... whether he saw anything’

(Gothic. Mk. VII.23, from Wackernagel (1892))

- c. **for**=*don itge* *Brigte* **bet**
 PV-us prayers.M.NOM.PL Brigid.F.GEN.SG be.PRS.SBJ.3PL
 ‘May Brigid’s prayers be upon us’

(Old Irish. Thes. II.348.4, from Thurneysen (1946, 257))

In the Latin of (63a), we find two rare examples, explicitly noted by Festus as the language of prayers²², of elements that are usually prefixed, here separated from their verb. Old Irish tmesis, as in (63c), may be a similar archaism preserved in poetry and the alliterative *rosc* style, if it is not instead an innovation influenced by Latin (cf. Breatnach, 1984; Corthals, 1996). Given this broad distribution, preverb fronting must be an archaic order, reconstructible back to the ancestral grammar, in which preverbs had more word-order freedom, and the separate classes of adpositions for nouns and prefixes for verbs were yet to emerge (Fortson, 2006, 230-1). To be precise, only one preverb is fronted, and it occupies a position before WL1 pronouns (e.g. *me* in (61b) and *don* in (63c)). It can also precede the relative *yá-* element in Sanskrit, as in (61d).

²²“*Ob vōs sacrō in quibusdam precātiōnibus est prō vōs obsecrō*” (‘*Ob vōs sacrō* is in certain prayers in place of *vōs obsecrō*’, 190.8-9)

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

All this prompts the question of what this fronting achieved in early Indo-European and Proto-Indo-European itself; an answer is necessary if we are to integrate them into the syntactic structure. Yet the purpose is not at all obvious; a preverb may have moved to fulfill a prosodic requirement, to be emphasised, to secure a particular interpretation, or to be pragmatically marked.

For Ram-Prasad (2022), fronted preverbs are focal. This analysis builds on the observation of Lowe (2014) for Sanskrit that, while this fronting of preverbs is possible in relative *yá*-clauses like (61d), it never appears in interrogative *ká*-clauses. This is unexpected in light of the analysis adopted herein that *yá*- and *ká*- move to CP uniformly. In both relative and interrogative clauses, their movement leaves the higher TopicP free to host the preverb. In other words, the syntax (60) offers no reason to block *PV-*ká*-orders. Lowe (2014) and Ram-Prasad (2022) have different accounts for this divergence. Lowe's analysis relies on the view that relative *yá*-pronouns have the option to be enclitic. If so, in clauses like (61d), *yá*- and the WL1 pronoun move to the C head, and the preverb stands in Spec,CP. Meanwhile, *ká*-pronouns are non-enclitic and only move to Spec,CP. Ram-Prasad (2022, 98-9) however refutes Lowe's evidence for the optional enclitic status of *yá*-, arguing that the evidence from sandhi and the diachronic development of *yá*- do not support it. Instead, Ram-Prasad (2022, 100) derives the word order through the different syntactic positions of the Rizgian Split CP. His claim is that *yá*- and *ká*- do not in fact move to the same position (Spec,CP in Chapter 2); rather, *ká*- lands higher in the syntax, specifically in Rizzi's Spec,FocP. This is based on the aforementioned association between focus and interrogativity. If Spec,FocP is also the target of fronted preverbs, this explains

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

the lack of *PV-*ká-* orders: they are in competition for the same left-peripheral position. If *ká-* is present, it takes precedence and occupies Spec,FocP as a focal element, while the preverb remains low.

However, it remains highly unclear as to how fronted preverbs may be foci. They may be emphatic, but there is more to focus than prosodic prominence; how they would interact with the semantic composition of the clause needs addressing. There is an alternative analysis that would avoid this issue, and keep *wh*-movement in Sanskrit unified. The reason for the absence of *PV-*ká-* may not be syntactic, but rather pragmatic. Unlike *yá-*, the *ká-* stem has multiple meanings. Elsewhere in early Indo-European, its cognates (Latin *quis*, Ancient Greek *tís*, Hittite *kuiš*) serve not only as interrogative elements, but also as indefinites; the intended reading is expressed through a combination of word order and prosody. While not as common as in the other languages, an indefinite reading for Sanskrit *ká-* is attested too (Monier-Williams, 1899, 240). The lack of *PV-*ká-* may in fact be an avoidance tactic, to ensure the interrogative reading that we know its cognates gain by a clause-early position. While permitted by the syntax, if *ká-* appeared after a fronted preverb (itself in the TopicP), it might be interpreted as lower within the clause and therefore as indefinite. This is a speculative analysis, but it does allow us to maintain the structure in (60), and to motivate the word-order differences from the features of the lexical items themselves.

Strictly in terms of their relative word order, fronted preverbs do not contradict the syntactic structure in (60); numerically at least, its positions can accommodate orders with

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

fronted preverbs. We may locate them in TopicP or CP.²³ In terms of their interpretation, fronted preverbs seem ill-suited to the CP, which hosts markers of clause type and status; unlike interrogative *ká*-elements and relative *yá*-elements, preverbs have no association with any particular type of clause. TopicP is the better candidate for their syntactic position, but any suggested motivation for their position there would be speculative.²⁴ What is certain is that more dedicated work on fronted preverbs is needed; the matter of whether fronted preverbs support a high Focus must remain unresolved until further work convincingly explains why they are fronted at all.

A fourth phenomenon that may necessitate an addition to the syntactic structure in (60) is the exceptional placement of Wackernagelian pronouns, in particular those in Homeric Greek noted by Taylor (1990). According to Taylor, 90% of enclitic pronouns appear following the first prosodic word of their clause; yet we also find anomalous orders in which the WL1 pronoun comes after the second. Assuming the syntactic structure in (60) and the overall syntacto-prosodic approach of this thesis can account for many of the exceptions, but it struggles with those like (64).

- (64) a. *epeì ou̐ =sphi thalássia érga*
 since not they.DAT.PL sea.N.NOM.PL work.N.NOM.PL
memélei
 be-care.PPRF.IND.ACT.3SG
 ‘Since works of the sea had meant nothing to them’

²³For Sanskrit, having located relative *yá*- in Spec,CP, we are left with only TopicP for the preverbs that often precede *yá*-.

²⁴For one thing, it could not account for why interrogative *ká*- can appear after other initial constituents, such as in the questions with nominal topics identified in §3.2.1.

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

(Ancient Greek. Homer, *Iliad* 2.614, from Taylor (1990, 42))

- b. *all' hóte dè dekátē =moi epéluthe*
 but when indeed tenth.NOM.SG I.DAT.SG come-upon.AOR.IND.ACT.3SG
nùx erebenné
 night.F.NOM.SG dark.F.NOM.SG
 ‘But when the tenth dark night came upon me’

(Ancient Greek. Homer, *Iliad* 9.474, from Taylor (1990, 41))

- c. *oud' ei mála =min khólos híkoi*
 not if very she.ACC.SG anger.M.NOM.SG come.PRS.OPT.ACT.3SG
 ‘Not even if anger greatly came upon her’

(Ancient Greek. Homer, *Iliad* 17.399, from Taylor (1990, 40))

These are problematic orders; having located the subordinators *epei*, *hóte* and *ei*, and the WL1 pronouns *sphi*, *moi* and *min* all in a unitary C, the intervening words are unexpected. Consequently, Ram-Prasad (2022, 106-7) sensibly locates *oú*, *dekátē* and *mála* in the high Focus of his Rizgian structure, while subordinators like *epei* stand even higher in Force; it is the lower head that is the target of WL1 pronouns. A focal status is not an unreasonable reading for these words.

The word orders in (64) remain rare though, and we should consider alternative motivations before proposing an otherwise unnecessary high Focus for the Ancient Greek left periphery, and especially before relying on them to build a case for high Focus in PIE. For one, other syntactic derivations are available that do not require additions to the syntax. Taylor (1990, 52-63) proposes that WL1 pronouns move syntactically only as far as the left edge of the IP layer. For this reason, they may surface as the first or the second element of those located in the IP; this variation is conditioned by whether “another word precedes

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

without a pause”. For Taylor then, an order like (64a) is produced through the movement of *sphi* to the left edge of the IP, followed by its prosodic inversion with *ou*, presumably because the C-element *epei* is followed by an intonation break. This naturally prompts the question of why a subordinator should be prosodically so separated, but the main point to consider here is that there might be additional prosodic operations at work that produce orders like (64).

This is the view of Goldstein (2010), whose account for Classical Greek gives more weight to prosody in determining enclitic position. It includes a hierarchical syntactic structure similar to (54), with “strong topics” above a CP layer and “strong focus” beneath it (2010, 76-9). Yet, in Goldstein’s view, prosody can modify what the syntax has produced through post-syntactic processes (cf. Embick and Noyer, 2001). Most notably, this is true of metrical language, such as Homeric epic verse, in which fitting metrical templates trumps all other considerations.

“In the metrical texts, what we find is the ability of a clitic to occur much farther into the clause than would normally be permitted. Rather than occur at the left edge of an intonational phrase derived from a syntactic boundary, clausal clitics can occur at the left edge of a metrical unit (namely the caesural unit), which is not the product of the mapping between syntax and phonology.”
(Goldstein, 2010, 84)

WL1 pronouns may have hosts in Homeric language that they would not have in prose or the spoken vernacular, such as the first accented element after a caesura. Working

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

to a metrical template provides the clause with “more intonational phrases than it would otherwise receive in the mapping from phrase structure to phonological form” (Goldstein, 2010, 97), and therefore more potential hosts for a WL1 pronoun. All this serves to argue that we should not be quick to propose additional syntactic components, such as high Focus, to derive a minority of WL1 word orders in Ancient Greek; we should first examine the qualities of the particular genre of text and consider additional factors, before we modify the syntactic derivation that works so well for the majority of WL1 pronouns, namely their movement to C.

A final argument against the need for reconstructing high Focus comes from an absence of evidence in the rest of the language set for this thesis. If we follow Kiparsky and assume that Proto-Indo-European syntax included a high Focus component, we should expect to find traces of it in Old Church Slavonic²⁵, Old English, Old Norse and Old Irish. Granted, an absence of evidence is not evidence of absence, but it seems diachronically implausible for a high FocusP in the proto-language to be lost across multiple branches of the subsequent language family.

In Old English and Old Norse, the character of the preverbal *Vorfeld* in §3.2.2 in main clauses is overwhelmingly topical.²⁶ In the literature, no previous work seems to have claimed that foci appear in this initial zone too. This leaves little reason to add a FocusP alongside the TopicP and CP proposed in §3.2.4. Kiparsky’s response (1995) to this would be his claim that his Proto-Indo-European FocusP in (47) had actually

²⁵Our sources for Old Church Slavonic are characteristically unhelpful in identifying evidence for foci and movement within the clause, except perhaps the “predicate movement” noted by Willis (2000, 4-5).

²⁶With one exception being *wh*-material, accounted for with a CP.

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

transformed into CP by the time of Proto-Germanic. This diachrony is unconvincing. A high Focus might be a source for Indo-European *wh*-movement and some kinds of verbal fronting (e.g. imperative), but it is a less persuasive origin for the other phenomena argued in Chapter 2 to arise from a CP structure: subordinators, relative elements, interrogative particles, and Wackernagelian pronouns. Reconstructing a clause-typing CP back to Proto-Indo-European itself is a unified means of accounting for all of these early Indo-European phenomena.

Lastly, in Old Irish, as with early Germanic, the clausal left periphery likewise does not play a role in focus. Old Irish resorts to an alternative construction for the purpose of focusing and emphasising a constituent: clefting (DiGirolamo, 2018; Stifter, 2009, 55-6). It is especially common in the Old Irish glosses; this is not surprising, considering “the nature of the activity of glossing, which tends to focus on a particular word or phrase” (Russell, 2014, 287).

- (65) *is* [*tri* *Ísu*] *predchim=se* ...
 be.PRS.IND.3SG through Jesus.M.ACC.SG preach.PRS.IND.ACT.1SG.REL-I
is [*macc*] *míastar*
 be.PRS.IND.3SG son.M.NOM.SG judge.FUT.ACT.3SG.REL
 ‘It is through Jesus that I preach ... it is the Son who will judge’

(Old Irish. Wb. 1d9)

As (65) illustrates, Old Irish cleft constructions comprise the copula verb, the focused element and then a relative clause. Clefting therefore really involves two clauses, one main and declarative, the other subordinate and relative. The fact that Old Irish relies on this bi-

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

clausal construction suggests that fronting for focus within one clause is not an option, and that its high syntactic structure does not include a FocusP alongside the TopicP already established.

Instead, there is tentative evidence that intra-clausal focus in Old Irish occurs later in the linear order and further down in the syntax. Lash (2020) discusses clauses with the rightward dislocation of the subject, in which it appears last or later in a VOS order, in contrast with the usual VSO.

- (66) a. *ro-siacht* *corrici nem* *a trocaire*
reach.PST.IND.3SG up-to heaven.N.ACC.SG his mercy.F.NOM.SG
'His mercy reached up to heaven'

(Old Irish. Ml. 55d2, from Lash (2020, 29))

- b. *fo-sissetar* *a pectū* *ind* *firien*
confess.PRS.IND.3PL their sin.M.ACC.PL the.M.NOM.PL righteous.M.NOM.PL
'The righteous confess their sins'

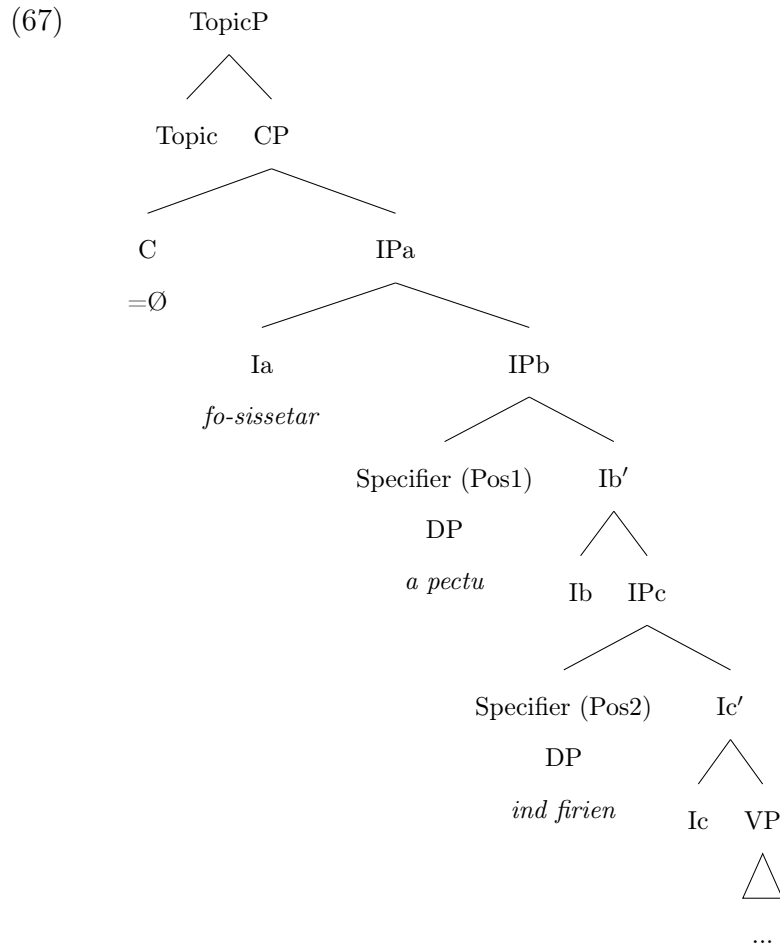
(Old Irish. Ml. 132a1, from Lash (2020, 34))

Lash suggests two reasons for why the subjects in (66) might stand not immediately after the verbal complex. One is the length and complexity of the subject; noun phrases that include a relative clause, a complex adjective or two appositive nouns tend to be clause-final. Their late position may be due to prosody and the rhythm of the overall clause. Yet such factors, Lash (2020, 22) argues, are hard to diagnose in an ancient language, and adds that it is not clear what exact length or degree of complexity constitutes a complex subject. Clause-late subjects therefore “might in fact be better understood by appealing to

3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE

information structure” (Lash, 2020). Lash posits two syntactic positions, Pos1 and Pos2, that arguments and adverbial adjuncts may occupy. Pos1, the hierarchically higher of the two, is used for old information. This is usually the subject, but not always. Pos2, the lower, is for new information. If the subject is not given or focused, it moves to Pos2, while something else may stand in Pos1. We can derive an order like (66b) with a ‘split IP’, containing the targets for the movement of the finite verb (IPa in (67)), old information (IPb) and new information (the lowest IPc).

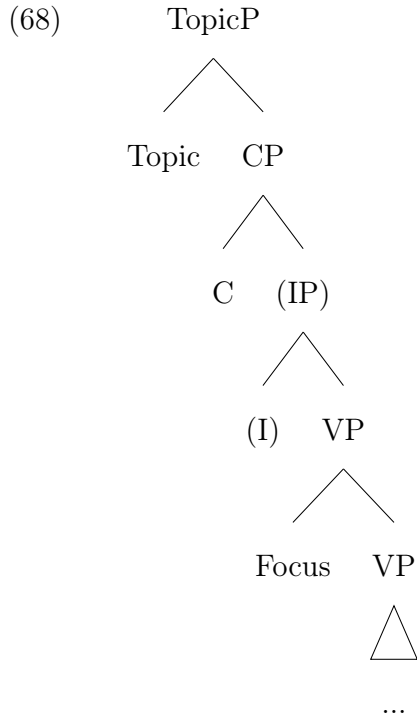
3.3. FOCUS IN EARLY INDO-EUROPEAN AND PIE



Further work can explore the possibility that Lash's Pos2 and IPc in (67) continue the low Focus position of Proto-Indo-European proposed in (54).

3.4 CONCLUSIONS: TOPICS, FOCI AND EARLY INDO-EUROPEAN

In sum, this chapter and the previous one demonstrate that reconstructing Topic, C and a low Focus position back to Proto-Indo-European successfully accounts for the syntactic behaviour that we observe in its descendants, including the relative positions of various kinds of topics and foci, instances of *wh*-movement, and lexical elements of the category C. Low Focus can successfully derive the widely acknowledged preverbal position of early Indo-European foci, the variation between narrow and broad focus, and the focal quality of *wh*-material. In contrast, high Focus does not reach the same benchmark of comparative support that would justify reconstructing it for the proto-syntax. The lack of a need for high Focus in the three more discourse-configurational languages in the set (Latin, Ancient Greek, Sanskrit) is compounded with an apparent lack of foci in the left peripheries of Old English, Old Norse and Old Irish. All of this results in the following abstract syntax for the early Indo-European and Proto-Indo-European finite clause:



The widely acknowledged surface adjacency of topics and foci in Latin, Ancient Greek and Sanskrit is understood to be a syntactic accident, produced by an absence of lexical exponents for the intervening syntactic projections. All of this is a deliberate departure from previous Rizgian analyses of early Indo-European syntax (e.g. Ram-Prasad, 2022), and notably from Kiparsky’s (1995) influential reconstruction of the proto-syntax.²⁷

The idea that the phenomenon of focus arises from the lower regions of the clausal

²⁷Consequently, it can help analyses that build on Kiparsky’s reconstruction. For one example, Erschler’s (2012) diachronic account of innovative *wh*-movement in Ossetic involves a stage in which Middle Iranian languages must develop a second low Focus position in addition to the older high Focus, adopted by Erschler following Kiparsky (1995). Yet, according to the account of this thesis, low Focus was there all along.

3.4. CONCLUSIONS: TOPICS, FOCI AND EARLY INDO-EUROPEAN

syntax may surprise some, since the high syntax and the left periphery that it generates are commonly associated with discourse information. Yet it is not so strange when we consider focus more generally; true, it plays a role in discourse information, but it has qualities beyond the pragmatic. As mentioned at the beginning of §3.3, focus sits at a crossroads between syntax, semantics, pragmatics and prosody, and contributes to the semantic interpretation of a proposition. As such, it takes scope over the constituents that comprise that proposition, namely the lexical components of the VP. Topics meanwhile do not share this semantic role, so it is inaccurate to consider topic and focus as simply two sides of the same discourse-informational coin, with the former providing old information, the latter new. Focus is a beast of a different nature, one that suits a low position in our reconstruction of Proto-Indo-European syntax.

A final word is appropriate on the considerable effects of Topic and Focus, and the consequences of integrating them into the syntax. From Delbrück (1888) onwards, the reconstruction of Proto-Indo-European word order has searched for the original basic arrangement of its subjects, objects and verbs. In the view of Delbrück and many scholars since, Proto-Indo-European word order was fundamentally SOV. Yet a syntactic reconstruction like (68) robs this search of its feasibility. Since Topic and Focus consistently motivate the movement of the lexical constituents of the VP (chiefly, the arguments of the verb), we can never be sure of whether our evidence displays an unmarked arrangement of these constituents. Even the SOV order in early Indo-European may be a product of discourse configurationality, and therefore no less affected; its frequent occurrence may not indicate its syntactic basic-ness, but rather result from the tendency of subjects to be top-

3.4. CONCLUSIONS: TOPICS, FOCI AND EARLY INDO-EUROPEAN

ical and objects to be focal. No arrangement of subject, object and verb, no matter how statistically common, can be confidently claimed in the earliest Indo-European languages to be a product of syntactic derivation unaffected by discourse-driven movement to Topic and Focus.

This thesis does not claim that the Proto-Indo-European VP was unstructured, but only that we today are not now able to look into that VP. Agreeing with Watkins (1976, 316), definitively categorising Proto-Indo-European as either an ‘OV’ or a ‘VO’ language seems to be an unobtainable goal, and the search for a fundamental word order of the clausal constituents may be essentially futile. Perhaps surprisingly to some, it seems more fruitful to hunt instead for syntactic components that determine the observable discourse-informational arrangements of lexical constituents, namely Topic and Focus.

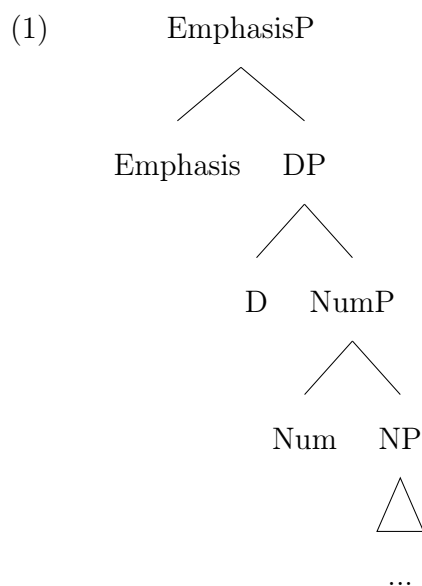
THE LEFT PERIPHERY OF THE NOUN PHRASE IN PROTO-INDO-EUROPEAN

4.1 INTRODUCTIONS AND CHAPTER AIMS

4.1.1 CHAPTER AIMS

This thesis now turns to its second object of study: the noun phrase in early Indo-European. This chapter builds a case for reconstruction based on the left-peripheral behaviour of noun phrases across the language set. It considers all typical constituents of the noun phrase, with a particular focus on functional lexical elements, such as determiners. From the com-

parative evidence of these seven languages, the ultimate goal is to reconstruct aspects of the syntax of the noun phrase in Proto-Indo-European, specifically its topmost components. This is again pursued according to the methodology set out in Chapter 1, namely, by diagnosing functional categories and heads in the underlying syntax that generated the attested word-order data. The main claims of this chapter are that Proto-Indo-European¹ had a configurational noun phrase, and that this crucially contained at least three functional categories: ‘Num’, ‘D’ and ‘Emphasis’. These are arranged hierarchically in the following manner:



The different projections of the syntax in (1) are each assumed to include a specifier position that can host moved elements. (1) is reconstructed as a sufficient common origin

¹Or at least Proto-*Core*-Indo-European, see §1.4.7.

4.1. INTRODUCTIONS AND CHAPTER AIMS

for various observed behaviour across the language set, such as the emphatic quality of initial constituents of the noun phrase, and the typical pre-nominal position of markers of (in)definiteness. An additional claim is that the configuration of the lexical components, namely the underlying relative positions of the head noun, adjectives and adnominal genitives, cannot be reconstructed to the same degree of cogency. As with the clause, these constituents show such considerable variation that, except for intra-theoretical reasons, it is difficult to ascertain a basic order that might have been inherited by the language set.

After introducing the noun phrase as a syntactic concept, the line of argument begins with the three older languages of the set: Latin, Ancient Greek and Sanskrit, which display considerable word-order flexibility, as well as apparently discontinuous nominal constituents with intervening material. The chapter then turns to the later four languages, at least three of which exhibit a more fixed word order. A final section ties up the different threads and summarises the evidence for the proto-syntax in (1).

4.1.2 THEORETICAL BACKGROUND: THE SYNTAX OF NOUN PHRASES

The noun phrase has so far been treated simply as one among many components of the clause, its inner workings neglected. However, the noun phrase has been considered in various languages to be almost, if not equally, as complex as the clause. In English, it contains constituents that we can roughly divide into the lexical and the functional. The former group include the head noun (the ‘core’ of the domain), adjectives, genitives and dependent prepositional phrases. These are understood to be lexical, since they refer to

4.1. INTRODUCTIONS AND CHAPTER AIMS

entities and qualities in the world, and involve predicates. Functional elements include determiners, which help to specify the reference of a head noun, distinguishing one or a subset of the set which it denotes; determiners may be definite, such as definite articles (e.g. English *the*) and demonstratives (e.g. *this, that*), or they may be indefinite, such as indefinite articles (e.g. *a, some*). Determiners are herein understood to be distinct from quantifiers (e.g. *many, all*), at least syntactically, since the two can co-occur in one noun phrase.

- (2) *all* *the* *ten* *good* *houses of the neighbourhood*
 QUANTIFIER DETERMINER NUMBER ADJECTIVE NOUN GENITIVE

Moreover, modern English does not assemble these constituents randomly; determiners, adjectives and nouns are arranged into an inflexible DAN order.

- (3) a. *a good house*
 b. **good a house*
 c. *these good houses*
 d. **houses these good*

English today is morphologically quite impoverished, compared with its ancestral forms and other European languages, but cases like (3c) nonetheless show agreement of a number feature between the head noun *houses* and the determiner *these*. In Old English, a morphologically richer language, this agreement of features extends to the adjective too.

4.1. INTRODUCTIONS AND CHAPTER AIMS

- (4) a. *þæt* *gōde* *hūs*
the.N.NOM.SG good.WK.N.NOM.SG house.N.NOM.SG
'The good house'

(Old English. Artificial example)

- b. *þāra* *gōdra* *hūsa*
the.N.GEN.PL good.WK.N.GEN.PL house.N.GEN.PL
'Of the good houses'

(Old English. Artificial example)

Because of the ordering and the percolation of the head noun's features within the noun phrase, noun phrases have been considered in the generative traditional to be configurational, formed according to an abstract structure. The noun phrase mirrors the structure of the clause; each of these two domains of syntax has a lexical core (the finite verb or the head noun), which is then modified and 'fleshed out' by additional elements, such as nominal arguments, adjuncts, auxiliary verbs and determiners. Indeed, the verbal or nominal feature of that lexical core itself may not be basic, but rather added (cf. Borer, 2014). Evidence for this fundamental non-categorial status comes from noun phrases that match the argument structure and propositional content of finite clauses.

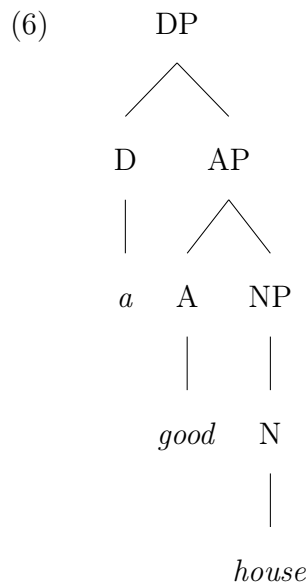
- (5) a. *the Romans destroy Carthage*
(Clause)
b. *the Romans' destruction of Carthage*
(Noun Phrase)

In (5), the Romans are in both examples the agent of the action of destruction, and

4.1. INTRODUCTIONS AND CHAPTER AIMS

Carthage its patient, but the patterns of agreement result from the categorial feature given to the action, that is, whether it be verbal or nominal. The clause and the noun phrase are treated in generative thinking as “extended projections” of the lexical core, the categories V and N, above which are located “functional shells” (Grimshaw, 2000) that offer positions for additional lexical items, such as adverbs and adjectives, and contribute grammatical features.

Following the influential analysis of Abney (1987), the syntax of the English noun phrase is widely considered to contain the high functional component D(eterminer), which may be lexically overt (e.g. *the*, *a*) or null. A noun phrase like (3a) therefore for Abney (1987, 208) has the following underlying structure:



Abney (1987), Szabolcsi (1987) and Horrocks and Stavrou (1987) draw explicit parallels

4.1. INTRODUCTIONS AND CHAPTER AIMS

between clausal and noun-phrase syntax. For Abney, D corresponds in function to the I head of the clause. For Szabolcsi (1987, 168), the Hungarian definite article *a(z)* realises the head “CN”, which is “the nominal counterpart of C”, while both the clausal and the noun-phrase syntax include a functional I(nflection) head (Szabolcsi, 1984, 1987). This is responsible for the morphological agreement of person and number between the subject and the verb in clauses, but also between a possessor and a possessum in noun phrases.

- (7) a. *én látom*
I.NOM.SG see.PRS.IND.DEF.1SG
'I see'
(Hungarian. Artificial example)
- b. *az én vendégem*
the I.NOM.SG guest.POSS.1SG
'My guest'
(Hungarian. From Szabolcsi (1984))

Because of parallel word-order phenomena in some languages, we can reasonably talk about a ‘left periphery’ for the noun phrase as well as the clause, and propose the existence of functional categories, like D, that produce it. The purpose of this chapter is to demonstrate that this analysis applies to the set of early Indo-European languages and to Proto-Indo-European. Chapter 5 discusses the parallels between the two reconstructed domains of syntax.

4.2 ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

4.2.1 THE LATIN NOUN PHRASE

The noun phrase in Latin is morphologically and syntactically typical of earlier Indo-European languages; hence, it is worth discussing in detail before moving on to the others. Its noun phrases contain various categories of word associated by the agreement of case, number and gender. Classical Latin has six cases, two numbers and three genders, which are morphologically realised as fusional endings on nouns, adjectives, lower numbers, determiners and quantifiers.

- (8) a. *perversam* *atque impiam* *religiōnem*
perverse.F.ACC.SG and impious.F.ACC.SG piety.F.ACC.SG
'Perverse and impious superstition'

(Latin. Cicero, *For Sulla* 25.70)

- b. *omnium* *f̄nitimārum* *c̄vitātum*
all.F.GEN.PL bordering.F.GEN.PL city.F.GEN.PL
'Of the neighbouring cities'

(Latin. Caesar, *Gallic War* 5.58)

Some other possible members of the noun phrase though do not show agreement with the head noun, namely modifying genitives and indeclinable cardinal numbers from four upwards.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

- (9) a. *post* [*hōram* *prīmam* *noctis*]
after hour.F.ACC.SG first.F.ACC.SG night.F.GEN.SG
'After the first hour of night'

(Latin. Cicero, *For Sextus Roscius* 19)

- b. *hās* *septem stēllās*
this.F.ACC.PL seven star.F.ACC.PL
'These seven stars'

(Latin. Varro, *On the Latin Language* 7.73)

Since agreement signals the relationship between a head noun and its different modifiers, it affords the Latin noun phrase the capacity for considerable word-order variation. This includes different orderings of constituents relative to the noun phrase, and instances of hyperbaton. The latter involves 'discontinuous' noun phrases, in which two or more elements agree in reference and therefore in features, but do not stand adjacent in the order. Hyperbata are common in poetry, in which morphological agreement keeps the grammatical information clear while metrical templates determine the word order.

- (10) a. [*tantae*] =ne [*animīs* *caelestibus*] [*irae*]
such.F.NOM.PL Q soul.M.DAT.PL heavenly.M.DAT.PL anger.F.NOM.PL
'Do heavenly souls have such angers?'

(Latin. Virgil, *Aeneid* 1.11)

- b. [*mūtam*] *nēquīquam adloquerer* [*cinerem*]
mute.F.ACC.SG in-vain speak-to.IPRF.SBJ.DEP.1SG ash.F.ACC.SG
'I address in vain speechless ash'

(Latin. Catullus 101)

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

However, we should not conclude that the Latin noun phrase is chaotic and randomly ordered; there are clear patterns in its composition.

One pattern concerns determiners. In its classical standard, Latin does not exhibit definite or indefinite articles, but it still displays a wide array of determiners. These elements can all be used pronominally or as part of a noun phrase (like English *that*), and include demonstrative and anaphoric elements (e.g. distal *ille* ‘that’, medial *iste* ‘that’, proximal *hic* ‘this’), primarily anaphoric elements (e.g. *is* ‘that’, *idem* ‘the same’, *ipse* ‘the very’), interrogative determiners (e.g. *quī* ‘which’) and indefinite determiners (e.g. *aliquis* ‘some’). Although their position with respect to the head noun is variable, a pre-nominal position before all other constituents is very common (with 94% of examples in the corpus of Iovino (2012)). Pinkster summarises the word order of these elements thus:

“Determiners most often precede the head noun. Interrogative and relative determiners almost always precede; demonstrative, anaphoric, and indefinite determiners show more variation.”

(Pinkster, 2021, 1067)

Demonstratives very frequently appear pre-nominally, in particular at the far left edge of the phrase. Devine and Stephens (2006, 511) report that *hic* is always pre-nominal in Cato and Caesar, though a post-nominal position is possible and observed in the Latin of Cicero.

- (11) a. ... *dē* [*eā* *rē*] *verba* *ūtī*
about that.F.ABL.SG matter.F.ABL.SG word.N.NOM.PL use.PRS.INF.DEP

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

f̄iant. [*huic* *operī*] *pretium* *ab*
become.PRS.INF.DEP this.N.DAT.SG work.N.DAT.SG price.N.NOM.SG from
dominō *bonō*...
master.M.ABL.SG good.M.ABL.SG
‘... Prayers on this matter should be made. For this work, from a good
homeowner...’

(Latin. Cato, *De Agriculture* 14)

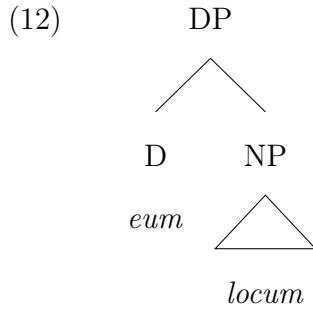
- b. *id* [*eā* *maximē* *rationē*]
that.N.ACC.SG that.F.ABL.SG greatest.F.ABL.SG reason.F.ABL.SG
fēcit, *quod nōluit* [*eum*
do.PRF.IND.ACT.3SG that not-want.PRF.IND.ACT.3SG that.M.ACC.SG
locum] *unde Helvētīū* *discesserant*
place.M.ACC.SG whence Helvetii.M.NOM.PL leave.PPRF.IND.ACT.3PL
vacāre
be-empty.PRS.INF.ACT
‘This he did with the main reason that he did not want the place which the
Helvetii had abandoned to be empty’

(Latin. Caesar, *Gallic War* 1.28)

Following Oniga (2014) and Giusti and Iovino (2016) *inter alios*, by identifying determiners as exponents of a D head within a left-headed DP², higher than the noun and its lexical modifier, we derive their usual initial position.

²We can treat determiners like *ille* and *hic* either as syntactically simplex heads or as complex phrasal elements in Spec,DP, following Leu (2008). Until diagnostics demonstrate otherwise, the former, simpler analysis is adopted. As for the use of the label ‘D’, an appropriate alternative is ‘Dem’, since Latin has demonstratives but no articles. However, for Latin and the other languages, the label D is maintained; it is intended to be more generic, characterised by (in)definiteness, and so able to include definite and indefinite determiners in its category, and to interact with the argument structure of the clause (see the conclusion in §4.5).

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK



However, this structure needs modifying if it is to generate different orders with determiners.

A key factor in the variation is “the semantics of the demonstrative” (Devine and Stephens, 2006, 514), by which they mean the choice between an ostensive or a discourse-anaphoric meaning. When ostensive, demonstratives like *hic* are assumed to be accompanied with a gesture, singling out an entity in the speaker’s environs; if used so, the demonstrative is pre-nominal. When anaphoric, they instead refer to something not necessarily present in the external circumstances but rather signal to something already given in the discourse; if so, they tend to be post-nominal.³

- (13) *quamdiū etiam [furor iste tuus] nōs*
 how-long still madness.M.NOM.SG that.M.NOM.SG your.M.NOM.SG we.ACC.PL
ēlūdet
 mock.FUT.IND.ACT.3SG
 ‘How long will this madness of yours still mock us?’
 (Latin. Cicero, *Catilinarian Orations* 1.1)

³This distinction is not consistent across determiners; the demonstrative *is* only has anaphoric and cataphoric functions, yet is rarely post-nominal.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

- (14) a. *egō interim [hanc āram]*
 I.NOM.SG meanwhile this.F.ACC.SG altar.F.ACC.SG
occupābō
 occupy.FUT.IND.ACT.1SG
 ‘Meanwhile I’ll occupy this altar’
 (Latin. Plautus, *Mostellaria* 5.1, from Pinkster (2021, 1068))
- b. *tange [āram hanc Veneris]*
 touch.PRS.IMP.ACT.2SG altar.F.ACC.SG this.F.ACC.SG Venus.F.GEN.SG
 ‘Touch this altar to Venus’
 (Latin. Plautus, *Rudens* 5.2, from Pinkster (2021, 1068))

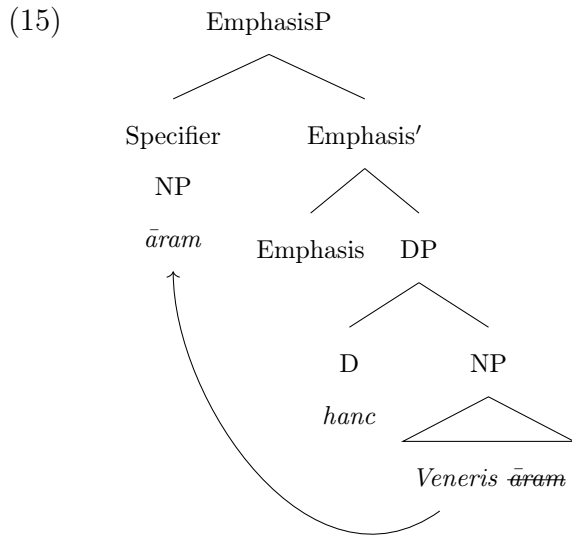
The difference between the altars in (14) is whether or not the altar is known to the audience and part of the ongoing conversation. In the first, the altar is an entity introduced into the discourse, and presumably Tranio, the speaker, gestures towards it. In the latter, the altar has already been mentioned and *hanc* is anaphoric, perhaps better translated as *the*. An initial or early pre-nominal position within the noun phrase seems therefore to add force to the demonstrative, linking it to current extra-linguistic circumstances, while post-nominal demonstratives are weaker, referring back to previous elements within the discourse.

Even when not initial, the earliness of determiners’ positions is consistent; only one constituent precedes the determiner, while others follow. This gives us grounds to analyse its underlying location as stable in the syntax, while subsequent operations produce its surface second position. One option is to bring in post-syntactic factors to derive orders like *furor iste tuus* in (13). Prosodic inversion (Halpern, 1992, 1995) could explain non-initial *iste*, if we first assume that it is enclitic. However, there does not seem to be any evidence from orthography or metre that points to a prosodic difference in non-initial

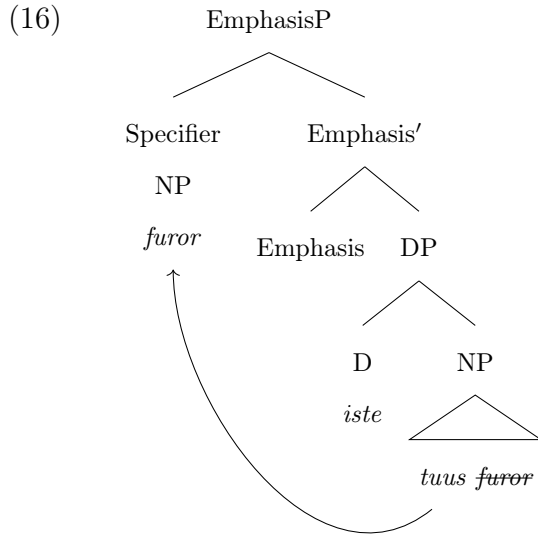
4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

determiners.

Instead, this thesis proposes that the Latin noun phrase involves the additional movement of constituents to an even higher syntactic location, namely ‘EmphasisP’. For instance, the determiner-second orders in (13) and (14b) derive from the movement of *āram* and *furor* higher than the DP of which *hanc* and *iste* are the head. This movement and resultant word-order variation produce the ostensive/anaphoric distinction in demonstratives.



4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK



This is understood to involve movement specifically to the specifier of Emphasis. So as not to contravene the Head Movement Constraint (Travis, 1984) and produce unattested word orders, the moved constituent (such as the nouns *āram* in (15) and *furor* in (16)) is considered to be underlyingly phrasal. Consequently, to achieve the movement of only the noun in (15) and in (16), and not its genitive or possessive modifiers, we must additionally posit that the latter constituents are underlyingly higher than the NP of which the noun is the head. It is this NP alone that moves to Spec,EmphasisP, leaving the modifiers in their basic positions. This suggests aspects of configurationality in the lower “NP” structure, but further comment is outside the left-peripheral scope of this thesis.

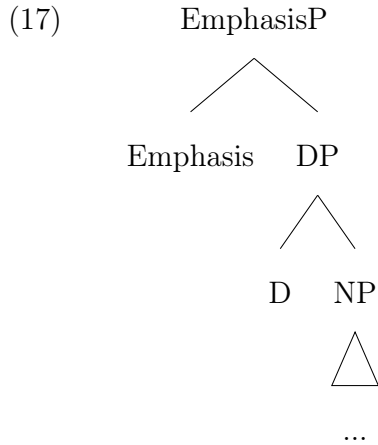
The key characteristic of the high functional head is not immediately clear; to call it ‘Topic’ or ‘Focus’, as with the clausal syntax, does not seem right, since these functions are

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

part of the discourse configuration of the clause as a whole. It is not apparent how a noun phrase could have a topic, about which the rest of the phrase provides information. A noun-phrase Focus, which provides new information and implies alternatives, seems more plausible (cf. Rooth, 1992).⁴ These focal functions are not identified though in the philological literature; in most cases, the interpretation given is that the fronted element is simply afforded a certain prominence, perhaps with emphatic prosody, which communicated its importance and encouraged the audience to relate it to or contrast it with other things in the ongoing discourse. In the case of phrase-second demonstratives like *ille*, the fronting of another constituent ‘backgrounded’ it, reducing its force and signalling its anaphoric function. Therefore, to summarise the qualities of the initial position in the philological literature, and to avoid committing to a too restrictive definition of this high syntactic component, this thesis opts for the label of ‘Emphasis’.

In sum, the left-peripheral behaviour of the Latin noun phrase arose from the following high syntax:

⁴Emphasised constituents in an English noun phrase imply alternatives, e.g. *an **English** linguist* (as opposed to a Scottish one) or *an English **linguist*** (as opposed to another kind of scholar).



We can detect the workings of Emphasis even in noun phrases without determiners. Adjectives for instance typically stand closer to the head noun than other constituents. They may be pre-nominal or post-nominal; both Adjective-Noun and Noun-Adjective orders are very well attested in Classical Latin. There is considerable disagreement over which might be the basic base order. Oniga (2014) argues that AN is the norm in archaic Latin, whereas Sleeman and Perridon (2011) instead identify NA as the more frequent order; following Pinkster, they state that Latin adjectives are usually post-nominal “unless pragmatic factors such as Focus cause them to be preposed” (Pinkster, 2021, 186). Adams (1976) agrees, saying that AN was “was reserved for adjectives which were emphatic or which carried a strong emotional content”, such as those that expressed a personal opinion, like *bonus* ‘good’. Other scholars though decline to offer a definitive view for either word order. For Allen and Greenough (1903) “the determining and most significant word comes first”, which could be a noun or its adjective. They admit at least adjectives of size or

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

quantity, like *multus* ‘much’ or *magnus* ‘great’, usually do precede, which Pinkster (2021, 1077) agrees with. Devine and Stephens (2006) likewise equivocate; in their opinion, there is no one factor that can explain the word order variants, because a mixture of discourse-information, semantics and prosody determines the order. This debate over the basic position of adjectives in Latin is substance for the claim in §4.1.1 that the configuration of lexical constituents in the reconstructed noun phrase is very difficult to diagnose, perhaps totally obscure. However, within the debate, a repeated theme is that the pre-nominal position had qualities of emphasis.

Likewise, there is similar great variation in the position of genitives with respect to the head noun; both post-nominal and pre-nominal positions are available to them, although there are preferences according to author, style and time. Noun-Genitive orders are rare in the oldest Latin sources, but if Genitive-Noun was the default at an early point, it had begun to give way by the time of Plautus (Magni, 2009). Devine and Stephens (2006) claim again that this equivalence of orders in Classical Latin has many contributing factors, including “conceptual individuation, the degree to which the genitive plus head combination expresses a single recognizable concept” (Devine and Stephens, 2006, 388). Yet again, along with this semantic pattern, Magni (2009) also mentions a pragmatic tendency of the Genitive-Noun order, namely that in this order the genitive is predictable information.⁵

Latin’s possessive pronouns, such as *meus* ‘my’ and *tuus* ‘your’, again appear pre- and post-nominally with near-equal frequency. Different writers may have their own pref-

⁵Magni (2009) also identifies a syntactic tendency, in that pre-nominal subjective genitives mirror the pre-verbal position of clausal subjects.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

- (19) a. *trēs* *nōbilissimī* *Haeduī*
three.M.NOM.PL most-noble.M.NOM.PL Haedui.M.NOM.PL
'Three most noble men of the Haedui'

(Latin. Caesar, *Gallic War* 7.67, from Devine and Stephens (2006, 494))
- b. *ex* [*tribūs* *ōrātōris* *officiīs*]
from three.N.ABL.PL orator.M.GEN.SG duty.N.ABL.PL
'From three duties of the orator'

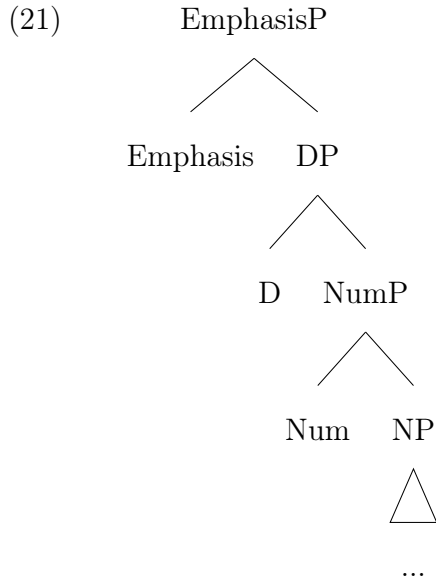
(Latin. Cicero, *Brutus* 197, from Devine and Stephens (2006, 494))

If a determiner element is present though, it typically precedes the number.

- (20) *hārum* *trium* *partium*
this.F.GEN.PL three.F.GEN.PL part.F.GEN.PL
'Of these three parts'

(Latin. Cicero, *Orator* 2.129, from Devine and Stephens (2006, 494))

Such behaviour of cardinal numbers in Latin can be derived from a NumP component to the syntax, in between DP and NP, as shown in (21).



Noun phrases with numbers also offer evidence for EmphasisP. A number may precede any determiners present if it is itself the more salient constituent, while in (22), the noun *templ̄is* precedes its number *tribus* in the dependent ablative noun phrase. This can be analysed as the fronting of *templ̄is* for contrast, since the temples are among several buildings that Cicero lists.

- (22) *Capitol̄ium* *illud* [*templ̄is* *tribus*]
 Capitol.N.NOM.SG that.N.NOM.SG temple.N.ABL.PL three.N.ABL.PL
inl̄ustr̄atum
 illustrate.PRFP.N.NOM.SG
 ‘That Capitol, adorned with three temples’

(Latin. Cicero, *For Aemilius Scaurus* 46, from Devine and Stephens (2006, 494))

A problem for the syntactic structure in (21) is the behaviour of quantifiers. *Omnis* ‘all’

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

and *tōtus* ‘all’ stand before the head noun “most often” (Pinkster, 2021, 1076). While pre-nominal and post-nominal *omnis* are roughly equal in the older Latin of Cato, *omnis* is pre-nominal in 80% of its uses in Cicero and more so in Caesar (Devine and Stephens, 2006, 509).

- (23) a. *omnēs* *hae* *trēs* *partēs*
 all.F.NOM.PL this.F.NOM.PL three.F.NOM.PL part.F.NOM.PL
pūrgātiōnis
 purification.F.GEN.SG
 ‘All these three parts of the purification’
 (Latin. Cicero, *Rhetoric for Herennius* 2.24)
- b. *omne* *id* *medium* *tempus*
 all.N.ACC.SG that.N.ACC.SG middle.N.ACC.SG time.N.ACC.SG
 ‘All the intervening time’
 (Latin. Gellius, *Noctes Atticae* 3.2.4, from Oniga (2014, 193))

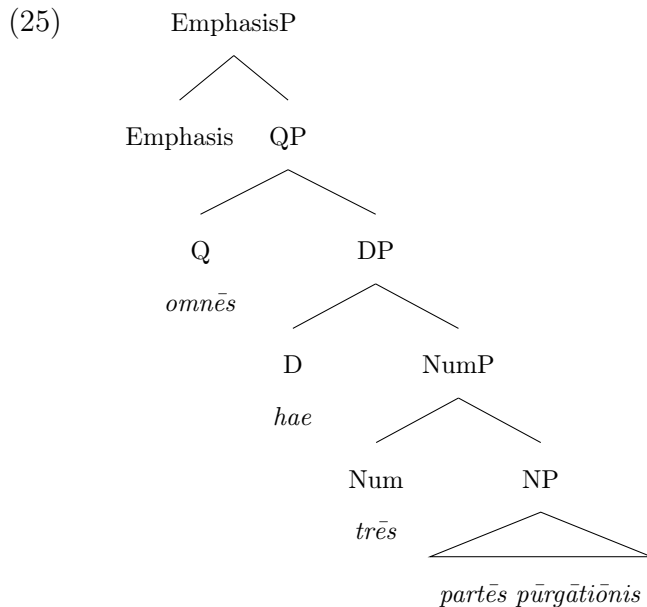
The quantifying adjective *multus* is typically pre-nominal too, although it also shows more variation as to whether it precedes or follows any determiners present in a particular noun phrase. The position of the quantifier is, once again, sensitive to its context. For example, if its head noun is contrastive, it precedes the quantifier, as in (24), in which *Gallia* is fronted to associate and contrast it with the Germans.

- (24) *sī* [*Gallia* *omnis*] *cum Germānīs*
 if Gaul.F.NOM.SG all.F.NOM.SG with German.M.ABL.PL
cōnsentīret
 conspire.IPRF.SBJ.ACT.3SG
 ‘If all Gaul conspired with the Germans’

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

(Latin. Caesar, *Gallic War* 5.29, from Devine and Stephens (2006, 510))

The key issue here is that quantifiers consistently precede determiners, as in (23). Two options for deriving this position present themselves. One is to add an additional QP into (21), between DP and EmphasisP, of which a quantifier like *omnis* is the head.



This configuration in (25) can not only derive the usual QDN order, but also leaves EmphasisP free to host fronted constituents, such as *Gallia* in (24).

However, other word-order evidence suggests that a QP is unnecessary, and that (21) is sufficient for orders with quantifiers too. There is the distinction of Devine and Stephens (2006, 517-9) between strong and weak quantifiers, illustrated in (26). Strong quantifiers, which are twice as common as weak ones in Cicero, appear before all other elements in the

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

noun phrase. Weak quantifiers appear after demonstratives, and occasionally even after DN combinations.

- (26) a. *praetermittō* [*omnem* *hanc* *ōrātiōnem*]
 omit.PRS.IND.ACT.1SG all.F.ACC.SG this.F.ACC.SG discourse.F.ACC.SG
 ‘I am omitting all this discourse’

(Latin. Cicero, *On the Agrarian Law* 1.7.21, from Devine and Stephens (2006, 517))

- b. *nōs* *autem* [*hanc* *omnem* *quaestiōnem*] ... *ā*
 we.NOM.PL but this.F.ACC.SG all.F.ACC.SG inquiry.F.ACC.SG by
nōbīs *explicātam* *esse* [*hīs* *litterīs*]
 us.ABL.PL explain.PRFP.ACC.SG be.PRS.INF this.F.ABL.PL letter.F.ABL.PL
arbitrāmur
 believe.PRS.IND.ACT.1PL
 ‘But we believe that this whole question ... has been explained by us in these words’

(Latin. Cicero, *On the Ends* 1.12, from Devine and Stephens (2006, 518))

The exact interpretational difference for Latin speakers between strong and weak quantifiers is not clear, but we can at least presume that there was a difference. If we assume that strong, pre-determiner quantifiers were somehow prosodically emphasised, we can locate them in EmphasisP. Post-determiner quantifiers, as in (26b), may instead reflect the basic position of quantifiers, lower in the noun-phrase structure. In this, they act like adjectives. Along the same lines, Devine and Stephens (2006, 518) propose a “SQP” structural position for strong quantifiers above all other projections, along with a “WQP” further down. Together the two capture the usual distribution of quantifiers. However, two QPs

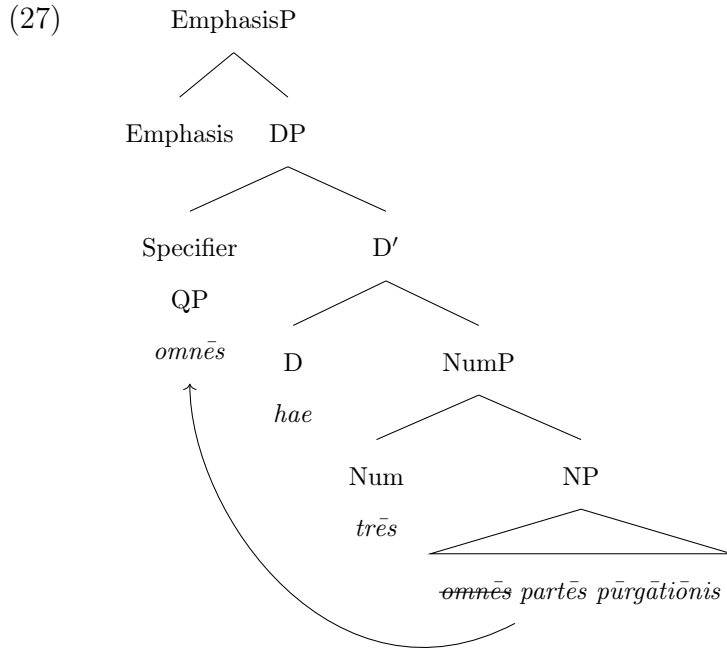
seem somewhat redundant. The same distribution can be explained by the movement of quantifiers out of an original adjectival position, and up to a higher position, which makes a quantifier strong. Naturally, this prompts the question of why quantifiers but not determiners are treated as fundamentally adjectival. The difference is supported by morphology; while quantifiers (e.g. *omnis*, *multus*) are first/second or third-declension adjectives, determiners belong to the so-called ‘pronominal’ declension (e.g. Sturtevant, 1913), which has several distinct endings not found with nouns and adjectives.⁶

In the syntax in (21), the target of moved quantifiers could be either Spec,DP or Spec,EmphasisP. If we locate in initial quantifiers in EmphasisP, we gain a motivation for their ‘strong’ interpretation, but this does force us to consider all pre-determiner quantifiers as somehow emphatic. To avoid this, we can move the quantifier to Spec,DP. This is a reasonable location; like determiners, quantifiers serve to specify the reference of a head noun and its lexical dependents.⁷

⁶This ‘functional’ paradigm includes the neuter nominative/accusative singular *-d* of *illud*, *istud* and *quid*, as well as the genitive singular ending *-ius*. There are exceptions to this morphological division, such as *sōlus* ‘only’ and *tōtus* ‘all’. In the nominal singular, these exhibit the *o/a*-stem declension typical of adjectives (i.e. *tōtus*, *tōta*, *tōtum*), but pronominal *-ius* in the genitive singular (*tōtius*). This looks like the “extension” of the pronominal ending in the classical language (Clackson, 2007, 15,17), perhaps to partly reassign these more functional words to a functional declension.

⁷In formal semantics, English determiners (such as *the*) and quantifiers (such as *every*) have even been analysed as the same thing, namely generalised quantifiers (cf. Montague, 1973).

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK



This analysis does predict that we should find examples with a XQD order, in which X is a fronted constituent, since the derivation in (27) leaves EmphasisP available. If such orders cannot be found in our copious Latin data, it indicates that the former analysis, with quantifiers and fronted constituents competing for the same position, is the stronger.

All in all, the structured syntax for the Latin noun phrase in (21) accounts well for the data in our Latin sources. To produce the acknowledged flexibility of its arrangement, the underlying syntax should include at least one component that can upset the basic order by triggering movement; this we can derive through Emphasis. (21) is a necessary and sufficient minimum; there is evidence for further syntactic ingredients, such as a QP,

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

but this behaviour can also economically derive through EmphasisP. (21) is deliberately ambiguous about the location of lexical constituents (adjectives and genitives) relative to the noun, since these exhibit greater variability. The evidence does at least suggest that numbers are generated from a NumP, below DP but above NP.

A final issue is how (21) works with instances of non-contiguous noun phrases. Ultimately, hyperbata are not problematic; there are various explanations at our disposal for this behaviour. One is to propose that such words do not at any stage of derivation constitute a single syntactic unit.

- (28) a. *partēs* *mihi* *Caesar* *hās*
 part.F.ACC.PL I.DAT.SG Caesar.M.NOM.SG this.F.ACC.PL
imposuit
 place-on.PRF.IND.ACT.3SG
 ‘Caesar has imposed this role on me’

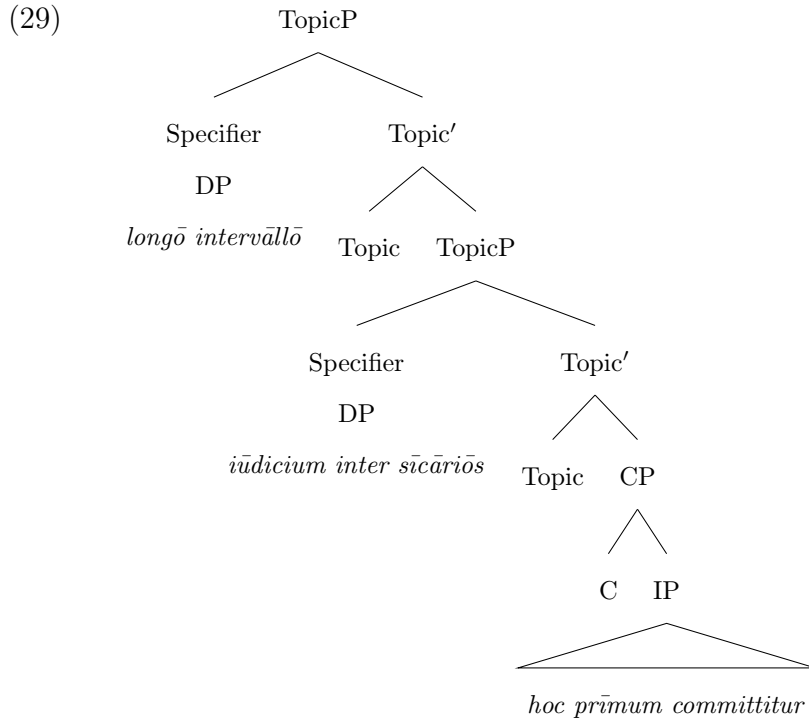
(Latin. Cicero, *Letters to Atticus* 10.10.2, from Devine and Stephens (2006, 531)

- b. *longō* *intervāllō* *iūdicium* *inter*
 long.N.ABL.SG interval.N.ABL.SG trial.N.NOM.SG between
sīcāriōs *hoc* *prīmum* *committitur*
 murderer.M.ACC.PL this.N.NOM.SG first.N.NOM.SG join.PRS.IND.PASS.3SG
 ‘After a long interval, this trial is the first to be convened on a charge of murder’

(Latin. Cicero, *For Sextus Roscius* 11)

Partēs, *longō intervāllō* and *iūdicium* are topical phrases in the syntax proposed in Chapter 3, while *hās* and *hoc* are resumptive pronouns.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK



Sometimes, one of the words may be a tail added on to the clause, like a clarifying afterthought, while its agreeing modifier belongs to the clause proper.

- (30) *id,* *ut dixī,* *exīlem* *amat*
 it.NOM.SG as say.PRF.IND.ACT.1SG meagre.F.ACC.SG love.PRS.IND.ACT.3SG
terram
 earth.F.ACC.SG
 ‘This, as I have said, likes it lean, the earth’

(Latin. Columella, *On Agriculture* 2.10.3, from Devine and Stephens (2006, 423))

At the syntactic level, topics, resumptive pronouns and tails can be explained as multiple and separate noun phrases, which may comprise only one word. Alternatively, some non-

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

continuous orders may arise from true syntactic phrases that are affected by the movement of a subset of their constituents. This again might be done to give that subset a particular interpretation, such as focus in the preverbal position.

- (31) *atque eī* *puerōs* *ostendērunt* *multōs*
and he.DAT.SG boy.M.ACC.PL show.PRF.IND.ACT.3PL many.M.ACC.PL
'And they showed him many boys'

(Latin. Cicero, *De Inventione* 2.2, from Devine and Stephens (2006, 532))

For (31) (and perhaps also (30)), we can view the final quantifier as maintaining a basic, unmarked position for the phrasal object *puerōs multōs*, the first component of which has undergone movement to low Focus to ensure that the noun *puerōs* alone gains the focal interpretation.

All this is to say that we can utilise a variety of analyses (namely movement out of (21) and multiple co-referential noun phrases) to account for Latin's frequent use of discontinuous noun phrases, without losing the structured syntax in (21). Having admitted that some agreeing nominal constituents of the clause are best explained as words in only apposition, an alternative view to adopting the given analysis would be to claim that there is no noun-phrase syntax at all. In other words, all nouns, adjectives, determiners, etc. are fundamentally independent of each other. However, such a non-configurational view risks throwing the baby out with the bath-water. To produce the relative orders that we consistently observe in Latin, it is better to adopt a structured syntax that is also capable of self-modification. This issue of identifying noun-phrase syntax from considerable word-

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

order variation pertains also to Australian languages, some of which (such as Warrongo) consistently display non-continuous nominal constituents and variable relative word order. Within Australian languages, Louagie (2023) proposes three groups, according to the degree to which their nouns and adnominal modifiers exhibit clear “unithood”. For the middle group:

“... the evidence is somewhat less straightforward, but which also seem to have NP construal as a major option: these are languages with fixed or restricted flexible order but any type of case marking (alternating phrasal and word marking, just word marking, or no marking). As a consequence, these languages probably have a wider range of construals available than the languages of the first group. An example is Guugu Yimidhirr, which has restricted flexible order: elements with a determining function (like a personal or possessive pronoun) can be placed at either edge, while adjectives (incl. numerals) have a fixed position following the head noun...”

(Louagie, 2023, 199)

From Louagie’s definitions, Latin clearly belongs to this middle group; its noun phrases are more often continuous than not, and the noun phrase is not chaotic in its composition. It does not follow that discontinuity is incompatible with continuity and phrasehood; “the presence of discontinuous expressions in a language does not necessarily provide evidence that other, contiguous expressions per definition could not form a syntactic unit ... discontinuous expressions can really be seen as a separate construction type” (Louagie, 2023,

200).

4.2.2 THE VEDIC SANSKRIT NOUN PHRASE

Much of what can be said for Latin pertains to the attested orders of the Vedic Sanskrit noun phrase too. Through its rich nominal morphology, countless orders of the members of a noun phrase are possible and attested. An agreeing adjective can stand before, after and even at a distance from a head noun, if one is present.

- (32) *[bṛhadúḁṣo* *Marúto* *viśvávedasaḥ]* *prá*
 greatly-raining.M.NOM.PL Marut.M.NOM.PL all-possessing.M.NOM.PL PV
vepayanti *párvatān* *[ádābhyāḥ]*
 shake.PRS.IND.ACT.3PL mountain.M.ACC.PL invulnerable.M.NOM.PL
 ‘The Maruts, greatly raining and possessing all wealth, invulnerable, they shake
 the mountains’

(Sanskrit. RV 3.26.4, from Schäufele (1991a, 58))

Such discontinuous noun phrases, in as much as they *are* phrases, are very common. As with Latin, the possible reasons for discontinuity are numerous: working to fit metrical patterns; discourse information, such as topicalisation or clausal tails; and even a syntactophonological reason, namely the movement of enclitic WL1 pronouns out of their noun phrase and to C, as in (33).

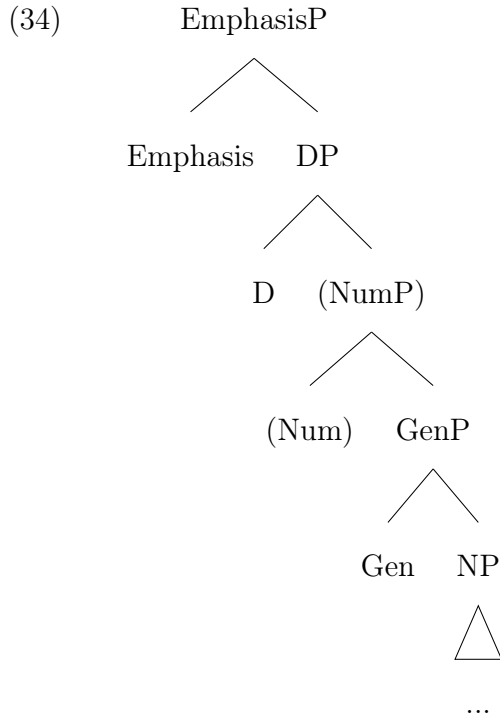
- (33) *kadá =naḥ* *śuśravad* *gíra* *Índro* *aṅgá*
 when we.GEN.PL hear.PRF.SBJ.ACT.3SG prayer.F.ACC.PL Indra.M.NOM.SG PRT

‘When might Indra have truly heard our prayers?’

(Sanskrit. RV 1.84.8, from Macdonell (1916, 354))

The apparent flexibility of noun-phrase elements is so great that we might, following Reinöhl (2016), be pessimistic about the constituency of multi-word strings. In her view, the syntax of nouns, adjectives, determiners and other similar elements is so “unconstrained” that all such categories “stand in relations of apposition, no element being obligatory or clearly super- or subordinate”. In other words, we should consider nouns and adjectives to be syntactically independent of each other, and there is no nominal unit larger than the inflected word. Contrary to this, this thesis posits that the Sanskrit noun phrase is better understood as configurational; Sanskrit suits a similar syntactic structure to that proposed for Latin in (21), with at least a DP layer that generates the basic position of determiners. (34) illustrates this configuration, with brackets indicating components for which less solid evidence has at present been found.⁸

⁸The inherent difficulty of working on the Sanskrit noun phrase is compounded by an apparent lack of previous work on its word order in general, in any European language, not just English. Morphology, agreement patterns and compounds dominate the discussion of Sanskrit syntax.



There is moreover limited evidence for a projection higher than the DP in (34). It can scramble the basic order that (34) generates, akin to EmphasisP in (21).

First, it is necessary to respond to Reinöhl's view of the Sanskrit noun phrase as being essentially non-configurational. This view is extreme and has some clear drawbacks; for one, it is theoretically very costly. It would analyse the four-word subject in (32) not as one argument of the verb and one joint recipient of its nominative case, but rather as four separate subjects and arguments. For cases of fronted topical phrases like (35), we would need to double the operations of movement.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

- (35) [*divyā āpo*] *abhī yād =enam āyan*
 heavenly.F.NOM.PL water.F.NOM.PL PV when he.ACC.SG go.IPRF.IND.ACT.3PL
 ‘When the heavenly waters came upon him’
 (Sanskrit. RV 7.103.2, from Lowe (2014, 2))

Most notably, it would ignore the clear examples of syntactic phrasehood in which enclitic elements respect the boundaries of a constituent phrase. In (36), we find the Wackernagelian ‘second-position’ elements *hī* and *enam* positioned later than after the first word.⁹

- (36) a. [*mahé kṣatrāya śāvase*] =*hī*
 great.N.DAT.SG dominion.N.DAT.SG power.N.DAT.SG for
jajñé
 be-born.PRF.IND.MED.2SG
 ‘For to great dominion and power were you born’
 (Sanskrit. RV 7.28.3, from Lowe (2014, 41))
- b. [*amṛtatvām rākṣamāṇāsa*] =*enam devā*
 immortality.N.ACC.SG protectPRSP.M.NOM.PL he.NOM.SG god.M.NOM.PL
Agnīṃ dhārayan draviṇodām
 Agni.M.ACC.SG maintain.IPRF.IND.ACT.3PL wealth-giver.M.ACC.SG
 ‘Protecting their immortality, the gods maintained him as Agni, the giver of
 wealth’
 (Sanskrit. RV 1.96.6, from Hale (2018, 1935))

If all parts of the initial phrases in (36) were appositional and syntactically independent, we would expect *hī* and *enam* to interrupt the phrase, as such elements sometimes do, rather than after their phrasal boundary. The order in (36b) is derived in §2 through the

⁹However we define ‘word’, e.g. prosodically or orthographically.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

movement of the WL1 pronoun to C, lower than the syntactic position of the initial phrase *amṛtatvám rākṣamāṇāsa*. It is because of examples like (36) that Lowe (2014, 44) states that “there is clear evidence for the existence of the phrasal categor[y] NP”.

Furthermore, as with Latin again, if Sanskrit lacks a nominal domain larger than the individual word, this predicts total word-order freedom, with no tendencies as to the position of nouns and adjectives. Yet there are such tendencies; nouns, adjectives and other categories do cluster together when they share a common reference, and in the view of Speyer (1886, 10), the typical position of attributive adjectives is immediately pre-nominal. A post-nominal position can be interpreted predicatively, as if they are two parts of a clause with an implicit copula. Delbrück (1893, 94-103) agrees, but adds that “*spezialisierende*” adjectives, which signify and emphasise some special status of the noun, can also be immediately post-nominal. In such cases, noun and adjective stand “in apposition”, according to Macdonell (1916, 285). All of this is to say that there is good support from nouns and adjectives for the idea of a distinct structured noun phrase in Vedic Sanskrit, although it is undeniably flexible.

Turning now to the components of (34), to support the NumP, the most common position of cardinal numbers for appears to be immediately pre-nominal and after any determiners, but this is only an impression; a dedicated review on this matter has not yet been found, so NumP is included in parentheses in (34).

- (37) a. [*imāni* *trīṇi* *viṣṭápā*]
 this.N.ACC.PL three.N.ACC.PL surface.N.ACC.PL
 ‘These three surfaces’

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

(Sanskrit. RV 8.91.5, from Wenthe (2012, 27))

- b. [**dáśa** áśvān] [**dáśa** kósān] [**dáśa**
ten.ACC.PL horse.M.ACC.PL ten.ACC.PL chest.M.ACC.PL ten.ACC.PL
vástrā ádhibhojanā] [**dáśa** u hiraṇyapiṇḍān]
robe.N.ACC.PL addition.N.ACC.PL ten.ACC.PL PRT gold-nugget.M.ACC.PL
'Ten horses, ten chests, ten robes as a bonus, ten lumps of gold too'

(Sanskrit. RV 6.47.23)

- c. [**ṣáḥ**] astabhñā [**viṣṭīrah**] [**pāñca**
six.ACC.PL set.IPRF.IND.ACT.2SG space.F.ACC.PL five.ACC.PL
saṃdṛśāḥ]
direction.F.ACC.PL
'Six spaces and five directions you have fixed'

(Sanskrit. RV 2.13.10)

As (37c) shows, numerals can also form part of discontinuous phrases. In all obvious respects, they therefore do not seem substantially different from other adnominal categories. One difference with other adnominal modifiers is that the cardinal number can be the head of the phrase, making the counted noun a partitive genitive instead (Petrocchi, 2022, 290). What is not clear is their position with respect to such categories; whether they typically stand before any adjectives present is something only a future dedicated survey can demonstrate.

Second, to support the DP in (34), we should review the behaviour of determiners in Sanskrit. Like Latin, Sanskrit lacks articles, but has a set of determiners, used for either ostentation or discourse anaphora, either pronominally or as part of a larger phrase. These include proximal *idám* and *etád*, medial *adás* and distal *tád*, all of which decline, although

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

features of their declensions differ from the declensions of adjectives. Doubled determiners are also attested (Schäufele, 1991a, 39), as in (38c).

- (38) a. [***ābhīh*** *śámābhir*] *maháyanta*
this.F.INS.PL work.F.INS.PL exalt.PRSP.M.NOM.PL
'Exalting (you) with these sacrifices'
(Sanskrit. RV 4.17.18)
- b. [***té*** *devā*] [***etád*** *yájur*]
that.M.NOM.PL god.M.NOM.PL this.N.ACC.SG worship.N.ACC.SG
apaśyan
see.IPRF.IND.ACT.3PL
'Those gods saw this worship'
(Sanskrit. *Taittiriya Samhita* 5.2.3.1, from Schäufele (1991a, 102))
- c. [***etát*** ***tyát*** =*ta* *indriyám*]
this.N.NOM.SG that.N.NOM.SG you.GEN.SG greatness.N.NOM.SG
aceti
notice.AOR.IND.PASS.3SG
'This greatness of yours has been noticed'
(Sanskrit. RV 6.27.4, from Wenthe (2012, 26))

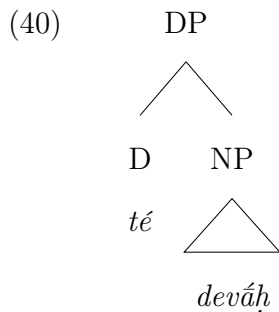
Ascertaining the basic word order of such determiners with respect to the rest of the noun phrase is challenging. In addition to the fronting of genitive pronouns to the left periphery, determiners like *sá-* and *syá-* sometimes appear fronted too and removed from their associated nouns and adjectives (Wenthe, 2012, 25). Nonetheless, the general pattern is that determiners are typically initial or second within the noun phrase; they “must be to the left of all the other members of the NPs they specify” (Schäufele, 1991a, 10), whether part of a discontinuous phrase or not. Wenthe (2012, 25-9) agrees with this generalisation,

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

and on this basis, proposes two “DP” projections in the underlying syntax, one or both of which may have overt material. Wenthe’s DP₁ and DP₂ are both higher than all other constituents of the nominal syntax:

- (39) [DP₁ [DP₂ [GenP [AP [NP]]]]]]
 (Wenthe, 2012, 28)

A noun phrase like the first in (38b) can therefore be straightforwardly derived through the following DP structure:



Wenthe furthermore identifies the movement of a noun or adjective within this structure into one of his DP layers “as a sign of emphasis” (Wenthe, 2012, 25); this movement can derive the minority of post-nominal determiners, and supports an additional EmphasisP in the syntax.

The syntax in (34) can be supported by Wenthe’s observation and derivation of the behaviour of adnominal genitives. Accented genitives can precede (as in (41a)) or follow their head noun (41b).

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

- (41) a. *syāma* *íd* [***Índrasya*** *śármaṇi*]
 be.PRS.OPT.3PL indeed Indra.M.GEN.SG protection.N.LOC.SG
 ‘May we live under Indra’s protection’
 (Sanskrit. RV 1.4.6)
- b. [*hárī* ***Índrasya***] *ní cikāya*
 horse.M.ACC.DU Indra.M.GEN.SG PV perceive.PRF.IND.ACT.3SG
 káh = *svit*
 who.MF.NOM.SG Q
 ‘Who then has perceived the two horses of Indra?’
 (Sanskrit. RV. 10.114.9, from Lowe (2015, 45))

Nonetheless, a pre-nominal position is acknowledged as usual; for a dependent genitive noun (phrase) “its position is the same as that of a simple adjective” (Delbrück, 1893, 102) and it “regularly stands in front” (Delbrück, 1888, 66). Macdonell (1916, 285) states that “the substantive belonging to a genitive is placed before the latter only when it is emphatic”, which offers tentative evidence for an EmphasisP in the noun-phrase syntax.

Non-accented genitives moreover offer useful diagnostics for (34). It seems that enclitic adnominal genitives¹⁰ have two syntactic options; they can move to C, producing a clausal second position, as in (33), or they can remain within their noun phrase. In the latter case, they display a variable position relative to the other constituents; they stand either second (42) or first (43).

- (42) a. *ápo* *revatīḥ* *śṛṇutá* [*hávam*]
 water.N.VOC.PL rich.N.VOC.PL hear.PRS.IMP.ACT.2PL call.M.ACC.SG

¹⁰Or rather, genitive and dative. The two cases are frequently identical in the form of pronouns and both express possession.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

=*me*]
 I.GEN.SG
 ‘Rich waters, hear my call!’

(Sanskrit. RV 10.30.8, from Wenthe (2012, 169))

- b. *ási* [*priyó* =*no* *átithih*]
 be.PRS.IND.2SG dear.M.NOM.SG we.GEN.PL guest.M.NOM.SG
 ‘You are our dear guest’

(Sanskrit. RV 6.2.7, from Wenthe (2012, 169))

- (43) a. *áver* *Indra* *prá* [=*no*
 further.PRS.OPT.ACT.2SG. Indra.M.VOC.SG PV we.GEN.PL
dhíyah]
 thought.F.ACC.PL
 ‘Oh Indra, further our thoughts’

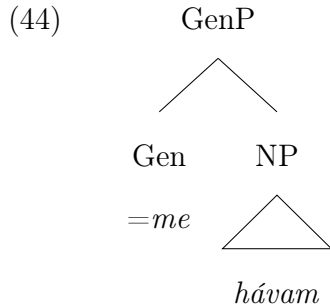
(Sanskrit. RV 8.21.12, from Wenthe (2012, 170))

- b. *Agním* *vardhantu* [=*no* *gírah*]
 Agni.ACC.SG increase.PRS.IMP.ACT.3PL we.GEN.PL hymn.NOM.PL
 ‘May our hymns increase Agni’

(Sanskrit. RV 3.10.6)

In (43a), the possessor *naḥ* is enclitic on and retroflexed by the preceding preverb *prá*.

Hale’s (2009) observation is that the pronoun-first order is “disproportionately common immediately following verbs” (Wenthe, 2012, 173), as in (43b). Adopting Wenthe’s analysis, we can derive this variation through the combined syntacto-prosodic approach of this thesis. The variation in position is best captured if the genitive clitic is located in a higher position in the noun phrase than the noun and thus ‘within touching distance’ of the preceding word in the linear order.



Yet, being enclitic, the pronoun requires a host; it is the prosodic organisation of the overall clause that determines whether the pronoun will be enclitic on the preceding word (as in (43)) or trigger prosodic inversion (Halpern, 1992, 1995) with the next word in the noun phrase. The latter produces the pronoun-second orders in (42).

Naturally, this prompts the question of why some possessor pronouns do not trigger prosodic inversion and are content with the preceding verb as their host, and vice versa, since *nah* is not enclitic on the verb *śṛṇava* in (46a). Hale’s explanation is that such instances of prosodic inversion indicate the presence of “a prosodic boundary (at least a weak prosodic break)” (Wenthe, 2012, 174). In other words, a slight pause between the preceding word and the noun phrase stops any enclitic pronouns from forming a prosodic unit across the edge of the syntactic phrase. Verbs and their objects might normally form a prosodic phrase as well as a syntactic one, but a prosodic break leaves inversion within the noun phrase as the only option for providing the pronoun with a host.¹¹

¹¹Why such a break may occur is a further issue; it may be because the noun phrase is emphasised, or functions as supplementary material to the clause proper, or to fit a metric template that includes a caesura.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

Notably, however, enclitic genitives follow all determiners, even double determiners, as in (38c). Examples of noun phrases that unambiguously contain a determiner and are enclitic-first have not been found; some enclitics that immediately precede the determiner of their phrase can be explained alternatively as instances of pronominal movement to C and the clausal left periphery, as in (45), in which the genitive clitic stands in the second position of the overall clause, and so can be treated as enclitic on the imperative verb and outside its associated noun phrase.

- (45) *śṛṇutám* =*ma* [*imám* *hávam*]
hear.PRS.IMP.ACT.2DU I.GEN.SG this.M.ACC.SG call.M.ACC.SG
‘Hear this my call’

(Sanskrit. RV 8.73.10, from Wenthe (2012, 181))

This enclitic-second pattern with determiners even occurs in discontinuous noun phrases, such as in (46a), in which the head noun *bráhmāṇi* appears separated from its associated demonstrative and possessor.

- (46) a. *úpa* [*bráhmāṇi*] *śṛṇava* [*imá* =*naḥ*]
here prayer.N.ACC.PL hear.PRS.SBJ.ACT.2SG this.N.ACC.PL we.GEN.PL
‘May you hear these prayers of ours’

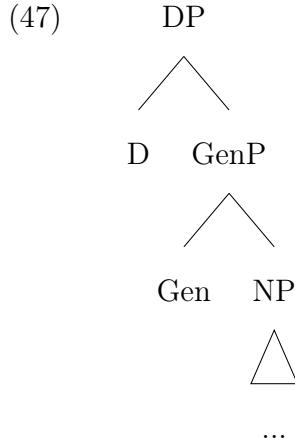
(Sanskrit. RV 6.40.4, from Wenthe (2012, 171))

- b. [*etád* =*asyā* *ánaḥ*] *śaye*
this.N.NOM.SG this.F.GEN.SG chariot.N.NOM.SG lie.PRS.IND.MED.3SG
[*súsampiṣṭam*] *vípāśy* *á*
wrecked.N.NOM.SG Vipas.F.LOC.SG on
‘This chariot of hers lies wrecked on the Vipas’

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

(Sanskrit. RV 4.30.11)

Admittedly, word orders like *etád asyā ánaḥ* in (46b) are syntactically ambiguous; it can be analysed as a single continuous phrase [*etád =asyā ánaḥ*], or as [*etád ánaḥ*] with the pronoun *asyā* intervening through prosodic inversion around *etád*, or even as completely discontinuous, with both *etád* and *asyā* having moved higher in the syntax and away from the head noun *ánaḥ*. However, (46a) at least follows the rule that enclitic genitives do not precede determiners. Those genitive pronouns that remain within the noun phrase therefore seem to have no choice but to use a determiner as their host; the presence of a determiner blocks the external clisis seen in determiner-less noun phrases like (43a). Thus, on the basis of both accented and unaccented genitives, we may tentatively conclude that the basic position of genitives is higher than NP, but lower than that of DP.



Having located determiners in DP, it clear that one more projection in the syntax is theoretically necessary. For one example, again like Latin, Sanskrit quantifiers (e.g. *sárvaḥ* and

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

vīśvah, both meaning ‘all, every, whole’) can occur in continuous noun phrases or stand removed from what they quantify. Quantifiers can be fronted to the left periphery, which Ram-Prasad (2022, 99-100) analyses as movement to a FocusP. Either way, they most frequently precede their quantified head noun, if one is present, as in (48a).

- (48) a. [*sárvam* *parikrośám*] *jahi*
all.M.ACC.SG scorner.M.ACC.SG kill.PRS.IMP.ACT.2SG
‘Kill every scorner’
(Sanskrit. RV 1.29.7)
- b. [*sárvam*] *sá* [*pūtám*] *aśnāti*
all.M.ACC.SG he.NOM.SG purify.PRF.P.M.ACC.SG eat.PRS.IND.ACT.3SG
‘He eats all food purified’
(Sanskrit. RV 9.67.31, from Ram-Prasad (2022, 89))

Because of all the usual word-order possibilities, this thesis agrees with Wenthe (2012, 25) that “the position of quantifiers is unclear”. They do at least seem to precede consistently the lexical constituents of the noun phrase, namely nouns, adjectives and genitives, competing with determiners for the initial position.

- (49) a. *átha* [*idám* *vīśvam*] *Pavamāna*
so this.N.NOM.SG all.N.NOM.SG Pavamana.M.VOC.SG
‘So all this, oh Pavamana...’
(Sanskrit. RV 9.86.28, from Wenthe (2012, 73))
- b. [*sárvam* *tád*] *Indra*
all.N.NOM.SG that.N.NOM.SG Indra.M.VOC.SG
‘All that, oh Indra...’

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

(Sanskrit. RV 8.93.4, from Wenthe (2012, 73))

It may be that the pre- and post-determiner positions had different interpretations for speakers; such a difference is something only ascertainable now through a thorough quantitative investigation. At the juncture, it does not seem possible to propose the precise position for quantifiers within (34), but we can claim that they, like determiners, belong to the higher structure. Since this thesis has identified determiners like *tád* as D, then we can locate quantifiers either in Spec,DP, in a QP of their own, or in EmphasisP. Whichever we choose, it should be higher than D. Sanskrit exhibits the same morphological split identified for Latin, namely that quantifiers exhibit adjectival declensions, while determiners have an equivalent ‘pronominal’ declension with some quirks across the paradigm (e.g. the *-d* of neuter nom./acc. *tád*). The movement analysis of Latin in §4.2.1 may therefore derive Sanskrit quantifiers too, but this must remain speculative until a review of quantifiers’ distribution provides the necessary facts.

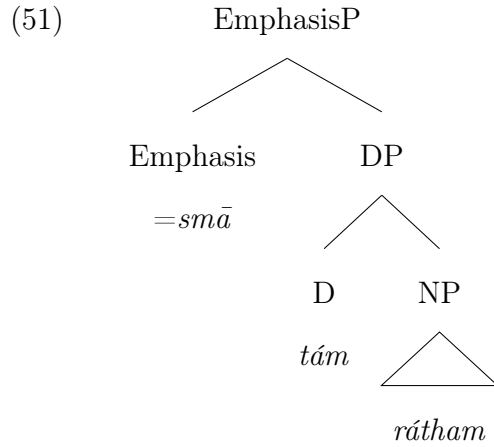
In favour of an EmphasisP component specifically for the syntax are the various remarks already given that the initial position is used for emphasis, and that an emphatic noun occupies it in a minority of orders. Furthermore, a highly speculative source of evidence comes from emphatic particles; these include *u*, *gha*, *cid*, *sma* and *evá* (Giannakis, 2021, 419). Various structural analyses offer themselves for these particles; some cases are most simply analysed as adverbs within the clause proper. However, in other cases, we can consider them to be lexical exponents of the Emphasis head of a noun phrase. Often these particles emphasise only a pronoun, but in noun phrases, they consistently stand after the

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

first word.

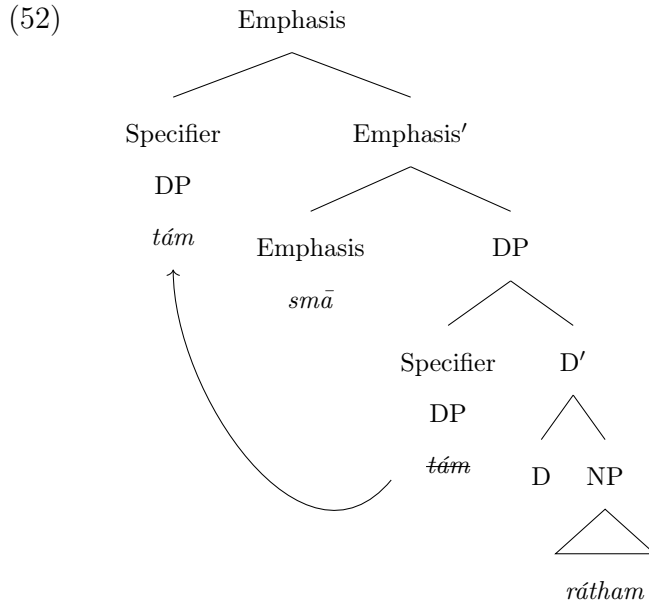
- (50) a. [*tám smā rátham*] ... *prá ava*
that.M.ACC.SG PRT chariot.M.ACC.SG PV support.PRS.IMP.ACT.2SG
'Follow that very chariot'
(Sanskrit. RV 1.102.3)
- b. [*asmākaṃ smā rátham*] *á tiṣṭha*
we.GEN.PL PRT chariot.M.ACC.SG PV mount.PRS.IMP.ACT.2SG
'Mount our chariot'
(Sanskrit. RV 1.102.5)
- c. [*samānāṃ cid rátham*] *ātasthivāṃsā*
common.M.ACC.SG PRT chariot.M.ACC.SG mount.PRFP.ACT.M.NOM.DU
'Two mounted upon one chariot'
(Sanskrit. RV 2.12.8)

Having first assumed that these particles belong underlyingly to the noun phrase that they appear within, two derivational operations can explain their phrase-second position. One is to posit that the particles are the highest head in the syntax, but, being enclitic, trigger prosodic inversion with the first viable host.



Alternatively, we can derive their position through the movement of one of the constituents into EmphasisP. This could be to the Emphasis head itself or to Spec,EmphasisP, if we want to obey the Head Movement Constraint (Travis, 1984), although the latter requires that we treat constituents of the Sanskrit noun-phrase as underlyingly phrasal; in (52), the moved determiner *tām* is base-generated as a phrasal DP in the specifier position of the main DP.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK



A strength of the derivation in (51) is that it accounts for why noun phrases with these emphatic particles otherwise display a typical word order, with determiners, numbers and genitives preceding nouns, as they do in (50) and (37b). However, its weakness is that we must accept that the preceding word is not especially emphatic; *tām*, *asmā́kam* and *samā́nām* in (50) are simply the first accented word in the linear order. If we adopt this analysis then, we may consequently claim that the Emphasis head in these cases emphasises the whole noun phrase, rather than one constituent of it. If we wish to maintain the latter view, then the derivation through constituent movement in (52) is better. The principal benefit though of all this discussion is that those emphatic particles considered to be noun-phrase-internal offer lexical evidence for the high Emphasis head in (34).

In sum, the Vedic Sanskrit noun phrase is something of a syntactic jungle; it lacks

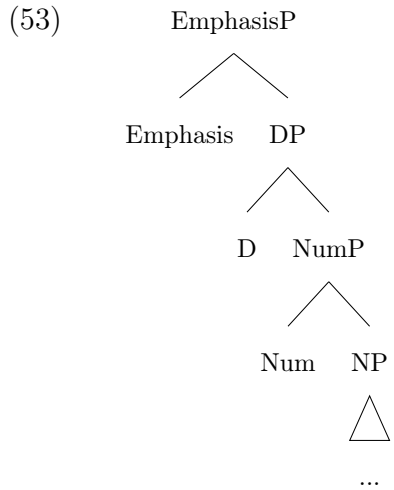
4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

stable elements, such as articles, that could serve as an anchor for the word order of a noun phrase, and demarcate its edges within the larger clause. However, as with Latin, the response to this extreme word-order variation is not to be overly pessimistic and deny the possibility of identifying anything of the underlying syntax of the domain, or even its existence outright. The little that we have to go off, namely word-order tendencies and the position of enclitics, indicates configurationality. If the Sanskrit nominal domain were totally non-configurational, it would look different; we would not expect adjectives, genitives, determiners and quantifiers to precede their head nouns with the frequency that we observe.

What is absolutely needed are dedicated quantitative surveys of the relative positions of nouns and adnominal modifiers in the language of the Vedas; in the absence of such a survey, we can only go off impressions and apparent tendencies. On this basis, the syntax in (34), with its GenP, NumP, DP and EmphasisP, can generate the patterns noted. Overall, the Sanskrit noun phrase has a higher basic position for functional elements like determiners and quantifiers, and a lower structure for lexical constituents like adjectives, genitives and the head noun. What this configuration produces can be modified by subsequent operations, namely: fronting for emphasis, prosodic inversion to satisfy the requirements of enclisis, movement of enclitic pronouns out of the noun phrase to C, as well as the apposition of multiple noun phrases. This conspiracy of factors is theoretically able to derive the majority of the attested data, and it is still a better approach than the alternative, that is, Reinöhl's denial that there is syntax beyond the individual word, since it can account for the patterns that we do find.

4.2.3 THE ANCIENT GREEK NOUN PHRASE

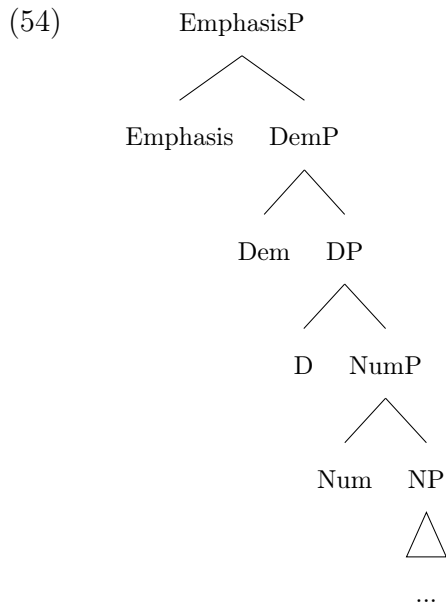
The Ancient Greek noun phrase displays distinct syntactic patterns. A necessary first step though is to draw a line between the earlier Homeric language and the later Classical Greek. While both contribute to the reconstruction, the development of a definite article is responsible for certain differences in their syntax. In Mycenaean Greek, there is no trace of definite articles; in Homeric, there are signs of the early stages of development out of a particular definite determiner; and in Classical Greek, the definite article *ho* is noticeably common. In general, this development has consequences for the surface word order of the noun phrase, which displays less flexibility in the classical language; the definite article seems to have an anchoring effect on the overall domain. (53) shows the proposed syntax of the earlier Homeric noun phrase.



The exact configuration of the lower structure, comprising the head noun and lexical mod-

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

ifiers, is again argued to be unknowable, on the basis of their word-order variation. (54) meanwhile shows the later syntax, in which the original DP layer has split into two: DP and DemP.



It should be noted that, like Latin and Sanskrit, Ancient Greek clearly displays both continuous and discontinuous phrases (cf. Devine and Stephens, 2000); as with Latin and Sanskrit, hyperbata are best derived through a combination of movement and apposition, rather than refuting phrasal syntax outright.

Firstly, to support the NumP in (53) and (54), we must review adnominal numbers. Cardinal and ordinal numbers do not differ much from other adnominal modifiers of the noun; their only major point of distinction is cardinal numbers' lack of agreement from five (*pénte*) and up. In terms of syntax, their usual position is before the noun; the Herodotean

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

data of Bakker (2009, 34-5) are that 64% of numerals precede the noun in indefinite noun phrases, and 87% in definite ones. As well as the classical language, Homeric numbers likewise seem stable in their pre-nominal, but post-determiner position.¹²

- (55) a. *atâr [tâs pénte néas kuanoprōireíous]...*
 but that.F.ACC.PL five ship.F.ACC.PL dark-prowed.F.ACC.PL
 ‘But those five dark-prowed ships...’
 (Homeric Greek. *Odyssey* 3.299)
- b. [*hoî*] = *dé* = *moi* [*heptâ kasígnētoi*] *ésan*
 that.M.NOM.PL and I.DAT.SG seven brother.M.NOM.PL be.IPRF.IND.3PL
en megároisin
 in hall.N.DAT.PL
 ‘But those seven brothers of mine were in the halls’
 (Homeric Greek. *Iliad* 6.421)

However, it is possible for the noun to precede if it is prominent and contrastive; Bakker (2009, 59) gives the following example to illustrate this.

- (56) *haûtai = *dè* [treîs muriádes kai triékonta*
 this.F.NOM.PL and three.F.NOM.PL hundred-thousand.F.NOM.PL and thirty
orguiéōn] *gínontai,* [*stádioi = *dè**
 fathom.F.GEN.PL happen.PRS.IND.MED.3PL stade.M.NOM.PL and
triēkósioi kai triskhílioi]
 three-hundred.M.NOM.PL and three-hundred.M.NOM.PL
 ‘This makes three hundred and thirty thousand fathoms, or three thousand three
 hundred stades’

¹²A review taken personally of the cardinal numbers *pénte* ‘five’ and *heptá* ‘seven’ in both Homeric epics, excluding discontinuous phrases and compounds, found that the two numbers precede their head noun in all thirteen instances.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

(Classical Greek. Herodotus, *Histories* 4.86, from Bakker (2009, 59))

In (56), the two numeral phrases, *treîs muriádes kai triékonta* and *triēkósioi kai triskhílioi*, respectively precede and follow the nouns that they quantify. This difference is “perfectly explicable” by pragmatic principles, since the context has led the audience to anticipate the first measurement, which is then converted into and contrasted with the second unit of measurement, *stádioi* (Bakker, 2009, 59). If the noun phrase contains a definite article, the numeral will appear in between the article and the head noun. In his study of Plato’s Republic, Frederick (2009, 59) found that the cardinal number *dúo* ‘two’ usually appears in a Determiner-Number-Noun order, sometimes DNNum, but never preceding the article in a *NumDN order. Numbers should substantiate not only NumP in (53) and (54), but also EmphasisP and DP.

Turning now to the main evidence for that DP, the definite article should be properly reviewed. On the basis of both textual and comparative evidence, we can see that the Ancient Greek definite article *ho* develops out of a demonstrative (Beekes, 2009, 1041). In Homeric Greek, *ho* can be either a pronoun or a determiner, and it serves anaphoric, contrasting and (re-)topicalising functions; it can highlight and re-introduce a referent that is assumed to be known but is absent from the preceding discourse, and help the audience to track and distinguish between entities.

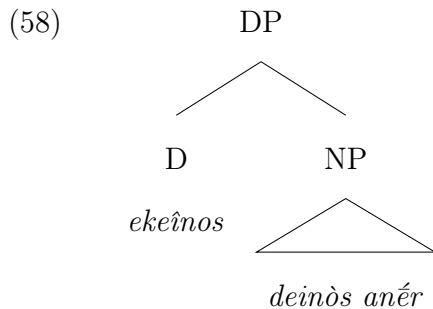
- (57) *hoúneka* [tòn *Khrúsēn*] *ētímasen*
 wherefore that.M.ACC.SG Chryses.M.ACC.SG dishonour.AOR.IND.ACT.3SG
arētêra *Atreídēs.* *hò* =gàr
 priest.M.ACC.SG Atreus-son.M.NOM.SG that.M.NOM.SG for

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

êlthe [*thoàs*] *epì* [*nêas* *Akhaiôn*]...
 go.AOR.IND.ACT.3SG quick.F.ACC.PL on ship.F.ACC.PL Achaean.M.GEN.PL
 ‘Because the son of Atreus had dishonoured the priest Chryses;¹³ he had come to
 the swift ships of the Achaeans...’
 (Homeric Greek. *Iliad* 1.11-12)

Note how in (57), *ho* appears as a determiner and a pronoun for Chryses¹³, preceding the proper name, but is not used with the known Achaean ships.

On the basis of the Homeric and Classical syntactic evidence, we can posit the DP layer of (53) for definite noun phrases.



By the time of Classical Greek, *ho* has parted ways with other demonstratives and gained the frequency and functions of an article, “used virtually obligatorily with common/count nouns to convey pragmatic definiteness” (Manolessou and Horrocks, 2007, 229). It is not used with vocative noun phrases though, despite the inherent definiteness implied by their addressee (Sansone, 1993, 205). Compared with Latin and Sanskrit, this element gives

¹³NB: with the exception of its masculine/feminine nominative singular/plural forms, *ho* begins with *t-*, as with the masculine accusative singular *tón* here.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

the Ancient Greek noun phrase a distinct rigidity. The most common orders are DN, DA and DAN, with the definite article preceding the head noun and lexical modifiers; certain word orders with *ho* are not attested and presumably were therefore not permitted by the grammar.

- (59) a. *tò* *krátos* [*tês* *gês*
the.N.NOM.SG power.N.NOM.SG the.F.GEN.SG land.F.GEN.SG
hekásēs]
each.F.GEN.SG
‘The possession of each bit of land’

(Classical Greek. Thucydides, *History* 4.98)

- b. [*têi* *prôtēi* *hēmérai*] ... *kai* [*têi*
the.F.DAT.SG first.F.DAT.SG day.F.DAT.SG and the.F.DAT.SG
deutérai] ... *kai* [*têi* *trítēi*] ...
second.F.DAT.SG and the.F.DAT.SG third.F.DAT.SG
‘On the first day ... and on the second ... and on the third ...’

(Classical Greek. Thucydides, *History* 3.96, from Smyth (1920, §1157))

From the abundant evidence of the definite article, we can propose that Classical Greek maintained the DP-syntax in (58), with the definite article as a D head positioned above the head noun and any lexical modifiers. That D head must be lexically null though in indefinite noun phrases (see §4.5 for discussion of indefinite NP), and also in phrases with a proper noun, such as a personal name. With these the definite article is still optional, and its appearance is “popular... emphatic, or anaphoric” (Manolessou and Horrocks, 2007, 229).¹⁴

¹⁴We may even understand the (somewhat) complementary distribution between the definite article and

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

There are further relevant patterns and restrictions on the definite noun phrase. A *DNA order, with the adjective last, does not seem permitted. Instead, an adjective may follow its noun in definite noun phrases, but this requires a second article, resulting in DNDA.

- (60) a. [*hoi ónoi hoi ágrioi*]
 the.M.NOM.PL ass.M.NOM.PL the.M.NOM.PL wild.M.NOM.PL
 ‘The wild asses’

(Classical Greek. Xenophon, *Cyropaedia* 1.4.7, from Bakker (2009, 12))

- b. [*tèn arkhèn tèn hēmetéran*]
 the.F.ACC.SG rule.F.ACC.SG the.F.ACC.SG our.F.ACC.SG
 ‘Our rule’

(Classical Greek. Thucydides, *History* 6.20)

This ‘determiner spreading’ (cf. Alexiadou, 2014) is obligatory for post-nominal adjectives in Classical Greek. Curiously, an order with determiner spreading and the reverse positions of adjective and noun, *DADN, is not attested. Moreover, it is not necessary for genitive and prepositional modifiers; genitives can appear after the head noun both with (61a) or without (61b) a doubled definite article. In (61a), this is the masculine nominative singular *ho*, agreeing with the head noun *patér*.

- (61) a. *ho patèr ho [toû*
 the.M.NOM.SG father.M.NOM.SG the.M.NOM.SG the.M.GEN.SG
Polemárkhou]
 Polemarchus.M.GEN.SG

proper names as evidence for the ‘N-movement’ of the inherently definite name up to the D head, following Longobardi (1994).

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

‘The father of Polemarchus’

(Classical Greek. Plato, *Republic* 1.328b)

- b. *tò* *plêthos* [*tôn* *pheugóntōn*]
 the.N.NOM.SG multitude.N.NOM.SG the.M.GEN.PL flee.PRSP.ACT.M.GEN.PL
 ‘The multitude of the fleeing’

(Classical Greek. Thucydides, *History* 4.96)

- c. *hoi* *alloi* *hoi* [*perì tèn*
 the.M.NOM.PL other.M.NOM.PL the.M.NOM.PL around the.F.ACC.SG
límnen̄n]
 lake.F.ACC.SG
 ‘The others around the lake’

(Classical Greek. Thucydides, *History* 4.93)

We can derive instances of determiner spreading through apposition of two noun phrases. This presumably had a distinct interpretation; DNDA and DNDG orders like (60a) and (61a) may have worked like a clarification, specifying the reference of the noun, akin to English ‘*the asses, the wild ones*’ and ‘*the father, the one of Polemarchus*’. This does not explain though the apparent ban on DADN, and in turn raises the question of why determiner spreading is obligatory for adjectives, but not for other post-nominal modifiers. Two details point towards a syntactic explanation for the latter issue.

One is that adjectives inflect and agree with the head noun, while other adnominal modifiers do not. The other is that the absence of *DNA and *ADN orders means that adjectives always stand after the definite article and before a head noun. On the basis of these details, we can suggest that the Ancient Greek adjective is fixed in a post-articular and pre-nominal position that is responsible for agreement features, akin to “AGR” in

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

earlier generative literature (cf. Punske, 2014). An AGRP beneath DP and above NP would further give a position for the aforementioned pre-nominal genitives, including those that function as subjects. This is only brief speculation, but it could be confirmed with further statistical research, specifically into the position of numerals, some of which decline to agree.

Turning now to demonstratives, these also offer support for the DP and EmphasisP above it, and additionally an innovative functional projection (DemP) by the time of Classical Greek. In Homeric language, the future definite article *ho*, along with *hóde* ‘this’, *hoûtos* ‘this’ and *ekeînos* ‘that’, is still a demonstrative. These elements have the possible functions both of ostension towards objects in the world and of anaphoric (and cataphoric) reference within the discourse; they can be used pronominally; and none of them are obligatory in any definite contexts. In a review for this thesis of the distal demonstrative *ekeînos* in the *Iliad* and the *Odyssey*, sixteen examples of continuous noun phrases were found with pre-nominal (and pre-adjectival) *ekeînos*, compared with six examples of post-nominal *ekeînos*. Of the latter six, five are the same phrase, *émati keínōi* ‘on that day’, which therefore seems like a set construction. Across the two Homeric epics, the evidence therefore indicates that these demonstratives, including *ho*, typically precede nouns and their modifiers.

- (62) a. *hoîos* [*ekeînos* *deinòs* *anēr*]
what-sort.M.NOM.SG that.M.NOM.SG terrible.M.NOM.SG man.M.NOM.SG
‘Of what sort that fearful man is’
(Homeric Greek. *Iliad* 11.653-4)

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

- b. *hóte* [*ekēinos* *hupérthumos* *Diòs*
 when that.M.NOM.SG high-spirited.M.NOM.SG Zeus.M.GEN.SG
huiòs] *épleen*
 son.M.NOM.SG sail.IPRF.IND.ACT.3SG
 ‘When that high-spirited son of Zeus sailed’

(Homeric Greek. *Iliad* 14.250-1)

Homeric demonstratives can vary in their position; they occasionally follow a head noun in a NDem order, as in the aforementioned phrase *émati keínōi* or the noun phrase in (63) (if it is in fact underlyingly a single noun phrase).

- (63) *líssōm’* [*anéra* *toûton* *atásthalon*
 beg.PRS.SBJ.MED.1SG man.M.ACC.SG this.M.ACC.SG wicked.M.ACC.SG
obrimoergón]
 mighty-working.M.ACC.SG
 ‘I may beg that wicked man, that worker of violence’

(Homeric Greek. *Iliad* 22.418)

The elements that remain demonstratives after Homeric times (*ekēinos* ‘that’, *hóde* ‘this’ and *hoûtos* ‘this’) occur in the classical language with the definite article, clearly indicating its development into a new marker of definiteness. Classical demonstratives stand in what is traditionally called the ‘predicative’ position: either before the definite article, noun and any adjectives (DemDN), or after them all (DNDem). If the latter, the demonstrative does not require determiner spreading.

- (64) a. *taútēn* *tēn* *pólin*
 this.F.ACC.SG the.F.ACC.SG city.F.ACC.SG

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

‘This city’

(Classical Greek. Herodotus, *Histories* 5.49.7)

- b. *tèn* *timèn* ***taútēn***
the.F.ACC.SG honour.F.ACC.SG this.F.ACC.SG
‘This honour’

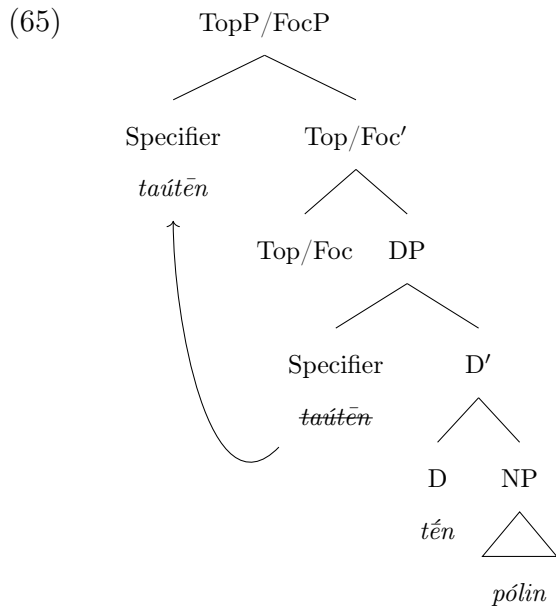
(Classical Greek. Thucydides, *History* 2.36.1)

Devine and Stephens (2000) state that the typical position for demonstratives is initial, and argue, as does Bakker (2009, 77), that their word order is motivated by emphasis and pragmatic markedness; the pre-nominal, pre-articular position of demonstratives is used for ostentation, while a post-nominal position is used when it is the noun that is emphasised, or when the demonstrative has a less emphatic, discourse-anaphoric function for a known individual (Guardiano, 2009). For example, in (63), the speaker is King Priam and the man referred to is Achilles; if *toûton* here modifies the noun *anéra*, then it suits a non-ostensive, discourse-given interpretation, since we know that Priam is behind the Dardanian Gate, with Achilles out of view on the other side.

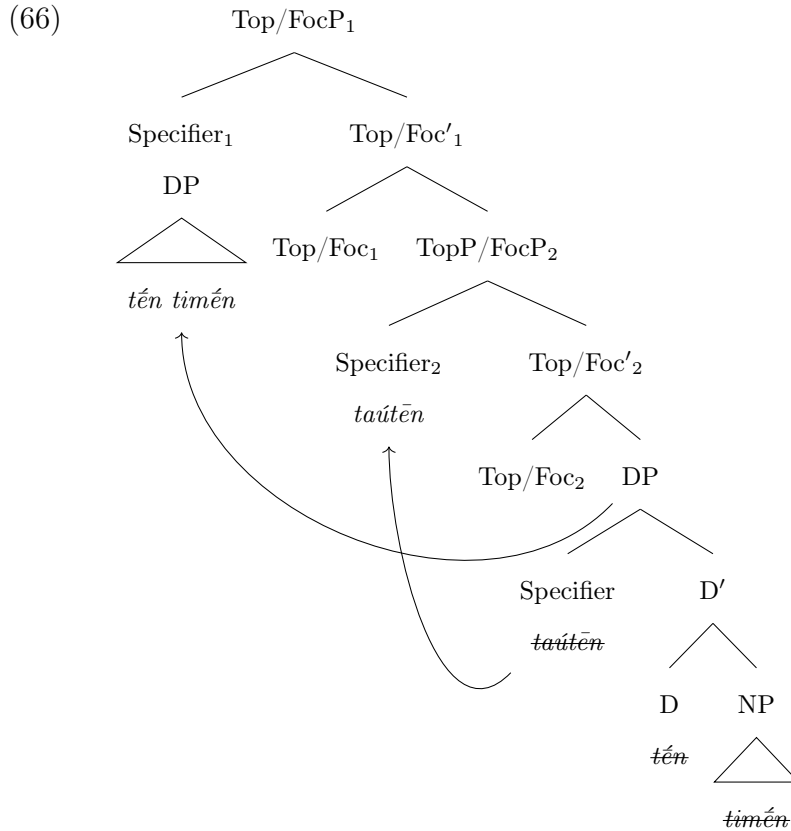
As for Latin in §4.2.1, the alternation between Dem(D)N and (D)NDEM is cogently derived again through the movement of certain constituents up to a high EmphasisP. One precedent analysis is that of Kirk (2007, 117-20), in which the alternation is likewise produced through movement. This analysis involves an articulated syntactic structure with a DP, TopicP and FocusP. More accurately, the syntax contains two functional projections (“FP”), which seem to be optionally either a TopicP or a FocusP. Kirk’s derivation is that the article and the demonstrative are both first-merged in DP, as its head and specifier

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

respectively. This is reasonable, since both are lexical items associated with definiteness. From here, they can move up to one of the two possible FPs. For Kirk, a DemDN order like (64a) therefore has the following syntax:



Meanwhile, a DNDem order like (64b) would involve the movement of both the demonstrative and the remnant of the DP into the two FPs.

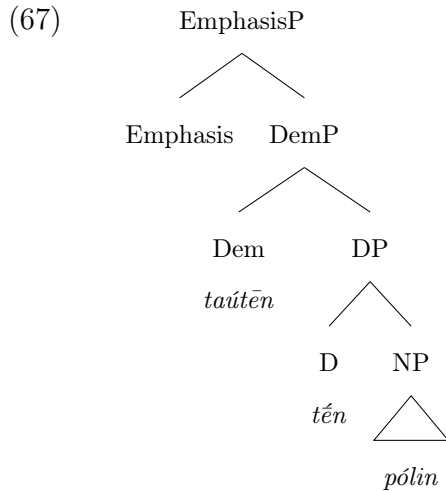


Kirk's articulated syntax is successful in producing the attested orders of DemDN, DNDem and others with modifiers. However, a problem is that it commits the constituents involved to certain discourse-informational features, for which the evidence is limited. This difficulty is perhaps responsible for the explicit optionality of the characteristic of the FPs (either TopicP or FocusP) in (65) and (66). A topical quality for the article and noun in DNDem orders like (64b) is plausible, since post-nominal demonstratives like *taútēn* are understood to be given, and so the preceding *tēn timēn* may be taken to be given information. Yet

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

it is not so apparent why an initial, pre-nominal demonstrative is (on its own) topical. Topicality is really a quality of noun phrases as a whole, rather than of a subset of their constituents.

This thesis builds on, but also reduces Kirk’s syntax, through the classical syntax given in (54). In DemDN orders, we can locate the noun in NP, the definite article as D, and the demonstrative as the head of its own projection, DemP.¹⁵

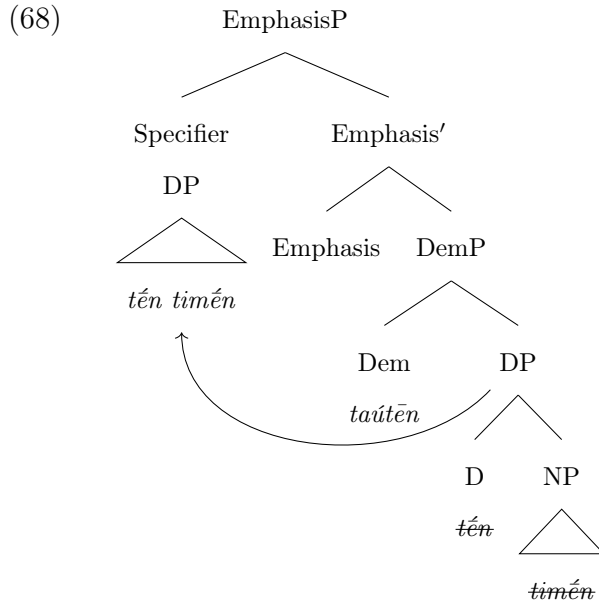


The demonstrative is identified as the head of Dem, not in the Spec,DP, for intra-theoretical reasons; it permits us then to move the DP as a phrase to Spec,EmphasisP, which produces the alternative DNDem order.¹⁶

¹⁵The label and category ‘DemP’ follow Frederick (2009), but ‘DP1’ or neutral ‘FP’ are alternatives, depending on how we chiefly characterise this projection.

¹⁶If we assume that proper names are DPs, then we can derive Name-Demonstrative orders like *Theódōros hóde* ‘this Theodoros’ (Plato, *Theaetetus* 147, from Frederick (cf. 2009, 16-7)) in the same way.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK



This syntactic derivation, moving the DP to Spec,EmphasisP, produces the attested order of DNDem and the interpretation proposed for it, without committing the noun-phrase constituents to a topical or focal quality. The fronting of the DP had the consequence of ‘backgrounding’ the demonstrative, which weakened it and distinguished it in function from initial (ostensive) demonstratives. Unlike in Kirk’s syntax, the demonstrative remains in its basic position in (68).

The issue of quantifiers now raises its head. These (such as *pâs* ‘all’) typically also occupy the predicative position, appearing in QDN (69a) and DNQ orders (69b).

- (69) a. *pantòs* *toû* *stratoû* *tôn*
 all.M.GEN.SG the.M.GEN.SG army.M.GEN.SG the.M.GEN.PL
Athēnaiōn
 Athenian.M.GEN.PL

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

‘Of all the army of the Athenians’

(Classical Greek. Thucydides, *History* 4.96)

- b. *tèn* *emèn* *skeuèn* ***pâsan***
the.F.ACC.SG my.F.ACC.SG apparel.F.ACC.SG all.F.ACC.SG
‘All my apparel’

(Classical Greek. Herodotus, *Histories* 7.15.3, from Bakker (2009, 120))

Like demonstratives, we can derive the predicative position of quantifiers by locating them higher than the DP, meaning that they do not move with it to Spec,EmphasisP in cases of DNQ orders like (69b). This prompts the question of whether quantifiers can be located in the same functional projection in (67), or whether the syntax needs an extra component to accommodate them when demonstratives and quantifiers co-occur in QDemDN orders. Following the analysis for Latin in §4.2.1, an expansion of the syntax does not seem necessary; quantifiers like *pâs* display word-order variation and morphology that allows us to posit that they begin life as adjectives, but in some cases undergo movement up to Spec,DemP. Bakker (2009, 250-2) notes that quantifiers can also occupy the ‘attributive’ position, like a lexical adjective, and the traditional view has been that this has a different interpretation.

- (70) *ho* ***pâs*** *khronos*
the.M.NOM.SG whole.M.NOM.SG time.M.NOM.SG
‘The whole time’

(Classical Greek. Plato, *Apology* 40e, from Frederick (2009, 7))

Smyth (1920, §1174) states that *pâs* in the attributive position in (70) “denotes the whole

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

regarded as the sum of all its parts”, while “in the predicate (and usual) position *pâs* means ‘all’ ... [being] often emphatic”. Likewise, Bakker (2009, 251) describes attributive *pâs* in (70) as “an adjunct of state”, whereas predicative *pâs* in (69) “clarifies which referent is referred to”. Again, an argument from morphology can support a different derivations for quantifiers and demonstratives; as with Latin, the Ancient Greek definite article *ho* and the demonstratives built from it (*hóde*, *hoûtos* ‘this’) have an idiosyncratic declension, differing from adjectives in the ending-less neuter nom./acc. singular (e.g. *tó*, *toûtou*), while quantifiers like *pâs* belong to the second/third declension. In sum, we can locate predicative quantifiers after movement in Spec, DemP, the same projection as demonstratives; the primary purpose of DemP is the ‘reference specification’ for the overall noun phrase.

Additional evidence for EmphasisP in (54) comes from the fronting of other subsets of the noun phrase. For one, amidst the word-order variation of adjectives, genitives and the head noun, a common thread is the association between an initial position and emphasis. Beginning with adjectives, both Adjective-Noun and Noun-Adjective orders are frequent, although Homeric and Platonic Greek at least give the impression that AN is the more common.

- (71) a. [*poluphloísboio* *thalássēs*] ... [*arguréoio* *bioîo*]
 loud-roaring.F.GEN.SG sea.F.GEN.SG silver.M.GEN.SG bow.M.GEN.SG
 ‘Of the loud-roaring sea ... of the silver bow’
 (Homeric Greek. *Iliad* 1.34, 49)
- b. [*nêa* *mélainan*] *erússomen* *eis* [*hála*]
 ship.F.ACC.SG black.F.ACC.SG drag.AOR.SBJ.ACT.1PL into sea.F.ACC.SG

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

dîan]

heavenly.F.ACC.SG

‘We should drag a black ship into the heavenly sea’

(Homeric Greek. *Iliad* 1.141)

- c. [***mounophthalmous anthrôpous***] *kai* [***khrusophûtolakas***
one-eyed.M.ACC.PL man.M.ACC.PL and gold-guarding.M.ACC.PL
grûpas]
griffon.M.ACC.PL
‘One-eyed men and gold-guarding griffons’

(Classical Greek. Herodotus, *Histories* 4.27, from Bakker (2009, 50))

- d. [*basilêos* ***amûmonos***] ... [*gaîa* ***mélaina***]
kind.M.NOM.SG blameless.M.NOM.SG earth.F.NOM.SG black.F.NOM.SG
‘A good king ... the black earth’

(Classical Greek. Plato, *Republic* 2.363)

Different factors have been proposed to motivate the variation. For example, for Brunel (1964, 16), the difference is semantic, specifically the degree of ‘unity’ between the noun and the adjective; in an AN order, the adjective is an inseparable part of the noun’s reference; in a post-nominal NA order, their relationship is looser, more coincidental. However, Dik (1995) disagrees; semantics plays no role in variant orders, only discourse context. It is when an adjective is contrastive or otherwise salient that it is preposed. In the language of Herodotus, Bakker (2009) likewise states that, in noun phrases with one noun and one modifying element, it is discourse-informational importance that principally determines word order, preposing the element that is more important or relevant to the discourse. For instance, Bakker’s analysis of (71c) is that the adjectives *mounophthalmous* and *khrusophûlakas* precede their nouns simply because they provide “the most prominent

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

information” in their phrases. Bakker (2009, 62) does however also acknowledge the role of prosody, syntax and rhetorical and poetic techniques, such as when the adjective is complex (“heavy”), with multiple dependent components, in which case the noun will precede the complex adjective, regardless of pragmatic markedness.¹⁷

The behaviour is similar with genitive modifiers. In general, genitives appear close to their head noun, either immediately after it or before it. According to Viti (2008), Genitive-Noun and Noun-Genitive orders are equally common and Greek shows the “highest rate of flexibility” among early Indo-European languages. Bakker’s findings from Herodotus (2009, 34-5) agree with this, reporting 55% pre-nominal and 45% post-nominal in indefinite noun phrases, as well as 58% to 42% in definite ones. In Viti’s study of the Greek of Homer and Herodotus, both semantic and discourse-informational reasons appear to be competing motivations for genitive placement. Her view is that a pre-nominal position is more likely when the genitive is singular and animate; generic groups and inanimate entities are less likely to be preposed. Furthermore, a GN order also often occurs when the genitive is new information, while a post-nominal genitive “refers to the topic of the immediately preceding clauses” (Viti, 2008, 1).

- (72) a. *ekpérsai* [*Priámoio* *pólin*]
 destroy.AOR.INF.ACT Priam.M.GEN.SG city.F.ACC.SG
 ‘... that you sack the city of Priam’

¹⁷A difference to note is between indefinite and definite noun phrases in the presence of the later definite article. In indefinite noun phrases without it, NA and AN orders almost match each other in frequency; in the Classical Greek of Herodotus, Bakker (2009, 34) reports percentages of 42% AN and 58% NA in noun phrases with one noun and one adjective. However, with the definite article, the adjective does not enjoy the same freedom; Bakker (2009, 35) reports a majority of AN orders of 79% in Herodotean Greek.

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

(Homeric Greek. *Iliad* 1.19)

- b. [*purai* ***nekúōn*** *kaíonto* [*thameiaí*]
 pyre.F.NOM.PL dead.M.GEN.PL burn.IPRF.IND.MED.3PL crowded.F.NOM.PL
 ... [*kēla* ***theoío***]
 arrow.N.NOM.PL god.M.GEN.SG
 ‘The crowded pyres of the dead burned ... the arrows of the god’

(Homeric Greek. *Iliad* 1.52-3)

- c. *kai dè kai* [***gunaikòs*** *eídōlon* *khruíseon*
 and PRT and woman.F.GEN.SG statue.N.ACC.SG golden.N.ACC.SG
trípēkhu]
 three-cubits.N.ACC.SG
 ‘And also a golden statue of a woman three cubits high’

(Classical Greek. Herodotus, *Histories* 1.51, from Giovanni and Celano (2013))

- d. *kai dè kai* [*tò* *eídos* [***tês*** ***gunaikòs***]]
 and PRT and the.N.ACC.SG form.N.ACC.SG the.F.GEN.SG woman.F.GEN.SG
huperepainéōn
 praise-beyond-measure.PRSP.ACT.M.NOM.SG
 ‘And praising the appearance of the woman beyond measure’

(Classical Greek. Herodotus, *Histories* 1.8, from Giovanni and Celano (2013))

For example, in (72d), the genitive phrase *tês gunaikós* refers to a known entity, while *gunaikós* in (72c) is new information and is the subject matter of the following sentence. The importance of the second *gunaikós* is therefore reflected in its pre-nominal position, which for Viti contributes an emphatic or contrastive quality. For Classical Greek, Alexiadou (2002, 100) agrees with Viti, in that pre-nominal genitives tend to be definite and denote humans, and so are often proper names. Moreover, Alexiadou notes that genitives that function as a subject for the noun tend to precede it, while object genitives

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

follow, much like the clause. These analyses of the Ancient Greek noun phrase therefore detect syntactic factors in the placement of genitives¹⁸, but also pragmatic motivations like contrast and newness.

Moreover, possible examples exist in which the genitive precedes a demonstrative or the definite article. Having identified these as D heads, a preceding element gives evidence for movement; the fronting of full phrases like *tôn basiléōn* in (73b) indicates that this movement is specifically phrasal movement, targeting Spec,EmphasisP.

- (73) a. [**Héktoros** *hède* *gunè*], *hòs*
Hector.M.GEN.SG this.F.NOM.SG woman.F.NOM.SG who.M.NOM.SG
aristeúeske *mákhesthai*
be-best.IPRF.IND.ACT.3SG fight.PRS.INF.MED
‘This wife of Hector, who was best at war’

(Homeric Greek. *Iliad* 6.460)

- b. *epeì timân* *eóthasi* *Pérsai*
since honour.PRS.INF.ACT be-wont.PRF.IND.ACT.3PL Persian.M.NOM.PL
[[**tôn** **basiléōn**] *toùs* *paîdas*]
the.M.GEN.PL king.M.GEN.PL the.M.ACC.PL child.M.ACC.PL
‘Since Persians are accustomed to honouring the children of kings’

(Classical Greek. Herodotus, *Histories* 3.15, from Devine and Stephens (2000, 105))

Likewise, occasional instances too of ND and NDA orders, with the noun first, offer support for EmphasisP. We can understand this ordering as the fronting of the noun for prominence;

¹⁸The animate/inanimate and subject/object genitive patterns identified by Viti and Alexiadou could additionally be derived through a layered syntax in the manner of Cartography (cf. Cinque and Rizzi, 2015), with dedicated positions for the different kinds of adjective and genitive.

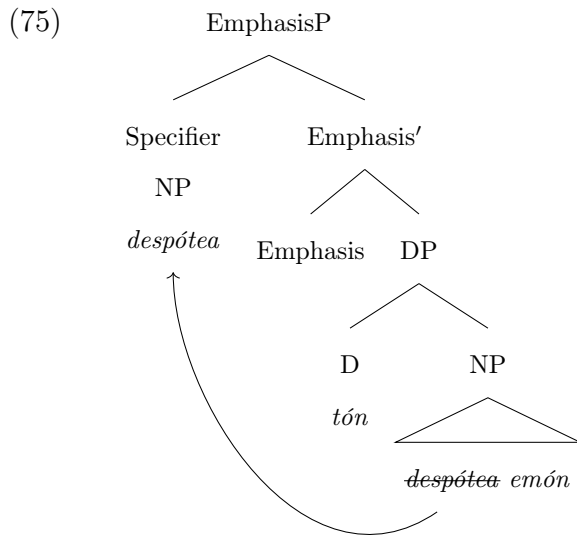
4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

in (74), the head noun *despótea* is preposed before the definite article to emphasise the man that Gyges must kill, or contrast it with the alternative (himself).

- (74) *epeí =me anankázeis [despótea tòn emòn] kteínein*
 since I.ACC.SG force.PRS.IND.ACT.2SG master.M.ACC.SG the.M.ACC.SG
 my.M.ACC.SG kill.PRS.INF.ACT
 ‘Since you are forcing me to kill my master’

(Classical Greek. Herodotus, *Histories* 1.11, from Viti (2008))

A noun phrase like that of (74) can be derived through the movement of the noun higher than the DP, namely EmphasisP.



Finally, some additional support for Emphasis could come from emphatic particles; following their analysis for Sanskrit in §4.2.2, it might be possible to explain common

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

particles like *dě* as lexical exponents of Emphasis (Giannakis, 2021, 418-9). *Dě* has a variety of interpretative functions, but a key quality is adding intensity to the word immediately preceding it (Smyth, 1920, §2840-7).

- (76) [*kartístēn* ***dē*** *tēn* *ge* *mákhēn*]
 strongest.F.ACC.SG PRT that.F.ACC.SG PRT battle.F.ACC.SG
pháto *dúmenai* *andrôn*
 say.AOR.IND.MED.3SG enter.AOR.INF.ACT man.M.GEN.PL
 ‘He said that it was the mightiest battle of warriors to have entered’

(Homeric Greek. *Iliad* 6.185)

As with Sanskrit though, there is great difficulty in distinguishing instances of *dě* that belong to a given noun phrase, from those that belong instead to the clause proper and only happen to intervene in the noun phrase through enclisis. *Dě* is often found second in its clause, thus displaying Wackernagelian behaviour, as it does in (76). We must first identify sure examples of *dě* and other emphatic particles within noun phrases, if we are then to claim that these phrases are evidence of the Emphasis head, and additionally of movement to EmphasisP.

In sum, two historical stages corroborate the overall syntactic structure of Ancient Greek, given at the start of this section in (53) and (54). If we posit a DP component across the history of the Ancient Greek noun phrase, this gives a motivation both for the typical pre-nominal position of Homeric determiners (including pre-articular *ho*) and for the later definite article. In later Greek, however, we have evidence for two functional projections: DemP and DP. The earlier DP ‘split’ over time into DemP and DP, in tandem with the

4.2. ANCIENT NOUN PHRASES: LATIN, SANSKRIT AND ANCIENT GREEK

development of the definite article. Originally, the Homeric DP was been the position of all functional elements that qualified the reference of the noun phrase, not only definite determiners as its head, but also quantifiers. These functions were divided between the two new heads; the new classical DP specifically serves to encode the (in)definiteness of the phrase, while DemP maintains the functions of deixis and quantification. This diachrony is speculative, but it should at least demonstrate that both Homeric and Classical Greek noun phrases are explained well through positing high functional components like DP, as well as an EmphasisP that can scramble the arrangements that the lower structure produces.

4.2.4 THE ANCIENT NOUN PHRASE: INTERMEDIATE SUMMARY

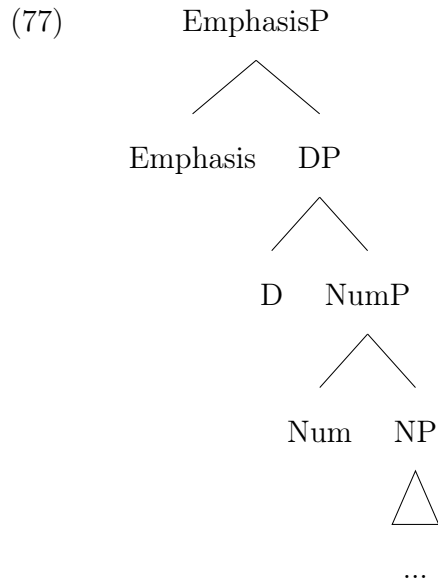
At this juncture, a summary of the behaviour of the early Indo-European noun phrase may be beneficial. Latin, Sanskrit and Ancient Greek (specifically Homeric) noun phrases show striking similarities in their word order, in that there are clear tendencies and patterns, but also considerable variation. Their rich agreement morphology allows the languages to associate discontinuous nouns and adnominal modifiers, many of which are most simply analysed as independent at the syntactic level. Similarly, some continuous phrases may actually be instances of the apposition of independent words or phrases. Barely any syntactic analysis can be proposed without doubt and without alternatives. However, to claim that there is no syntactic structure greater than that the individual noun, adjective or other modifiers would ignore all the patterns and evidence for phrasehood that we find. For instance, adjectives and genitives tend to be the closest words to their head noun, with

numbers in turn preceding them. Determiners meanwhile prefer to stand initially, preceding the relative positions of the aforementioned constituents. This all offers the evidence to propose tentatively that the early Indo-European noun phrase arose from a structured syntax, which included a DP component.

However, such a syntactic proposal gains more explanatory power if we include one higher position that can modify the constituents arranged below it. This functional projection is herein provisionally named ‘EmphasisP’, since it affords a certain prominence to whatever it hosts. It was not necessarily accompanied by distinct prosody. Fronting to EmphasisP was determined by extra-syntactic factors, such as the views of the producer of the noun phrase, and the wider discourse context. In this regard, it is a discourse-informational projection, similar to TopicP and FocusP, but these more specific labels are eschewed to avoid committing the fronted constituents to a particular pragmatic quality for which there is insufficient evidence. While the ‘prominence’ of the initial position is often mentioned in the literature for the different languages, a topical or focal quality is not often recognised by modern scholars.

So far, the noun phrase in three early Indo-European languages has been shown to be derivable through the syntax set out in (1).

4.3. THE EARLY GERMANIC NOUN PHRASE



The task now is to examine the evidence from elsewhere in Indo-European; the next branch of the family is Germanic.

4.3 THE EARLY GERMANIC NOUN PHRASE

4.3.1 EVIDENCE FOR THE EARLY GERMANIC DP: WORD ORDER

In brief, the noun phrase in the representatives of early Germanic, Old English and Old Norse, can be well derived through adopting a syntax identical to (77) above. More accurately, there is strong support for DP and NumP components, and weak evidence for an EmphasisP above them. The analysis of the Germanic evidence begins with determiners,

4.3. THE EARLY GERMANIC NOUN PHRASE

since they have distinct word-order and morphological interactions with adjectives and genitives.

First, within the Old English noun phrase, it is uncontroversial that the language has definite determiners, such as the proximal demonstrative *þes* ‘this’ and the common lexeme *se*. These decline to agree in gender, case and number with a head noun, and are very consistent in their pre-nominal position.

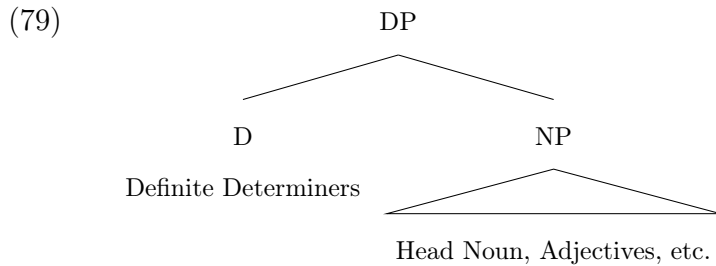
- (78) *Þis* *was* [*þes* *feorðes* *gēares*
 this.N.NOM.SG be.PST.IND.3SG the.N.GEN.SG fourth.N.GEN.SG year.N.GEN.SG
his *rīces*], *and on* [*þȳs* *ȳlcan* *gēare*]
 he.GEN.SG reign.N.GEN.SG and in this.N.INS.SG same.N.INS.SG year.N.INS.SG
gewearð [*se* *mycela* *hunger*] *on*
 become.PST.IND.3SG the.M.NOM.SG great.WK.M.NOM.SG hunger.M.NOM.SG in
Siria *þe* *Lucas* *rēcð* *on* [*þāre*
 Syria.F.DAT.SG REL Luke.M.NOM.SG reck.PRS.IND.3SG in the.N.DAT.SG
bōc *Actus* *Apostolorum*]
 book.N.DAT.SG act.M.NOM.PL apostle.M.GEN.PL
 ‘This was the fourth year of his reign, and in that same year there was the great
 famine that Luke mentions in the book of Acts’

(Old English. *Anglo-Saxon Chronicle* A 30)

The exact status of *se* is much debated; it is often referred to as the Old English definite article, since it both behaves like and is the etymon of the modern definite article *the*. Yet there are differences; in some texts, *se* does not display the obligatoriness in definite contexts expected of an article, and sometimes behaves more like an anaphoric demonstrative. It can also function as an anaphoric and relative pronoun. For such reasons, Ringe and

4.3. THE EARLY GERMANIC NOUN PHRASE

Taylor (2014, 447) state that “it is universally agreed that OE lacked dedicated articles, both definite and indefinite, although the lexical items which later become the articles ... are in some cases arguably used with the force of an article already in OE”. Similarly, the opinion of Hejná and Walkden (2022, 269) is that “Old English has a definite article, at least in later texts and in prose”.¹⁹ Regardless of its interpretation across different places and times, it is not necessary to assume that the emergence of the definite article involved a substantial change to the underlying syntax. All in all, the behaviour of Old English definite determiners can be straightforwardly derived through positing a high DP projection, above the positions of any adjectives, genitives and the head noun.



Old Norse differs in certain respects. Like Old English, the language has demonstratives: *sá* and *sjá*, which may also be pronominal or the initial determiner of a noun phrase. Additionally, Old Norse has *(h)inn*, which is generally referred to as its definite article. Specifically, *(h)inn* acts like a definite article for a head noun on which it is enclitic (Faarlund, 2009). This produces an ND order. As well as phonological indicators of its articular status, *(h)inn* also consistently occurs in definite contexts, or rather “contexts where an

¹⁹The issue of the status of *se* is unsurprising from a general diachronic point of view; our sources attest to later stages in the development of a demonstrative into an article, following the progression outlined by Greenberg (1978).

4.3. THE EARLY GERMANIC NOUN PHRASE

indefinite interpretation would otherwise be the natural” (Faarlund, 2009, 16), such as with common nouns.

- (80) a. *en* [*hamarrinn* *Mjöllnir*] *kom* *í*
but hammer.DEF.M.NOM.SG Mjöllnir.M.NOM.SG come.PST.IND.3SG in
[*mitt* *höfuð* *Hrungni*] *ok*
middle.ST.N.ACC.SG head.N.ACC.SG Hrungnir.M.DAT.SG and
lamði [*hausinn*] *í* [*smán*
hit.PST.IND.3SG skull.DEF.M.ACC.SG in small.ST.M.ACC.SG
mola]
crumb.M.ACC.SG
‘But the hammer Mjöllnir struck the middle of Hrungnir’s head and smashed
the skull into tiny pieces’
(Old Norse. *Prose Edda* 24)
- b. *ok kom* *í* [*hönd* *Hákoni* *konungi*]
and come.PST.IND.3SG in hand.F.ACC.SG Hakon.M.DAT.SG king.M.DAT.SG
upp í [*músinna*] *fyrir* *neðan* *ǫxl*
up in muscle.DEF.F.ACC.SG before below shoulder.F.DAT.SG
‘And it flew into the arm of King Hakon, into the muscle below the shoulder’
(Old Norse. *Heimskringla* 4.31)

Note in (80) how the enclitic definite article is not obligatory; the head of Hrungnir and the shoulder of the king are body parts of known individuals, and can be inferred to exist, *höfuð* and *ǫxl* do not bear the article.

It is also possible for *(h)inn* to precede the head noun. If only before the noun, it has a weak demonstrative or contrastive use (i.e. ‘that’, ‘the other’). This is considered by some to continue a separate lexical item, out of which the definite article emerged (Barnes, 2008,

4.3. THE EARLY GERMANIC NOUN PHRASE

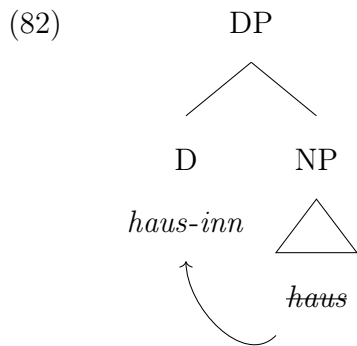
64, 84-6). Most notably, (*h*)*inn* must precede the noun when there is an adjective present.

- (81) a. *sem* [*sjá* *maðr*] *bjóði* *yðr*
as-if this.M.NOM.SG man.M.NOM.SG offer.PRS.SBJ.3SG you.DAT.PL
[*ina* *mestu* *sæmð*]
the.F.ACC.SG greatest.WK.F.ACC.SG honour.F.ACC.SG
‘As if this man offers you the greatest honour’
(Old Norse. *Njál’s Saga* 3)
- b. *er þér* *hafð* [*it* *mesta*
when you.NOM.PL have.PRS.IND.2PL the.N.ACC.SG greatest.WK.N.ACC.SG
blót *yðvart*] *ok sjá* *þar*
offering.N.ACC.SG your.N.ACC.SG and see.PRS.SBJ.1SG there
[*siðu* *yðra*]
custom.M.ACC.PL your.M.ACC.PL
‘When you have your greatest offering and there I will see your customs’
(Old Norse. *Heimskringla* 6.65)
- c. *ok á lagðr* [*inn* *helgi*
and on lie.PSTP.NOM.SG the.M.NOM.SG holy.WK.M.NOM.SG
kross] *með gulli* ... *hann* *hafði*
cross.M.NOM.SG with gold.N.DAT.SG. he.NOM.SG have.PST.IND.3SG
[*hvítt* *merki*]
white.ST.N.ACC.SG standard.N.ACC.SG
‘And laid on was the holy cross in gold ... he had a white standard’
(Old Norse. *Heimskringla* 7.49)

This DAN pattern offers evidence that Old Norse’s definite article, like its demonstratives and its Old English counterpart, has a basic position above that of the noun. This is to say, the article is generated in a DP above the noun, but a subsequent operation is responsible for the inverted ND linear order. We could identify this as a phonological phenomenon, in

4.3. THE EARLY GERMANIC NOUN PHRASE

the manner of Halpern’s prosodic inversion (1992; 1995). However, if so, we would expect the article to be less particular about the category of its host, and to appear enclitic on adjectives too, which it does not. There therefore must be a syntactic component to its word-order behaviour; Faarlund’s (2004, 57) account is that the noun moves up in the syntax to the same structural position as *(h)inn*, which he labels R. This thesis adopts this movement analysis, except with the label D for the functional head that the definite article realises.



When an adjective is present, however, it usually blocks this movement, keeping the noun low and requiring the independent form of the definite article. Rare examples without this blocking do occur, however, as in (83), in which we find an NDA order.

- (83) [Óláfr konungr] lætr þá setja fram
 Olaf.M.NOM.SG king.M.NOM.SG let.PRS.IND.3SG then set.INF forth
 [Orminn langa]
 snake.DEF.M.ACC.SG long.WK.M.ACC.SG
 ‘King Olaf then let Long Serpent set forth’
 (Old Norse. *Heimskringla* 6.93)

4.3. THE EARLY GERMANIC NOUN PHRASE

Furthermore, Old Norse occasionally attests noun phrases which contain both a demonstrative and the independent definite article, as in (84).

- (84) [*þeir* *inir* *íslenzku* *menn*]
these.M.NOM.PL the.M.NOM.PL Icelandic.WK.M.NOM.PL man.M.NOM.PL
fóru *þegar*
go.PST.IND.3PL at-once
‘Those Icelandic men went immediately’

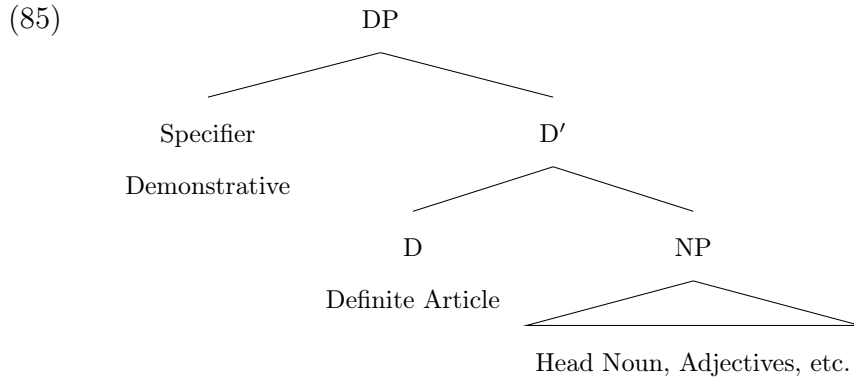
(Old Norse. *Heimskringla* 7.129)

Faarlund (2004, 82-4) proposes that these arise from two functional projections at the top of the nominal structure: a DP and an RP, with the demonstrative and the definite article as the D and R heads respectively. Alternatively, we could derive this phrasal behaviour with only DP, with the definite article as the head and the demonstrative in Spec,DP.²⁰ Locating both in the same projection accounts for their shared behaviour, such as ‘triggering’ the weak declension of adjectives (see §4.3.2). Why the two might co-occur is another matter; this may be taken as evidence for an intermediary stage in the development of (*h*)*inn* into the definite affix of the modern North Germanic languages (cf. Faarlund (2009)), in which the definite article is close to losing all demonstrative force, and so the stronger demonstrative element is brought in to reinforce some intended interpretation, such as contrast.

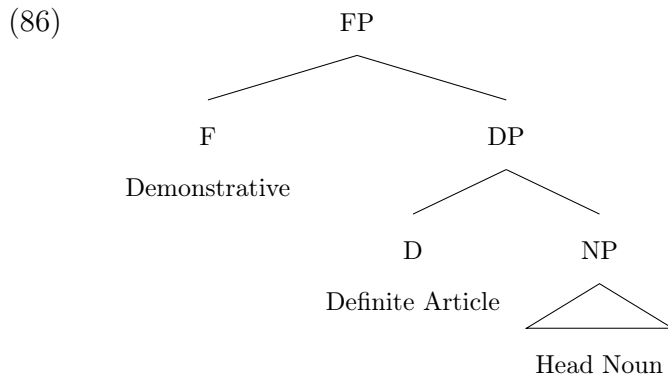
All this goes to say that definite determiners in Old Norse could be derived through the following DP-syntax in (85).

²⁰This relies on considering demonstratives to be phrasal and more complex than definite articles, for which there is precedent analysis (cf. Leu, 2008).

4.3. THE EARLY GERMANIC NOUN PHRASE



This affords a head position for the definite article and a specifier position for demonstratives, but this analysis depends on the syntactic status of Old Norse demonstratives. If we want to consider them as syntactically simple as the definite article, and therefore to be heads as well, then a double-projection structure for Old Norse is necessary, with an additional functional projection above DP, as as Faarlund proposes.²¹



Old Norse therefore fits the same diachrony proposed for Ancient Greek in §4.2.3, namely

²¹This neutral “FP” could be characterised as ‘DemP’, matching the analysis for Ancient Greek in §4.2.3, but the proposal below that *sumr* ‘some’ can also be the head of this FP complicates the identification of its category as solely demonstrative.

4.3. THE EARLY GERMANIC NOUN PHRASE

a split within an earlier unitary DP, which Old English still maintains in (79). It is uncontroversial that Old English has a less developed definite article than Old Norse, and is a later-attested language, so Old Norse has had both time and reason to modify the syntax.

It should also be noted that indefinite noun phrases do not contradict this analysis. Neither Old English nor Old Norse has an indefinite article, not even in the later sources. Bare noun phrases convey an implicit indefinite interpretation.

- (87) a. *hē* *āsette* *ðā* [*rāðels*]
he.NOM.SG set.PST.IND.3SG they.ACC.PL riddle.M.ACC.SG
'He set them a riddle'

(Old English. *Apollonius of Tyre*)

- b. *er* *hann* [*viṛ* *maðr*]
be.PRS.IND.3SG he.NOM.SG wise.ST.M.NOM.SG man.M.NOM.SG
'Is he a wise man?'

(Old Norse. *Njáls Saga* 13)

The etyma out of which the modern singular indefinite articles developed (*ān* and *einn*) still maintain their primary numeral sense of 'one', and lack the obligatoriness of their descendant articles today. Clearly though they have additional functions, including indefiniteness; they serve to introduce key entities into the discourse, like Modern English *a* or *a certain*, as in (88). *Ān* and *einn* behave syntactically like other cardinal numbers, in that they usually precede the head noun, and can themselves be preceded by definite determiners. Additionally, *ān* and *einn* sometimes follow the head noun to convey an alternative

4.3. THE EARLY GERMANIC NOUN PHRASE

meaning of ‘alone, only’ (89).

- (88) a. *þæt hīe gedydon on [ānre*
 that they.NOM.PL reach.PST.IND.3PL in one.ST.F.DAT.SG
wēstre ċeastre on Wirhealum]
 deserted.ST.F.DAT.SG fortress.F.DAT.SG in Wirral.N.DAT.PL
 ‘... until they came to a particular fortress in the Wirral’
 (Old English. *Anglo-Saxon Chronicle* A 894)
- b. *þar var [einn blindr maðr]*
 there be.PST.IND.3SG a.ST.M.NOM.SG blind.ST.M.NOM.SG man.M.NOM.SG
 ‘There was a certain blind man’
 (Old Norse. *Heimskringla* 7.236)

- (89) a. *mid [ūre sāwle ānre] wē*
 with our.F.DAT.SG soul.F.DAT.SG one.ST.F.DAT.SG we.NOM.PL
sculon riht aȝyldan
 must.PRS.IND.1PL rightly pay.INF
 ‘With our soul alone we must pay our dues’
 (Old English. *Vercelli homily* 4)
- b. *þeir gengu í [hús eitt]*
 they.NOM.PL go.PRS.IND.3PL in house.N.ACC.SG one.ST.N.ACC.SG
 ‘They went into one house’
 (Old Norse. *Heimskringla* 7.162)

We can locate *ān* and *einn* in the same syntactic position as either cardinal numbers: a NumP, between the NP and the DP, on the basis of their typical position between determiners and the noun.

4.3. THE EARLY GERMANIC NOUN PHRASE

- (90) a. [pā f̄if̄ ðing] ... þonne bið
the.N.NOM.PL five thing.N.NOM.PL then be.PRS.IND.3PL
þæt God; forðæm [pā f̄if̄]
that.N.NOM.SG God.M.NOM.SG consequently the.N.ACC.PL five
eall nān mennisc̄ man fullīce
all.N.ACC.PL no.M.NOM.SG mortal.ST.M.NOM.SG man.M.NOM.SG fully
habban ne mæg
have.INF not can.PRS.IND.3SG
‘These five things ... are then God; consequently, no mortal man can fully
possess all these five’
(Old English. OE *Boethius* 33)
- b. *Konungurinn mælti: þeir [tveir*
king.DEF.M.NOM.SG speak.PST.IND.3SG this.M.NOM.PL two.M.NOM.PL
ǰðlibornir menn]...
noble-born.ST.NM.NOM.PL man.M.NOM.PL
‘The king spoke: ‘they, two noble-born men...’
(Old Norse. *Heimskringla* 7.94)

Alternatively, for later stages of the languages, we can treat *ān* and *einn* as new exponents of an indefinite D head, as part of the emergence of indefinite articles in English and North Germanic.

A further element with indefinite functions is *sum/sumr* ‘some, a certain’; in the two languages, these are somewhere between an indefinite determiner and a quantifier.²² In both languages, they consistently precede the noun and any adjectives. In Old English, *sum* does not occur with a definite determiner, so we lack that diagnostic for locating *sum*

²²In the former role, akin to indefinite *some* with singular count nouns in Modern English, as *some man*. In the latter, akin to Modern English *some* in *some men*, quantifying and restricting the reference of *man* to a plural subset.

4.3. THE EARLY GERMANIC NOUN PHRASE

in the syntax; it could be in the DP layer, for instance, as an exponent of an indefinite D head, but we cannot securely claim so.

- (91) *þā lēdde mon forþ [sumne blindne*
 then lead.PST.IND.3SG one forth some.M.ACC.SG blind.ST.M.ACC.SG
mon of Angelcynne]
 man.M.ACC.SG of English.N.DAT.SG
 ‘Then a certain blind man of English origin was led forth’
 (Old English. Bede’s *Historia* 2.2)

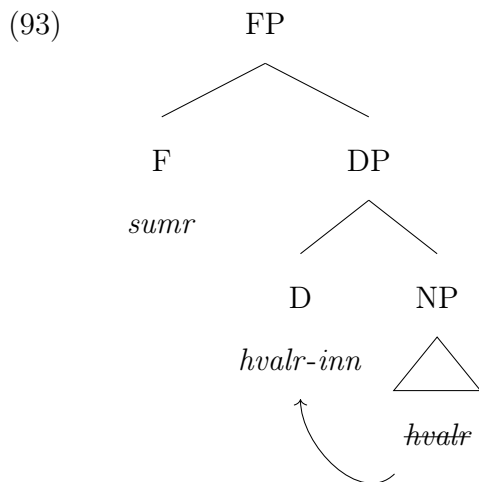
It appears though that Old Norse *sumr* can appear with the enclitic definite article (Nyggaard, 1905, 32).

- (92) a. *at hann sendi aftr [suman*
 that he.NOM.SG send.PST.IND.3SG back some.M.ACC.SG
mjóðinn]
 mead.DEF.M.ACC.SG
 ‘That he sent some mead back’
 (Old Norse. *Prose Edda* 3.6)
- b. *þar var stolinn í brott [sumr*
 there be.PST.IND.3SG steal.PSTP.ST.NOM.SG in away some.M.NOM.SG
hvalrinn]
 whale.DEF.M.NOM.SG
 ‘There was a certain whale stolen away’
 (Old Norse. *Saga of Hrafn Sveinbjarnarson* 12)

This co-occurrence of an indefinite and a definite determiner may seem contradictory, but this is likely a means to distinguish different kinds of indefiniteness. Definiteness is a

4.3. THE EARLY GERMANIC NOUN PHRASE

complex phenomenon, involving both the uniqueness of an entity and the familiarity of the interlocutors with that entity. In Old Norse, *sumr* is used for entities that are specific and known only to the producer of the noun phrase, introduced into the discourse for the first time, while an element like *noġkurr* is used for unknown and non-specific entities (Faarlund, 2004, 73). Assuming Faarlund’s derivation of the enclitic definite article, with the noun moving up to D, *sumr* must be in a higher position still. Identifying this position again depends on what we consider *sumr* to be: a head or a phrase. Since there is nothing obviously more complex about *sumr* than other elements, it is simplest to treat it as the head of the higher projection (the “FP” in (93)), where demonstratives are also located.



While Old English does not face this same word-order issue with its cognate *sum*, quantifiers in general in both languages do require consideration, since their quantifiers co-occur with definite determiners.

4.3. THE EARLY GERMANIC NOUN PHRASE

- (94) a. *hēo* *forsihð* *þonne* [*eall* *ðās*
 she.NOM.SG despise.PRS.IND.3SG then all.ST.N.ACC.PL the.N.ACC.PL
eorðlican *þing*]
 earthly.WK.N.ACC.PL thing.N.ACC.PL
 ‘She despises all these earthly things’

(Old English. OE *Boethius* 18)

- b. *fór* *konungr* *brot ok* [*allr*
 go.PST.IND.3SG king.M.NOM.SG away and all.ST.M.NOM.SG
herinn]
 army.DEF.M.NOM.SG
 ‘The king went away and all the army ...’

(Old Norse. *Heimskringla* 7.223)

The Latin and Ancient Greek data indicate that their quantifiers (e.g. *omnis*, *pās*) are base-generated as adjectives, but then move up to Spec,DP. For Old Norse, there is evidence for similar behaviour, given in §4.3.3. Old Norse quantifiers, namely *allr*, therefore can be derived by locating them in them in Spec,FP. This allows them to precede demonstratives.

- (95) [*ollum* *þessum* *kenningum*]
 all.F.DAT.PL this.F.DAT.PL kenning.F.DAT.PL
 ‘By all these kennings’

(Old Norse. *Prose Edda* 3.61)

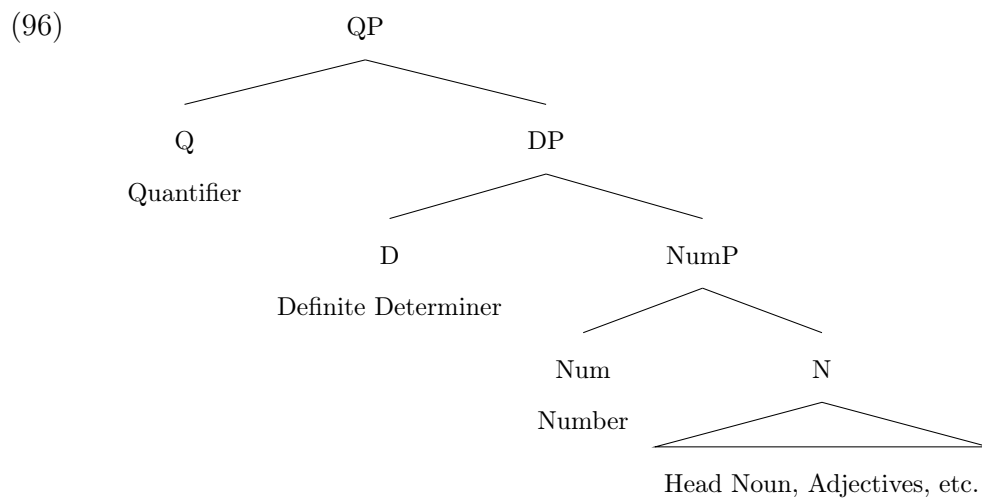
The Old English data with *eall* do permit the same analysis though; *all* consistently precedes definite determiners and has no apparent ‘attributive’ function.²³ Therefore, having located Old English definite determiners in DP, we require a further projection for

²³The phrase *þæt eall* ‘that all’ is attested (cf. *Beowulf* 2427, 3087), but this is insufficient evidence; only a full noun phrase with a D-*eall*-N order would support an adjectival base position for *eall*.

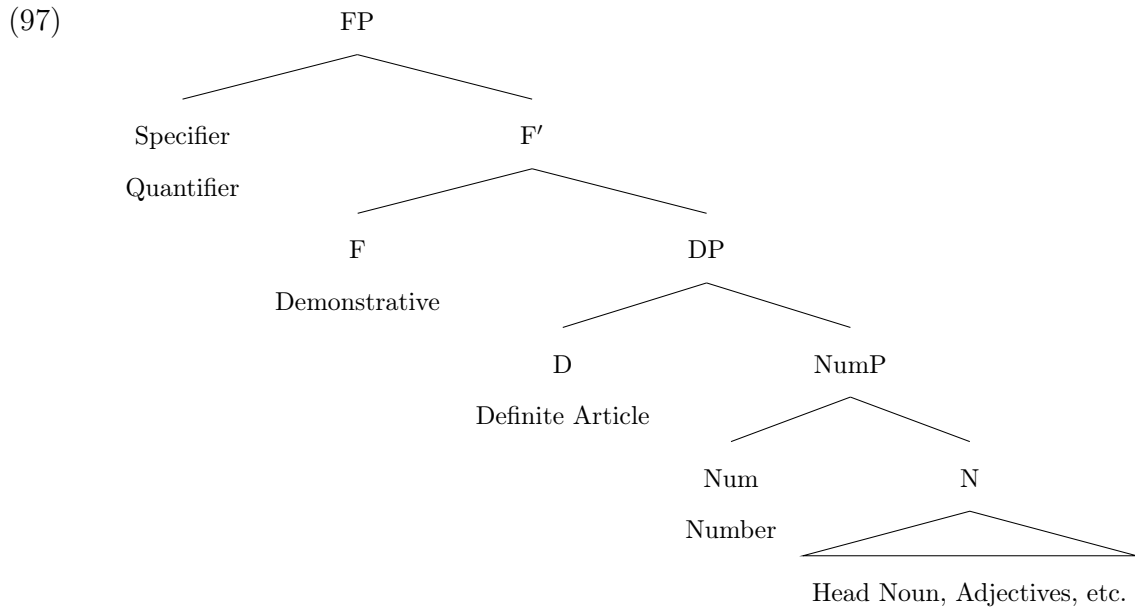
4.3. THE EARLY GERMANIC NOUN PHRASE

the quantifiers that precede them. The most efficient option at this juncture is to propose an additional functional projection for initial quantifiers like *eall*. Given its function, this is specifically a ‘QP’, in which quantifiers only are based generated.

With this addition to the syntax, we arrive at the following two syntactic structures for Old English (96) and Old Norse (97):



4.3. THE EARLY GERMANIC NOUN PHRASE



Having considered at length the word order of the early Germanic noun phrase, it is high time to turn to an alternative source of evidence for the DP: adjectival morphology.

4.3.2 EVIDENCE FOR THE EARLY GERMANIC DP: MORPHOLOGY

More support for the early Germanic DP, in both definite and indefinite noun phrases, comes from adjectival morphology. All early Germanic languages, including Gothic, make a morphological distinction between ‘strong’ and ‘weak’ adjectival endings, the choice of which depends on the syntactic context in which an adjective finds itself (Hogg and Fulc 2011, 146-7; Faarlund 2004, 95). The key factor is (in)definiteness. The weak or ‘definite’ adjectival declension occurs in the presence of a preceding definite determiner, such as a demonstrative or one of the developing definite articles, as well as after possessive pronouns

4.3. THE EARLY GERMANIC NOUN PHRASE

or a genitive phrase in Old English. In other contexts, the adjective displays the strong or ‘indefinite’ declension; it does so when it alone modifies a noun, or follows *ān/einn* and *sum/sumr*.

- (98) a. *þā lēdde mon forþ [sumne blindne*
 then lead.PST.IND.3SG one forth some.M.ACC.SG blind.ST.M.ACC.SG
mon of Angelcynne] ... þā æt nehstan wæs
 man.M.ACC.SG of English.N.DAT.SG then at last be.PST.IND.3SG
Agustinus [mid reohtrē nēðþearfnisse]
 Augustine.M.NOM.SG with right.ST.F.DAT.SG necessity.F.DAT.SG
gebæded ... bæd [God fæder
 bid.PST.M.NOM.SG bid.PST.IND.3SG God.M.ACC.SG father.M.ACC.SG
ælmih̄tigne], þæt hē [þām blindan
 almighty.ST.M.ACC.SG that he.NOM.SG the.M.DAT.SG blind.WK.M.DAT.SG
men] ġesyhðe forġefe
 man.M.DAT.SG sight.F.ACC.SG give.PRS.SBJ.3SG
 ‘Then a certain blind man of English origin was led forth ... then finally was
 Augustine commanded through right necessity ... [he] asked God the Father
 almighty to give sight to the blind man’
 (Old English. Bede’s *Historia* 2.2)

- b. *ok á lagðr [inn helgi*
 and on lie.PSTP.NOM.SG the.M.NOM.SG holy.WK.M.NOM.SG
kross] með gulli ... hann hafði
 cross.M.NOM.SG with gold.N.DAT.SG. he.NOM.SG have.PST.IND.3SG
[hvítt merki]
 white.ST.N.ACC.SG standard.N.ACC.SG
 ‘And laid on was the holy cross in gold ... he had a white standard’
 (Old Norse. *Heimskringla* 7.49)

The strong declension is the more varied of the two; its endings are more distinct and

4.3. THE EARLY GERMANIC NOUN PHRASE

show much less syncretism. Comparatively, the weak paradigm is less varied, with the predominant *-an* ending of Old English, and predominant *-a* and *-u* in Old Norse. The weak declension was able to display such minimal morphological distinction in gender, case and number because of the usual presence of preceding determiners with a more varied paradigm. As Strang (1970, 301) puts it, “so long as a preceding word carried the full *differentiae*, the adjective could appear in a less highly differentiated form”. In Old English, this preceding word is often the definite determiner *se*; in Old Norse, it is the independent, pre-nominal form of the definite article (*h*)*inn*.

The traditional account of the origins of this distinction is that the strong declension continues Indo-European adjectival morphology, while the innovative weak declension emerged at a Proto-Germanic or pre-Germanic stage from the addition of an ‘individualising suffix’ that created morphological *n*-stems (see Ringe 2006, 170; Hogg and Fulker 2011, 146-8). Allen (2008, 39) furthermore discusses how the suffix may also have caused a categorial change, creating nouns out of adjectives. Evans (2019) instead proposes the alternative theory that the endings developed out of linking vowels in compounds of the adjective and a noun.

However it started, it seems that the morpheme’s original purpose was to work as present-day definite articles do, marking the attribute or the whole noun phrase as given information (Hawkins, 2009, 5). In the earlier Gothic language, Ratkus (2017, 112) reports that the distinction is even more robust; “there can be no doubt about the robustness of the *n*-stem suffix as a definite marker ... the presence or absence of the determiner is unrelated to the weak inflection as a carrier of definiteness”. Likewise, Old English poetry, a

4.3. THE EARLY GERMANIC NOUN PHRASE

genre that often maintains conservative grammar, exhibits both weak and strong adjectives without any determiners, and so may preserve a distinction between the two closer to their original purpose (Fischer, 2000, 7). Vocative noun phrases for direct addressees have weak adjectives, which agrees with the definiteness of the weak declension (Hogg and Fulck, 2011, 147). However, adjectives in Old English vocative noun phrases are often accompanied by the definite determiner *se* (Pfaff and Walkden, 2024), as in the *men þā leofestan!* construction (literally ‘men the dearest’) noted by Porck (2020). This would be ungrammatical in present-day English, and is commented on in §4.3.3.

Determination in Germanic adjectives offers further evidence for a DP structure, because of the role that syntactic context plays. At least by the stages of Germanic grammar to which Old English and Old Norse attest, we can propose that there is a DP component, the D head of which has features of gender, case, number and (in)definiteness. Furthermore, it has a requirement that it must be somehow expressed by lexical items. In the case of definite noun phrases, this requirement may be satisfied by definite determiners, namely the lexemes *se* and *(h)inn*. A determiner moreover inflects to express the valued features of the D head. In such cases, as in (98), the D head licenses the syncretic, definite-associated endings of the weak declension.

Apart from a definite determiner, genitive nouns and possessive pronouns can also satisfy the requirement in Old English, since we find both standing initially before weak adjectives.

4.3. THE EARLY GERMANIC NOUN PHRASE

- (99) a. *tō* [[*þæs* *cyninges*] *untruman* *bearne*]
to the.M.GEN.SG king.M.GEN.SG weak.WK.N.DAT.SG child.N.DAT.SG
‘To the king’s sick child’
(Old English. Ælfric’s *Catholic Homilies* 1.8, from Allen (2008, 76))
- b. *þæt hēo* *ne mehton* *būton* [*heora*
that they.NOM.PL not can.PST.IND.3PL without they.GEN.PL
lēoda *gēþafunge* *and lēfnesse*] [*heora*
people.F.GEN.PL permission.F.DAT.SG and leave.F.DAT.SG they.GEN.PL
ealdan *þēawas*] *onscunian*
old.WK.M.ACC.PL custom.M.ACC.PL reject.INF
‘That they could not without the permission and leave of their peoples reject
the old customs’
(Old English. Bede’s *Historia* 2.2)

Triggering the weak inflection does not occur with post-nominal genitives (100a), nor genitives that immediately precede the head noun (100b).

- (100) a. *þæt þū* *Drihtne* *brohtest* [*micel*
that you.NOM.SG Lord.M.DAT.SG bring.PST.IND.2SG great.ST.N.ACC.SG
gestrēon *hāligra* *sawla*]
treasure.N.ACC.SG holy.ST.F.GEN.PL soul.F.GEN.PL
‘That you brought the Lord a great treasure of holy souls’
(Old English. Bede’s *Historia* 2.1)
- b. *ðā* *hirdas* *gehirdon* [*micelne*
the.M.NOM.PL shepherd.M.NOM.PL hear.PST.IND.3PL great.ST.M.ACC.SG
engla *sang*]
angel.M.GEN.PL song.M.ACC.SG
‘The shepherds heard a great angel song’
(Old English. OE *Martyrology* 1)

4.3. THE EARLY GERMANIC NOUN PHRASE

Adopting Yoon's (2002, 70-5) account for Old English genitives, we can understand them to be another means by which the functional projection headed by D is given morphophonological expression; this they achieve through movement into Spec,DP from a lower basic position, and in doing so they satisfy the features of D. In Yoon's account, genitives are (mostly) in complementary distribution with definite determiners because they are alternative lexical exponents of a DP with a definite feature value on its D head. Along with possessive pronouns, these two types of nominal constituent can lexically realise this [Definite] feature because they both "introduce referring expressions, in a sense that an individual or a set of individuals denoted by them is presupposed to exist" (Yoon, 2002, 71). This behaviour and analysis of genitive do not apply to Old Norse though; genitives and possessive pronouns usually appear post-nominally, unless the genitive and the noun have an established connection and form a "conceptual unit" (Faarlund, 2004, 59-60). In Old Norse then, it is only definite determiners that trigger weak-declined adjectives.

The strong declension also offers evidence for the early Germanic DP. We may propose that an indefinite D head exists and likewise requires morphophonological expression. In the absence of lexical material in the DP layer²⁴, D appears as morphological features of a pre-nominal adjective. Specifically, this can be only one adjective, since Old English and Old Norse avoid Adjective-Adjective-Noun orders with two or more pre-nominal adjectives; they instead tend to arrange multiple adjectives in ANA or NAA orders (cf. Carlton 1970, 178; Faarlund 2004, 73). Bech (2017) examines the AAN order in Old English and Old

²⁴Except, perhaps, *sum/sumr* 'some', although if we locate this, like other quantifiers, in a higher FP and not the DP, we gain a motivation for the appearance of *sumne* with strong-inflected adjectives, as in (91).

4.3. THE EARLY GERMANIC NOUN PHRASE

Norse in detail; limiting her analysis to “descriptive” adjectives only, thereby excluding functional modifiers like determiners and quantifiers, she confirms its rarity. The restriction may therefore be semantic, since it seems to apply only to lexical modifiers, those which “describe the noun independently of each other” (Bech, 2017, 15). There is an interesting interplay between declension (strong or weak) and the position of an adjective in Old English; Fischer (2012) notes that post-nominal adjectives are “practically always strong” in their declension, even in the presence of a definite determiner before the noun, and that they are “only weak when the determiner is repeated too”.

- (101) *ēode* [*se* *cyning* *sāriġ*] *tō*
 go.PST.IND.3SG the.M.NOM.SG king.M.NOM.SG sorry.ST.M.NOM.SG to
þām *sēaðe*
 the.F.DAT.SG pit.F.DAT.SG
 ‘The sorry king went to the pit’

(Old English. OE Ælfric’s Catholic Homilies 22.484, from Molencki (2012, 9))

It is not clear what interpretational qualities a strong adjective after a definite phrase like *se cyning* may have had.

We could derive strong inflection through the movement of the adjective up to the indefinite D; the features of the D head will produce the inflection, while the pre-nominal position (and post-quantifier, as in (91)) remains the same. This adjectival movement is the analysis for Old English of van Gelderen (2011, 207). However, additional constituents of the Old English noun phrase contradict this; in the absence of a definite determiner, numbers still precede adjectives, which decline strong (Hogg and Fulk 2011, 147; Fulk 2014,

4.3. THE EARLY GERMANIC NOUN PHRASE

37). In (102a), the number *fēower* precedes the strong adjective *wilde*.²⁵ This seems to be the case with Old Norse too (102b), although further examples of NumAN orders have been difficult to find.

- (102) a. *þā hēt se cāsere gēspōnnan*
 then order.PST.IND.3SG the.M.NOM.SG emperor.M.NOM.SG fix.INF
 [*fēower wilde hors*] *tō scride*
 four wild.ST.N.ACC.PL horse.N.ACC.PL to chariot.N.DAT.SG
 ‘Then the emperor ordered that four wild horses be harnessed to a chariot’
 (Old English. *Martyrology* 65)
- b. *en nú gangi [átta íslenskir menn]*
 but now go.PST.IND.3PL eight Icelandic.ST.M.NOM.PL man.M.NOM.PL
 ‘But now came eight Icelandic men’
 (Old Norse. *Gísels þáttur Illugasonar* 5)

If the adjective were moving to the DP, we would expect it to precede the number, itself base-generated in NumP below. The limited evidence from noun phrases with numbers suggests that the adjective does not move.

Instead, we could derive the strong declension in Old English and Old Norse through an operation of agreement between the null D head and the adjective. Alternatively, this can be modelled through a syntactic or post-syntactic operation of ‘lowering’, in which the D head is expressed as a gender-case-number morpheme affixed to the next suitable word in the linear order, akin to the ‘affix hopping’ behind the Modern English verb (see van Gelderen, 2016). If we assume that the adjective is underlyingly impoverished in features,

²⁵The *-e* is not the canonical ending for the strong neuter accusative plural adjective, but if it were weak, we would expect *-an*.

4.3. THE EARLY GERMANIC NOUN PHRASE

it becomes a viable target for the feature sharing of the D head. In the case of $D_{\text{Indefinite}}$, the adjective is the target for all the full features of D, namely gender, case, number and indefiniteness; the combination of the two produces the strong declension. In the case of D_{Definite} , the adjective receives only a definite feature, producing the weak declension. This feature sharing of D will not target any cardinal numbers preceding the adjective²⁶, since numbers do not share the same featural impoverishment of adjectives.

Post-nominal adjectives and ANA orders can be analysed as instances of apposition, that is, with a separate syntactic structure for the following adjective. This accords with the views of Carlton (1970, 177-8) and Faarlund (2004, 68) that the basic position of the adjective in the two languages is higher than the noun and therefore pre-nominal. Likewise, Fischer (2012, 2) proposes that a post-nominal adjective “forms a separate constituent (i.e. it does not function as a modifier governed by the head noun)”. Pysz (2009, 287) similarly argues that some post-nominal adjectives are in fact “reduced relatives”. The strong declension of post-nominal adjectives noted for Old English by Fischer (2012) is explicable as the interaction between the adjective and the D head of the second syntactic structure. This does not explain why Old English and Old Norse have their apparent ban on *AAN orders, but this can remain unresolved.

The primary goal of this section has been to demonstrate that the adjective in Old English and Old Norse suits a syntactic account of the noun phrase that involves a DP layer. Its D head encodes features of definiteness in the syntax. It may be realised by a definite

²⁶Most cardinal numbers in noun phrases do not inflect, while numbers 1-3 in Old English and 1-4 in Old Norse inflect mostly according to the strong declension and without a strong/weak distinction (Hogg and Fulk 2011, 185-8; Faarlund 2004, 34).

4.3. THE EARLY GERMANIC NOUN PHRASE

determiner, or may be lexically null. In the latter case, definite D gains lexical expression in Old English through hosting genitives or possessive pronouns in Spec,DP. Indefinite D is realised as an affix on a pre-nominal adjective. The claim is not, however, that this exact syntactic situation should be reconstructed back to a Proto-Germanic point. It understands the weak declension to be less complex than the strong, since the strong declension is derived through bearing all the features (gender, case, number, (in)definiteness) of an indefinite D head. This is the inverse of the philological view, in which the original state of Germanic determination was that weak adjectives were derived, while strong adjectives continued Indo-European adjectival endings. Further syntactic work could try to track the diachrony behind the Old English and Old Norse situations, specifically how the combination of weak adjectives with definite determiners diminished the original markedness of the weak declension.

4.3.3 EVIDENCE FOR AN EARLY GERMANIC EMPHASISP: NOUN FRONTING AND POSSESSOR-ARTICLE ORDERS

This final section concerns itself with deriving a minority of orders in Old English and Old Norse, which the syntax in (96) and (97) has difficulty deriving. One is the possible post-nominal position of numbers in Old Norse noun phrases. In some, we can detect the workings of discourse-informational factors on the word order.

4.3. THE EARLY GERMANIC NOUN PHRASE

- (103) a. *þau áttu [sex börn], [dætr þrjár]*
 they.NOM.PL have.PST.IND.3PL six child.N.ACC.PL daughter.F.ACC.PL
þrjár ok [sonu þrjá]
 three.F.ACC.PL and son.M.ACC.PL three.M.ACC.PL
 ‘They had six children, three daughters and three sons’
 (Old Norse. *Njál’s Saga* 20)
- b. *Útþrændir fóru [fjórum skipum]*
 Outer-Thronds.M.NOM.PL go.PST.IND.3PL four.N.DAT.PL ship.N.DAT.PL
suðr ... ok drápu þar [presta þrjá] ok
 south and slay.PST.IND.3PL there priest.M.ACC.PL three.M.ACC.PL and
brenndu [kirkjur þrjár]
 burn.PST.IND.3PL church.F.ACC.PL three.F.ACC.PL
 ‘The Outer Thronds went with four ships south ... and there slew three
 priests and burned three churches’
 (Old Norse. *Heimskringla* 4.18)
- c. *ok hugðu þeir vera í sundinu*
 and think.PST.IND.3PL they.NOM.PL be.INF in sound.DEF.N.DAT.SG
[kaupskip tvau]
 trade-ship.N.ACC.PL two.N.ACC.PL
 ‘And they thought there to be two merchant ships in the sound’
 (Old Norse. *Heimskringla* 7.30)

Alongside the usual and likely unmarked pre-nominal numbers, (103) attests the inverse order. In each case, the head noun contrasts with something else. In example (103a), daughters contrast with sons; in (103b), priests contrast with churches; and in (103c) the false belief that the ships are only merchants’ ships is contrary to the known fact that their purpose is military. The extra-clausal context therefore permits us to identify a special status for the noun in these orders. Either we derive through the apposition of

4.3. THE EARLY GERMANIC NOUN PHRASE

two structures, or more economically through the movement of the noun within one noun phrase, up past the number a discourse-associated position where the contrast can be communicated.

A related phenomenon is the order of late quantifiers in Old Norse, such as *allr* ‘all’. Carlson’s (1978) view is that quantifiers are of no different syntactic status to lexical adjectives; their usual initial position is derived through a general operation of movement available to all adjectives. This origin can explain the later position, after the demonstrative, that these words occasionally occupy.

- (104) a. *ek skal eignask [ríki þat*
 I.NOM.SG shall.PRS.IND.1SG own.INF.PASS realm.N.ACC.SG that.N.ACC.SG
alt
 all.ST.N.ACC.SG
 ‘I shall own that whole kingdom’
 (Old Norse. *Heimskringla* 7.35)
- b. *hann vill gefa Hákonni jarli ...*
 he.NOM.SG want.PRS.IND.3SG give.INF Hakon.M.DAT.SG earl.M.DAT.SG
[land þat allt]
 land.N.ACC.SG that.N.ACC.SG all.ST.N.ACC.SG
 ‘He wished to give Earl Hakon ... all of that land’
 (Old Norse. *Heimskringla* 7.171)

By adopting a movement analysis, we can understand the quantifier *allr* in (104) as syntactically low, below the level of the DP with the demonstrative *þat*. Instead, it is the noun that instead moves up to the highest position. If quantifiers were base-generated in their own dedicated and high FP, these low positions would be harder to account for. We can

4.3. THE EARLY GERMANIC NOUN PHRASE

derive the orders in (104) through the same analysis as for (103): the fronting of the noun, in which position it can gain prominence and express contrast. These noun-initial phrases therefore offer evidence for a comparable EmphasisP component in Old Norse, above the DP layer, to which the noun moves.

It is noteworthy that in noun phrases like (104), the noun is fronted, while *allr* stays lower, below DP; this gives the impression that fronted nouns and initial quantifiers are in fact competing for the same position. In §4.3.1, quantifiers are proposed to be in Spec,FP, yet this analysis of later *allr* suggest that quantifiers move to Spec,EmphasisP instead. What would refute this idea, keeping quantifiers in Spec,FP, would be the attestation of NQD orders in Old Norse, in which both the fronted noun and the quantifier precede the definite determiner in DP. Further study is necessary to see if the sources attest such orders.

A related diachronic idea is that proposing an EmphasisP for an earlier stage of Old Norse offers us an origin for the enclitic definite article. The typical ND of Old Norse may have arisen through the fronting of the noun to EmphasisP; like Ancient Greek demonstratives, the fronting of the noun may have backgrounded the originally demonstrative (*h*)*inn*, giving it a desired anaphoric interpretation. This arrangement was over time reanalysed into the movement of the noun to D, producing the typical ND order and occasional NDA orders (as in (83)) by the time of the Old Norse sources.

Another phenomenon deserving of attention is the combination of definite determiners and possessive pronouns in Old English. Examples of noun phrases exist which begin with either *se* and then a possessive pronoun (DPoss), or the other way around (PossD),

4.3. THE EARLY GERMANIC NOUN PHRASE

as in (105). These two orders are not simply inversions of each other; the PossD order is comparatively choosier about the context in which it appears. Allen (2007, 5) makes two key observations: first, that only *se* can be used in PossD (as opposed to *þes* ‘this’); second, that PossD always precedes an adjective, while adjectives are infrequent after DPoss. Sommerer (2018, 235-9) confirms these observations, and adds that PossD is in general the more common order of the two. PossD is exemplified in (105) below.

- (105) a. *mid* [*eallum* *mīnum* *goldhordum*] *for* [*þīnum*
with all.N.DAT.PL my.N.DAT.PL gold-hoard.N.DAT.PL for your.N.DAT.SG
ðāem *hālġum* *naman*]
the.N.DAT.SG holy.WK.N.DAT.SG name.N.DAT.SG
‘With all my treasures for your holy name’
(Old English. *Martyrology* 71)
- b. *mid* [***heora*** ***þām*** *swētestan* *ġehātum*]
with they.GEN.PL the.N.DAT.PL sweetest.WK.N.DAT.PL promise.N.DAT.PL
‘With their sweet promises’
(Old English. Bede’s *Historia* 1.16)
- c. *hē* *sealde* [***his*** ***þone*** *readan*
he.NOM.SG sell.PST.IND.3SG he.GEN.SG the.M.ACC.SG red.WK.M.ACC.SG
ġim], *þæt* *wæs* [***his*** ***þæt***
jewel.M.ACC.SG that.N.NOM.SG be.PST.IND.3SG he.GEN.SG the.N.ACC.SG
hālġe *blōd*]
holy.WK.N.ACC.SG blood.N.ACC.SG
‘He sold his red jewel, namely his holy blood’
(Old English. *Blickling Homilies* 1, from Sommerer (2018, 237))

Within the DP so far proposed for Old English, one possibility is that in phrases like those

4.3. THE EARLY GERMANIC NOUN PHRASE

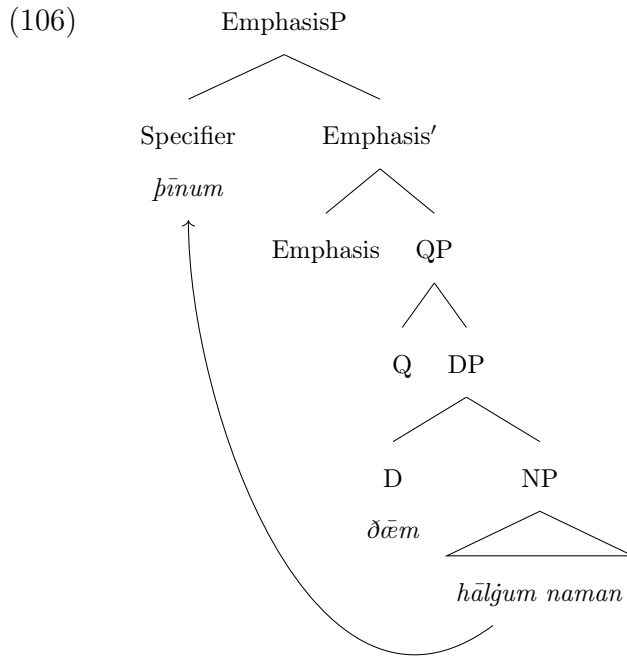
of (105), the definite determiner *se* is the exponent of the D head, leaving Spec,DP free for the possessive pronoun to occupy. Another possible analysis is that these orders arise from apposition, with two underlying syntactic phrases. This analysis works well for the DPoss orders, since these lack the apparent restrictions of PossD (namely the presence of an adjective), and we can claim that the initial *se* or *þes* works as a cataphoric determiner, anticipating the noun phrase proper, akin to Modern English *this: his cat*. PossD does not suit the same interpretation though, and neither of these analyses account for its co-occurrence with adjectives.

Allen's (2007, 6-11) proposal is that there is only one DP projection at the top of the structure, and that only the possessor occupies its specifier, while the definite determiner forms its own complex phrase with the adjective, and together they stand as an adjunct to the NP. In other words, noun phrases like *þīnum ðæm hālgum naman* comprise a possessive pronoun (in Spec,DP), a complex phrasal adjective and the noun. The phrase *ðæm hālgum* is an embedded DP; like an adnominal genitive, it is another noun phrase embedded within the extended projection of the head noun. The theory neatly provides a derivational reason for the co-occurrence of adjectives after an initial PossD sequence. However, Allen's analysis does not offer a reason for why the adjective needs its own definite determiner; the possessive pronoun and the weak declension are usually sufficient to express definiteness.

Instead, this thesis suggests that the word order, restrictions and interpretation can be explained through a noun-phrase syntax with DP and EmphasisP, with the tentative interpretation of the possessive pronoun in the PossD order as emphatic or contrastive. This works at least for (105a), in which *þīnum* 'your' contrasts with the preceding *mīnum*

4.3. THE EARLY GERMANIC NOUN PHRASE

‘my’.



On this basis, EmphasisP can again be brought in to motivate the position for these initial possessives. Assuming that the possessive pronoun is *not* within the DP, the requirement for definite D to be lexically realised motivates the co-occurrence of the definite determiner *se*. More precisely, the presence of an adjective requires the D head to license its declension. The overall noun phrase is definite, so the adjective must appear with weak declension. For this to happen, the adjective requires a definite D head to provide the adjective with a [Definite] feature. Since D_{Definite} has the additional requirement to be lexically realised, adjectives force the appearance of the default definite determiner *se*. The idea that the

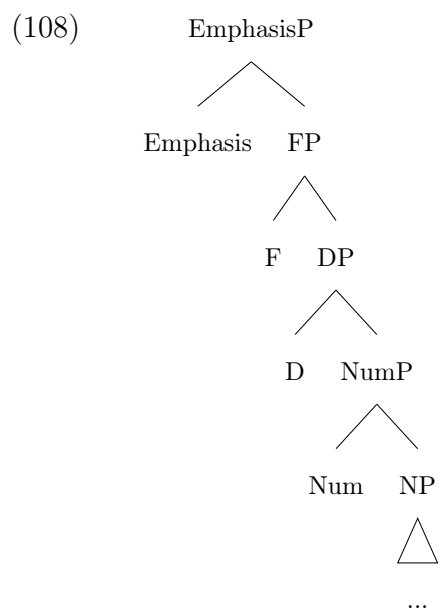
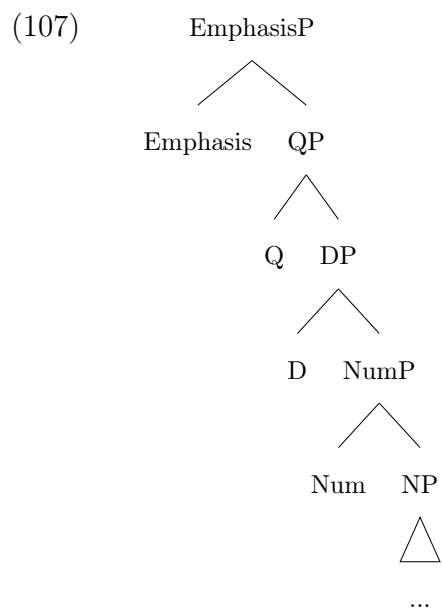
4.3. THE EARLY GERMANIC NOUN PHRASE

weak declension can ‘force’ the co-occurrence of *se* can also account for its appearance in vocative phrases that contain adjectives, as noted by Pfaff and Walkden (2024). This is all a very speculative analysis, but it demonstrates at least that a syntactic structure larger than a single DP successfully derives a notable minority of Old English noun-phrase word orders.

4.3.4 THE EARLY GERMANIC NOUN PHRASE: INTERMEDIATE SUMMARY

To summarise these arguments from early Germanic: the observable behaviour of the noun phrase in Old English (107) and Old Norse (108) can be derived through the two similar syntactic structures below.

4.3. THE EARLY GERMANIC NOUN PHRASE



4.3. THE EARLY GERMANIC NOUN PHRASE

This structure includes a DP component, which is responsible for encoding the (in)definiteness of the overall noun phrase. It is in this projection that we find constituents of the noun phrase associated with definiteness. The D head has certain requirements; if [Definite], it must be given lexical expression. This occurs either through a base-generated definite determiner in the DP (such as a demonstrative or definite article), or in Old English through triggering the movement of a genitive phrase or possessive pronoun. Furthermore, a [Definite] value triggers the weak declension of an adjective in the lower structure (“NP” in (107) and (108)). If the D is [Indefinite], and if an adjective is present, the D head interacts with the adjective to provide it with full features of gender, case and number, resulting in the forms of the strong declension.

However, a DP alone cannot derive all the attested word orders. Quantifiers typically precede definite determiners, and the Old English data require a projection for quantifiers above the DP; this is a dedicated ‘QP’ in (107). Old Norse on the other hand may be conservative in this regard, maintaining the derivation of initial quantifiers through movement, which we also see in Latin and Ancient Greek. Old Norse is innovative in its split between FP and DP in (108); this is the product of the continued grammaticalisation of *(h)inn* into a marker of definiteness. While *(h)inn* is the head of DP, and can trigger the movement of the noun, demonstratives like *sjá* stand in the higher FP. Moreover, there are fragments of evidence from Old English and Old Norse that the higher projection is one characterised by prominence, such as for conveying contrast. Furthermore, a case can be hesitantly made for Old Norse at least, on the basis of its NDQ orders (104), that quantifiers and fronted constituents compete for the same syntactic position. Overall, if

4.4. SLAVIC AND CELTIC NOUN PHRASES

we remove the syntactic innovations in the two languages, what remains is a syntactic structure with EmphasisP, DP, NumP and the lower lexical NP, a configuration identical to what has been proposed for the older languages of the set in §4.2.4.

4.4 SLAVIC AND CELTIC NOUN PHRASES

4.4.1 THE OLD CHURCH SLAVONIC NOUN PHRASE

This section addresses the two remaining members of the language set, beginning with the representative of the Slavic family, Old Church Slavonic. Within its noun phrase, we find evidence for a DP component, as well as very speculative support for an EmphasisP in an older stage of the language. As with the constituents of the clause, the word order of the OCS noun phrase closely matches that of Greek originals that it typically translates. However, divergences between the Greek sources and OCS exist that offer a window on the native Slavic syntax. One involves the language's set of definite determiners, namely its demonstratives: proximal *сь* 'this', distal *тъ* 'that', distal *онъ* 'that, yonder'. All three can appear either pronominally or as part of a noun phrase.

- (109) a. [*сь* *члѣкъ*] *начѣтъ* *зѣдати*
this.M.NOM.SG person.M.NOM.SG begin.AOR.3SG build.INF
'This man began to build'
(OCS. *C. Mar.* Lk. 14.30)
- b. *въ* [*тѣ* *нощѣ*] *бѣдете* *дѣва*
in that.F.ACC.SG night.F.ACC.SG be.FUT.3DU two.M.NOM.DU

4.4. SLAVIC AND CELTIC NOUN PHRASES

‘On that night there will be two’

(OCS. *C. Zogr.* Lk. 17.34)

- c. *prěšedъše učeniци ego na [onъ*
 reach.PSTP.NOM.PL disciple.M.NOM.PL he.GEN.SG on that.M.ACC.SG
polъ]
 half.M.ACC.SG
 ‘When his disciples reached the far side’

(OCS. *C. Mar.* Mt. 16.3)

These different demonstratives frequently parallel the word order of their Greek counterparts *hoûtos* ‘this’ and *ekeînos* ‘that’, although OCS does not have an obligatory definite article equivalent to Greek *ho*. These elements may follow the word or phrase that they specify, again in imitation of the Greek; in (110b), the demonstratives *togo* and *tø* appear after the quantifier and noun, matching the similarly postposed *ekeînou* and *ekeînen*, which have a weaker, anaphoric reading (see §4.2.3).

- (110) a. *hoi ándres [toû tórou*
 the.M.NOM.PL man.M.NOM.PL the.M.GEN.SG place.M.GEN.SG
ekeînou] *apésteilan eis [hólēn tēn*
 that.M.GEN.SG send.AOR.IND.ACT.3PL into all.F.ACC.SG the.F.ACC.SG
perikhōron ekeînen]
 surrounding-area.F.ACC.SG that.F.ACC.SG
 ‘The men of that place sent word into the surrounding area there’

(Koine Greek. Mt. 14:35)

- b. *møži [města togo] posъlaše vъ*
 man.M.NOM.PL place.N.GEN.SG that.N.GEN.SG send.AOR.3PL into
[vъsø stranø tø]
 all.F.ACC.SG country.F.ACC.SG that.F.ACC.SG

4.4. SLAVIC AND CELTIC NOUN PHRASES

‘The men of that place sent word into the countryside there’

(OCS. *C. Mar.* Mt. 14:35)

However, there are interesting instances of *s̆b* and *t̆b* that do not match the Greek word order exactly; in fact, their position would be unacceptable in Greek. In these, *s̆b* and *t̆b* appear in the second position of the phrase, following a noun, adjective or other category of word.

- (111) a. *labòñ* [*toùs* *heptà ártous*]
 take.AORP.M.NOM.SG the.M.ACC.PL seven bread.M.ACC.PL
 ‘Having taken the seven loaves’

(Koine Greek. Mk. 8.6)

- b. *priem̆b* [*sedm̆b* *t̆o* *xlěb̆b̆*]
 take.PSTP.M.NOM.SG seven.ACC.SG that.F.ACC.SG bread.M.GEN.PL
 ‘Having taken the seven loaves’

(OCS. *C. Mar.* Mk. 8.6, from Dimitrova (2008, 45))

- (112) a. *blépeis* [*taútas* *tàs* *megálas* *oikodomás*]
 see.PRS.IND.ACT.2SG this.F.ACC.PL the.F.ACC.PL great.F.ACC.PL
 building.F.ACC.PL

‘Do you see these great buildings?’

(Koine Greek. Mk. 13.2)

- b. *vidiši* *li* [*velikaa* *si* *z̆bdaniě*]
 see.PRS.2SG Q great.N.ACC.PL this.N.ACC.PL building.N.ACC.PL
 ‘Do you see the great buildings?’

(OCS. *C. Mar.* Mk. 13.2, from Dimitrova (2008, 44))

4.4. SLAVIC AND CELTIC NOUN PHRASES

In these discrepancies between the Greek and the Slavic, we can identify traces of another native system of definiteness marking. While some have stated that at least *tъ* in these cases “retains in Old Slavic its full value of a demonstrative” (Vaillant, 1942, 1), Dimitrova (2008) has proposed instead that second-position *sb* and *tъ* are products of grammaticalisation and function as definite articles. Such definite articles are exponents of a D head.

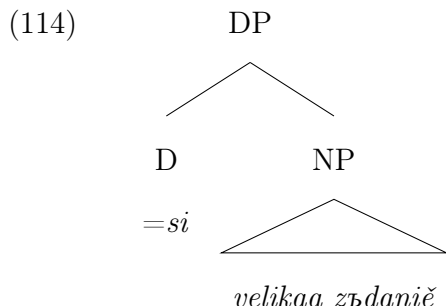
The second position of Old Church Slavonic definite articles Dimitrova (2008, 51-5) derives by N-to-D movement, a syntactic process. An alternative derivation, utilised for its clauses in Chapter 2, is that non-demonstrative *sb* and *tъ* have become enclitics, a byproduct of their grammaticalisation, and trigger prosodic inversion (Halpern, 1992, 1995) with whatever stressed constituent they precede, be it a noun or adjective. Although OCS manuscripts do not consistently include word breaks, there is even tenuous evidence for definite articles that trigger some kind of phonological fusion with their preceding host; in (113b), the noun *rodъ* appears not with its usual nominative singular ending, the ultra-short vowel *-ъ*, but as *rodo*. This upgrade to a longer vowel may be the result of the enclisis of *sb* and its combination with *rodъ* as one prosodic unit.

- (113) a. *tí* [*hē* *geneà* *haútē*]
 why the.F.NOM.SG generation.F.NOM.SG this.F.NOM.SG
zēteî *sēmeîon*
 seek.PRS.IND.ACT.3SG sign.N.ACC.SG
 ‘Why does this generation seek a sign?’
 (Koine Greek. Mk. 8.12)
- b. *čъto* [*rodo* *sb*] *znameniě* *ištetъ*
 why generation.M.NOM.SG this.M.NOM.SG sign.N.ACC.SG seek.PRS.3SG
 ‘Why does this generation seek a sign?’

4.4. SLAVIC AND CELTIC NOUN PHRASES

(OCS. *C. Mar.* Mk. 8.12, from Dimitrova (2008, 50))

If we agree with the evidence and the syntactic analysis, then Old Church Slavonic definite noun phrases like (112b) give us grounds to extrapolate a native DP structure out from out of the shadow of Greek influence. In this, the D could be realised by a non-enclitic demonstrative like *sb* or by its enclitic, article-like counterparts.



If we accept that second-position *sb* and *tb* are definite articles, or at least quasi-articular, we may wonder about the origin of this bit of syntax, specifically whether it is native to Slavic or yet another external influence. The majority of the Slavic languages have managed perfectly well without definite or indefinite articles, and article-less noun phrases are clearly an option for OCS too; countless examples (e.g. (117b) and (118b)) show that its authors were happy to leave Greek *ho* untranslated. However, there are some Slavic languages with definite articles, namely Bulgarian and Macedonian, the modern-day descendants of the South Slavic dialect of which OCS was a literary variety.²⁷ There

²⁷Dimitrova’s analysis fits well within the South Slavic (and wider Balkan) phenomenon of suffixal definite articles; in Bulgarian and Macedonian, the definite article is likewise “affixed to the leftmost inflecting member of a noun phrase”, be it a noun or adjective (Sussex and Cubberley, 2006, 235). Dimitrova

4.4. SLAVIC AND CELTIC NOUN PHRASES

is evidence too from northern, non-standard dialects of Russian (Sussex and Cubberley, 2006, 552), for which such a strong Greek influence cannot be proposed. Furthermore, across all Slavic languages, an initial or early position for any definite determiner is the rule (Sussex and Cubberley, 2006, 347-8). Perhaps the best argument for the native origin of the OCS definite article is simply its redundancy. The evidence from both the sources and comparative Slavic grammar shows that it is only optional for definite noun phrases, and it does not consistently translate the definite article or another lexical item of Greek. All this gives the impression that it was already a part of the syntax of the translators, and was something that they were free to utilise according to their personal discretion.

On the evidence of Old Church Slavonic, we can claim that the emergence of the South Slavic suffixal definite article was underway by the time of the earliest sources, and was not, at least directly, another consequence of contact with Greek Christian texts. Its emergence out of the demonstratives *sb* and *tb* followed a common and well-documented path of grammaticalisation (cf. Greenberg, 1978); this process presumably involved the weakening of both its deictic force and its phonology, leading to its enclitic status and second position. This may be yet another example in Indo-European of a late position in the noun phrase facilitating the creation of ‘weaker’ definite determiners, in the absence of different lexical items. It has been claimed in this thesis for Ancient Greek (§4.2.3) and Old Norse (§4.3.3) that word order and syntactic movement are used to distinguish strong (i.e. ostensive, emphatic, contrastive) demonstratives from weak ones (i.e. anaphoric, discourse-tracking, therefore provides the development of this now obligatory feature of Bulgarian and Macedonian with an earlier, nascent stage, and it is unlikely that this South Slavic behaviour is all the product of scholarly engagement with Greek texts.

4.4. SLAVIC AND CELTIC NOUN PHRASES

definite-marking) while using the same lexical elements. We can apply the same analysis to the diachrony of OCS too, as an origin for the attested determiner-second orders.

In fact, this may have happened twice in the history of Slavic: later with nouns, but earlier with adjectives. Old Church Slavonic uses adjectival declension to convey definiteness. Its adjectives have short and long forms; in OCS (and vestigially across Slavic today), their distribution corresponds to a difference in the definiteness of the overall noun phrase. Combined with a noun, a short-form adjective “denotes that the combination is presented as a new one”, while a long form “presents the quality as one already known and specifically known to belong to the particular substantive which it modifies” (Lunt, 2001, 142), such as with discourse anaphora and proper nouns.

- (115) a. *privěse* *къ нему* *слѣпа* ... *и*
 bring.AOR.3PL to he.DAT.SG blind.SF.M.GEN.SG and
емъ *за [ръко]* *слѣпаго*]...
 take.PST.M.NOM.SG by hand.F.INS.SG blind.LF.M.GEN.SG
 ‘They brought to him **a** blind man ... and having taken **the** blind man by
 the hand...’

(OCS. *C. Mar.* Mk. 8.22-3, from Lunt (2001, 142))

- b. *ити* *въ [Гѣонѣ]* *огньноѣ*
 go.INF in Gehenna.F.ACC.SG fiery.LF.F.ACC.SG
 ‘To go into fiery Gehenna’

(OCS. *C. Mar.* Mk. 9.47, from Lunt (2001, 142))

There is no evidence for a specific position for long and short forms with respect to the noun; Lindberg (2013, 72-3) demonstrates that their pre- or post-nominal position is clearly

4.4. SLAVIC AND CELTIC NOUN PHRASES

determined by the Greek order, with fewer than ten mismatches in her corpus. However, the choice of short or long does seem to be a deliberate decision on the part of the translators, based on productive pragmatic and semantic principles. It does not simply mimic definite markers in Greek; vocative adjectives in nominal phrases are typically in the long form (Lunt, 2001, 142), reflecting the known status of the addressee, even though this does not correspond to a particular element of the Greek prototype, namely the definite article *ho*. Definiteness marking on adjectives therefore seems to be a feature of the native Slavic grammar. We could therefore derive this behaviour through some functional category in the underlying syntax; this could be a case of agreement between the D head and the adjective (as proposed for Germanic in §4.3.2), or a definiteness head together with the adjective within a phrasal lexical constituent.

In terms of diachrony, the consensus seems to be that the long forms emerged through the fusion between adjectives and the third-person pronoun **jь* (Sussex and Cubberley, 2006, 264). This can be reconstructed back to the Proto-Slavic stage, and perhaps further back still to a Proto-Balto-Slavic point, since Lithuanian adjectives also exhibit long-form adjectives very similar in function and etymology (Šereikaitė, 2019, 87-8). However we account for their definiteness in OCS syntax, long-form adjectives may offer very tentative support for an EmphasisP in an earlier stage of the language. Following the analysis of phrase-final demonstratives in Ancient Greek (§4.2.3), through movement of the adjective proper to EmphasisP, this would have produced an Adjective-**jь* order; this over time fused into the definite long-form adjectives attested in OCS.

Additional, and very indirect, evidence for an EmphasisP in OCS comes from the

4.4. SLAVIC AND CELTIC NOUN PHRASES

considerable variation in noun-phrase constituent order. Adjectives, for instance, whether short or long, can be either pre-nominal (116) or post-nominal (117).

- (116) a. [*ho* *ponēròs* *ánthrōpos*] *ek* [*toû*
the.M.NOM.SG evil.M.NOM.SG man.M.NOM.SG from the.M.GEN.SG
ponēroû *thēsauroû*] *ekbállei*
evil.M.GEN.SG treasury.M.GEN.SG produce.PRS.IND.ACT.3SG
ponērá
evil.N.ACC.PL
‘An evil man produces evil things out of their store of evil’
(Koine Greek. Mt. 12.35)
- b. [*zъly* *čкъ*] *otъ* [*zъlaago* *съkrovišta*]
evil.M.NOM.SG person.M.NOM.SG from evil.N.GEN.SG treasury.M.GEN.SG
iznositъ *zъlaě*
carry-out.PRS.3SG evil.N.ACC.PL
‘An evil man produces evil things out of their store of evil’
(OCS. *C. Zogr.* Mt. 12.35)
- (117) a. *egó* *eimi* [*ho* *poimén*
I.NOM.SG be.PRS.IND.1SG the.M.NOM.SG shepherd.M.NOM.SG
ho *kalós*]
the.M.NOM.SG good.M.NOM.SG
‘I am the good shepherd’
(Koine Greek. Jh. 10.11)
- b. *azъ* *estъ*(sic) [*pastyrъ* *dobry*]
I.NOM.SG be.PRS.1SG shepherd.M.NOM.SG good.M.NOM.SG
‘I am the good shepherd’
(OCS. *C. Mar.* Jh. 10.11)

4.4. SLAVIC AND CELTIC NOUN PHRASES

The same goes for dependent genitive and dative words and phrases (118), possessive pronouns (119), numbers (120) and quantifiers (121).

- (118) a. *hóte hoi nekroì akóúsousin [tês*
 when the.M.NOM.PL dead.M.NOM.PL hear.FUT.IND.ACT.3PL the.F.GEN.SG
phônês [toû huiou [toû theou]]]
 voice.F.GEN.SG the.M.GEN.SG son.M.GEN.SG the.M.GEN.SG god.M.GEN.SG
 ‘When the dead hear the voice of the Son of God’

(Koine Greek. Jh. 5.25)

- b. *egda mr̥tvii uslyšet̥ [gls̥ [sna*
 when dead.M.NOM.PL hear.PRS.3PL voice.M.ACC.SG son.M.GEN.SG
b̥ziě]
 God’s.M.GEN.SG
 ‘When the dead hear the voice of the Son of God’

(OCS. *C. Mar.* Jh. 25)

- (119) a. *eíselthe eis [tò tameíon sou]*
 enter.AOR.IMP.ACT.2SG into the.N.ACC.SG room.N.ACC.SG you.GEN.SG
kai kleísas [tén thúran sou]
 and shut.AORP.ACT.M.NOM.SG the.F.ACC.SG door.F.ACC.SG you.GEN.SG
próseuxai [tôi patrí sou]
 pray.AOR.IMP.MED.2SG the.M.DAT.SG father.M.DAT.SG you.GEN.SG
 ‘Enter into your room and having closed the door, pray to your father’

(Koine Greek. Mt. 6.6)

- b. *v̥nidi v̥ [klět̥ tvojo] i zatvori*
 enter.IMP.2SG in room.F.ACC.SG your.F.ACC.SG and close.IMP.2SG
[dv̥ri tvoje] pomoli se [otcu
 door.F.ACC.PL your.F.ACC.PL pray.IMP.2SG self.ACC.SG father.M.DAT.SG
tvoemu]
 your.M.DAT.SG
 ‘Enter into your room and having closed the door, pray to your father’

4.4. SLAVIC AND CELTIC NOUN PHRASES

(OCS. *C. Mar.* Mt. 6.6)

- c. [hoi] dè [soi mathētai] ou
 the.M.NOM.PL but your.M.NOM.PL disciple.M.NOM.PL not
nēsteúousin
 fast.PRS.IND.ACT.3PL
 ‘But your disciples do not fast?’

(Koine Greek. Mk. 2.18)

- d. *a* [tvoi učenici] *ne postęť* *sę*
 but your.M.NOM.PL disciple.M.NOM.PL not fast.PRS.3PL self.ACC.SG
 ‘But your disciples do not fast?’

(OCS. *C. Mar.* Mk. 2.18)

- (120) a. [anthrōpós tis] *eikhen* [dúo
 man.M.NOM.SG something.NOM.SG have.AOR.IND.ACT.3SG two.ACC.PL
huiouís]
 son.M.ACC.PL
 ‘A man had two sons’

(Koine Greek. *C. Mar.* Lk. 15.11)

- b. [člvkъ edinъ] imě [dvva
 person.M.NOM.SG one.M.ACC.SG have.AOR.3SG two.M.ACC.DU
sna]
 son.M.ACC.DU
 ‘A man had two sons’

(OCS *C. Mar.* Lk. 15.11)

- (121) a. *kai idou* [polloi telōnai *kai hamartōloi*]
 and behold many.M.NOM.PL tax-collector.M.NOM.PL and sinner.M.NOM.PL
 ‘And behold, many tax-collectors and sinners...’

(Koine Greek. Mt. 9.10)

4.4. SLAVIC AND CELTIC NOUN PHRASES

- b. *i se [mъnodzi grěšъnici i mytare]*
and behold many.M.NOM.PL sinner.M.NOM.PL and tax-collector.M.NOM.PL
'And behold, many tax-collectors and sinners...'

(OCS. *C. Mar.* Mt. 9.10)

There are certainly differences in adnominal modifiers between the two languages, but most operate at the level of the individual word, such as the use of a specific adjective to translate a Greek genitive and the case of the verb's direct object (genitive vs. accusative) in (118), or the use of the numeral *edinъ* 'one' to translate the Greek indefinite determiner *tis*. Word-order deviations are attested too, but the relative positions of (accented) elements in the OCS nominal are once again unlikely to reveal much that is straightforwardly ascribable to the native Slavic syntax; the motivations behind the arrangements in the Biblical examples (118)-(121) lie in Greek. However, while they do illustrate that the attested OCS noun phrase is in thrall to Greek prototypes, we should invoke the principle of translation in §2.2.1: Greek could only determine what Slavic syntax allowed it to determine. The fact that OCS attests in (116) and (117) both pre-nominal and post-nominal adjectives suggests that these orders would not be unintelligible or badly received by their audience. This gives the impression that the OCS noun phrase was syntactically flexible and had the capacity for self-modification. By positing a component for the native syntax like EmphasisP, we can derive this flexibility. This is highly speculative, but could be strengthened through comparative work on Slavic in general.

In sum, the Old Church Slavonic evidence offers glimpses of a native Slavic noun-phrase syntax, which included both definite determiners and a degree of word-order flexibility.

4.4. SLAVIC AND CELTIC NOUN PHRASES

These together created post-posed quasi-articles out of demonstratives. All in all, the direct OCS evidence again supports an underlying syntax with at least a DP, while there is indirect evidence for the EmphasisP proposed for the syntax of Latin, Sanskrit, Ancient Greek in (77), and for early Germanic in (107)/(108).

4.4.2 THE OLD IRISH NOUN PHRASE

The final language to be discussed is Old Irish, representing the Celtic language family. As in other early Indo-European languages, the Old Irish noun displays the three features of gender, number and case, which it shares through agreement with the other constituents of the noun phrase. How those features are expressed though is quite distinct; an earlier system of inflection endings has transformed into a combination of endings, root-vowel alternations and mutations, which modify the word that follows. These mutations work within the boundaries of syntactic constituency; a noun may mutate a dependent adjective, and may itself be mutated by a preceding and governing determiner, number or preposition. Compared with its Indo-European sister languages, Old Irish is also considerably rigid in its noun-phrase arrangements. Within these restrictions, we can diagnose a hierarchical syntax that includes a high DP layer, as well as a NumP between DP and the lower lexical structure.

The evidence for this syntax comes from functional constituents, namely possessive pronouns, numbers, the definite article and quantifiers. The definite article *in* always appears before a head noun, adjectives, genitives and quantifiers like *uile*.

4.4. SLAVIC AND CELTIC NOUN PHRASES

- (122) [*int uile comthirchomracc sanctorum*] ...
 the.M.NOM.SG whole.M.NOM.SG congregation.M.NOM.SG holy.M.GEN.PL
 [*in chatlach [inna fer]*] ...
 the.F.NOM.SG totality.F.NOM.SG the.M.GEN.PL man.M.GEN.PL
 [*int oentu [inna æccalsa]*]
 the.M.NOM.SG unity.M.NOM.SG the.F.GEN.SG church.F.GEN.SG
 ‘The whole congregation of saints ... the totality of the men ... the unity of the
 church’
 (Old Irish. Wb. 7c8)

The article is not used with vocative noun phrases, despite their definiteness (Stifter, 2009, 71), which typically appear instead with the leniting vocative particle *a*.

Similarly, possessive pronouns always precede their possessum and are unstressed and proclitic; they usually fuse with a preceding preposition. To compensate for their prosodic weakness, their person feature can be emphasised with the addition of one of the emphatic enclitic pronouns, the *notae augentes*, to the end of the noun phrase, as in (123a).

- (123) a. [*mo gligernat*]=*sa* ... [*mo ben*]=*sa*
 my Glycerium.F.NOM.SG-I my woman.F.NOM.SG-I
 ‘My Glycerium ... my wife’
 (Old Irish. Sg. 61b14)
- b. *do chuil*
 your corner.F.ACC.SG
 ‘Your corner’
 (Old Irish. Sg. 229a)
- c. *do*[=*m thoschid*]
 for-my need.F.DAT.SG

4.4. SLAVIC AND CELTIC NOUN PHRASES

‘For my sustenance’

(Old Irish. Wb. 10d27)

- d. *gaibid* [*comarbus* [*for n-athar*]] *et*
 take.IMP.2PL heritage.M.ACC.SG your father.M.GEN.SG and
intamlid [*a béssu*]
 imitate.IMP.2PL his custom.M.ACC.PL
 ‘Take up your father’s succession and imitate his manners’

(Old Irish. Wb. 9a14)

These possessive pronouns derive from old genitive pronouns, hence their lack of inflection and agreement with the noun (Thurneysen, 1946, 276). If so, then it seems that genitives were once more flexible in their position with the noun phrase, able to precede and follow the noun. Possessive pronouns also precede other pre-nominal elements, such as compounded adjectives and more functional elements like quantifiers, as in (124). It is at present unknown whether they also precede numbers.

- (124) a. *tre* [*a n-uili aimsera*]
 through their all.F.ACC.PL time.F.ACC.PL
 ‘Through all their tenses’

(Old Irish Sg. 157b9)

- b. *is* *hé* [*in peccad*] *ro·géní*
 be.PRS.IND.3SG he the.M.NOM.SG sin.M.NOM.SG make.PRF.ACT.3SG.REL
 [*a n-uile comaccobor*]
 their all.N.ACC.SG concupiscence.N.ACC.SG
 ‘It is sin which has made their whole carnal desire’

(Old Irish. Wb. 3c25)

4.4. SLAVIC AND CELTIC NOUN PHRASES

Possessive pronouns are in complementary distribution with the proclitic definite article *in* (Thurneysen, 1946, 298). We can therefore unify possessive pronouns and the definite article, considering them as equivalent markers of definiteness (Thurneysen, 1946, 295-6).²⁸ We can syntactically analyse both as lexical exponents of a high D head, above the positions of the noun and its lexical modifiers.

It is also higher than a NumP for numbers. In NumP are base-generated both cardinal and ordinal numbers; these consistently stand pre-nominally in the surface order (125), even though ordinal numbers inflect to agree like adjectives.²⁹

- (125) a. [*inna* *teoir* *ranna* *sa*] *is*
the.F.NOM.PL three.F.NOM.PL part.F.NOM.PL this be.PRS.IND.3SG
[*óinrann* *far·dingrat*]
one-part.F.NOM.SG signify.PRS.IND.ACT.3PL.REL
‘These three parts, it is one part that they express’
(Old Irish. Sg. 26b13)
- b. *etir* [*inna* *cethri* *fersu*]
between the.M.ACC.PL four.M.ACC.PL verse.M.ACC.PL
‘Between the four verses’
(Old Irish. Ml. 58a11)
- c. *i* [= *sin* *chétné* *tuiste*]
in the.F.DAT.SG first.F.DAT.SG creation.F.DAT.SG
‘In the first creation’

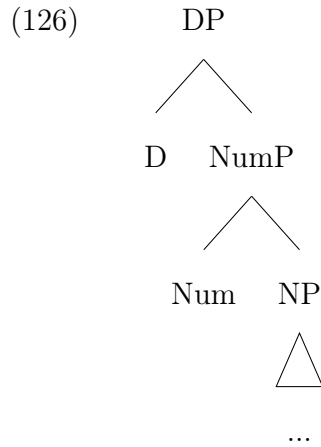
²⁸Although the functions of *in* extend to non-anaphoric definite and specific indefinite contexts (Goldstein, 2022).

²⁹Two exceptions are *tánaise* ‘second’ and *toísech*, which can mean ‘first’. These follow the noun, likely a product of their etymology; they do not derive from other numerals, but rather from two verbal roots (Greene, 1992, 17-8) and therefore occupy the regular position of adjectives.

4.4. SLAVIC AND CELTIC NOUN PHRASES

(Old Irish. Wb 1a1)

For the syntax then of the Old Irish noun phrase, we can propose the following structure:



Note that “NP” in (126) again is shorthand for the lower configuration of lexical modifiers. In this, the head noun must occupy a higher position than adnominal adjectives and genitives, since as a rule, adjectives follow the noun that they modify.

- (127) a. *hi [=sind libur romanach]*
in the.M.DAT.SG book.M.DAT.SG Roman.M.DAT.SG
‘In the Roman book’
(Old Irish. Sg. 4a12)
- b. *a geinti hireschu!*
oh gentile.M.VOC.PL faithless.M.VOC.PL
‘Oh faithless gentiles!’
(Old Irish. Wb. 3a13)

4.4. SLAVIC AND CELTIC NOUN PHRASES

- c. *is* *i* [*corp* *spirtálda* [*ind*
be.PRS.IND.3SG in body.M.ACC.SG spiritual.M.ACC.SG the.N.GEN.SG
eséirgi]] ... *do* [= *n* *chorp*
resurrection.N.GEN.SG to the.M.DAT.SG body.M.DAT.SG
marbdu]
mortal.M.DAT.SG
‘It is into a spiritual body of the resurrection ... to a mortal body’
(Old Irish. Wb. 3a14)

A pre-nominal position is also possible, but restricted. A handful of adjectives may precede the noun (some optionally, others obligatorily), in which case they lose their inflectional endings and form a compound with the noun (Stifter, 2009, 22).

- (128) a. *bésu* [*dagduine*]
perhaps-be good-man.M.NOM.SG
‘He may be a good man’
(Old Irish. Wb. 6b23)
- b. [*memmbrum* *naue*], [*drochdub*]
parchment.N.NOM.SG new.N.NOM.SG bad-ink.N.NOM.SG
‘New parchment, bad ink’
(Old Irish. Sg. 217a)
- c. [*senchassa* *rechta*] *ad·fiadat*
old-tales.M.NOM.PL law.M.GEN.SG declare.PRS.IND.ACT.3PL.REL
[*sáibapstil*]
false-apostle.M.NOM.PL
‘Old tales of the Law which false apostles tell’
(Old Irish. Wb. 28c23)

The adjectives that appear thus compounded are especially common words, expressing

4.4. SLAVIC AND CELTIC NOUN PHRASES

'basic' vocabulary, such as *dag* 'good', *droch* 'bad', *sen* 'old' and *óc* 'young'. There is little reason to deny that the default position for the adjective is post-nominal; pre-nominal adjectives are few and form a closed set.³⁰ Like adjectives, genitives consistently follow the head noun in the prose language of the glosses, as in (127c), although they sometimes precede in *retoiric*, "a stylized language found in poetry and early laws" (Stifter, 2009, 51), which may be another syntactic archaism.

At a pre-Old Irish stage, possessive pronouns were still genitive pronouns and the definite article was the demonstrative element **sindo-* (Thurneysen, 1946, 293). If we wish to consolidate these pronouns with nominal genitives (that is, having once had the same basic syntactic position), we can claim that the possessive pronouns once moved to Spec,DP, before later undergoing Spec-to-Head syntactic reanalysis (van Gelderen, 2004), accompanied by a reduction in their morphophonology.

A possible issue for the syntax in (126) is the use of demonstrative particles. As a proclitic article, *in* can no longer convey the meanings or force of a demonstrative. Consequently, additional indeclinable elements are added to contribute both emphasis and deictic direction (Thurneysen, 1946, 299-300). These include proximal *so*, medial and anaphoric *sin* and distal *tall* and *ucut*. Most notably, unlike other functional constituents, these attach themselves to the end of the head noun, which itself must be preceded by the

³⁰It is interesting though that the specific adjectives that may be pre-nominal and compounded are reminiscent of pre-nominal adjectives in Romance languages; for example, the French equivalents of four above (*bon*, *mauvais*, *vieux* and *jeune*) likewise usually precede their head noun. It could be interesting to explore whether compounded adjectives in Old Irish are relics of an older grammar, one with a hierarchy of adjectival projections, in the manner of Cinque (1994), within which *dag* and *droch* were once the product of a high projection like Cinque's "speaker-oriented AP".

4.4. SLAVIC AND CELTIC NOUN PHRASES

definite article. If a dependent adjective or genitive is present within the noun phrase, the particle follows that modifier too and remains last (130a).

- (129) a. *in* *testimin* *so*
the.M.NOM.SG text.M.NOM.SG this
‘This text’
(Old Irish. Sg. 4a12)
- b. *i* [= *snaib* *consonaib* *ucut*]
in the.F.DAT.PL consonant.F.DAT.PL that
‘In those consonants’
(Old Irish. Sg. 9a22)
- (130) a. *for* [= *sa* *cenélae* *metir* *sin*]
on the-n.acc.sg kind-n.acc.sg metre.M.GEN.SG that
‘In that kind of metre’
(Old Irish. Sg. 8a13)
- b. *do* [= *n* *dlúim* *máir* *sin* [*inna*
to the.F.DAT.SG mass.F.DAT.SG great.F.DAT.SG that the.M.GEN.PL
pecthach]]
sinner.M.GEN.PL ‘To that great mass of sinners’
(Old Irish. Wb. 9d5)

An exception to the phrase-finality of these elements is when the dependent genitive phrase is itself definite and includes the definite article; in cases like (130b), the demonstrative precedes the genitive. This exception to the phrase-final position of the demonstrative is reasonable from a semantic point of view; if it also follows the definite genitive phrase, it is likely to be interpreted as a member of that dependent phrase instead.

4.4. SLAVIC AND CELTIC NOUN PHRASES

These phrase-final elements do not pose a problem for the syntax in (126) though; we can integrate them by analysing them as low adverbial adjuncts to the noun. Such a status for Modern Irish demonstratives is the view of Kane (2013), who follows Bernstein (1997) and her treatment of demonstrative “reinforcers” in Germanic and Romance languages, such as French *ci* and *là*. Under this view, Old Irish *so*, *sin*, *ucut* and the rest are “post-nominal locative (place denoting)” elements (Kane, 2015, 148), recruited into the noun phrase as adjuncts to reinforce the definite article, which provides the feature of definiteness. The many elements used in this role support this analysis; these locative particles (mostly) do not inflect, and “in principle, any adverb of place can be added after article + noun for local qualification” (Stifter, 2009, 27). In terms of their syntax, they stand with adjectives and genitives within the low NP structure.

However, the locative particles themselves have been reduced to enclitics; thus, to emphasise a locative particle, the stressed particle *í* is used before it.

- (131) a. *in* *dóini* ***hi*** *siu*
the.M.NOM.PL person.M.NOM.PL PRT this
‘These people’

(Old Irish. Wb. 12d16)
- b. *for* [= *sin* *persin* ***í*** *sin*]
on the.F.DAT.SG person.F.DAT.SG PRT that
‘In that person’

(Old Irish. Sg. 197a11)

The particle *í* can additionally form complex pronouns with either the definite article

4.4. SLAVIC AND CELTIC NOUN PHRASES

alone or with the article and a locative particle (Thurneysen, 1946, 301) (132a). It is clearly stressed, since it can host the two clitic elements on either side. The combination of the article and *í* is often used as an antecedent for a relative clause (132b).

- (132) a. *cenmithá* [*a* *n-í* *siu*]
besides the.N.ACC.SG PRT this
'Except for this'

(Old Irish. Wb. 24a18)
- b. *ind* *í* *crettes*
the.M.GEN.SG PRT believe.PRS.IND.ACT.3SG.REL
'Of the one who believes'

(Old Irish. Wb. 2b19)

Although its etymology is not clear, it is best to consider *í* as some kind of pronoun or light noun in its own right, which can form its own noun phrase with the definite article. In orders like (131), *í* and the preceding noun are in fact syntactically independent of each other. Constructions like (131b) consequently contain two phrases in apposition; the second comprises only *í* and the locative element *sin*, and the whole construction is akin to English 'in the person, that one'. This apposition of phrases, such as with *í* or the aforementioned *notae augentes*, is clearly a productive means in Old Irish of emphasising constituents that cannot be emphasised themselves. Under this view, constructions with *í* do not pose difficulties for the syntax proposed in (126).

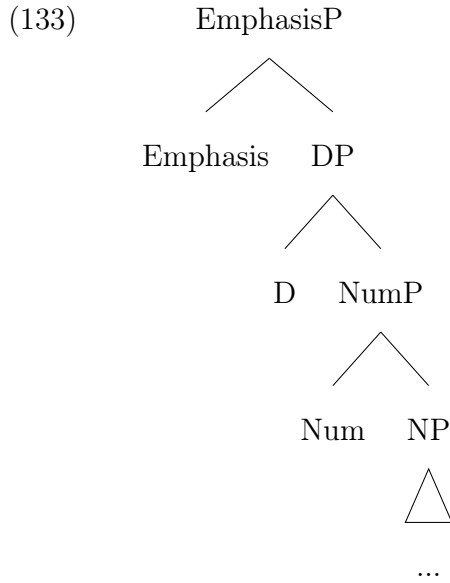
All things considered, the Old Irish noun phrase is a syntactically inflexible domain, with little optionality. Its restrictive word order can be easily derived through a syntax

4.5. CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

with at least a DP, a NumP, and a position of the head noun higher than adjectives, genitives and adjunct locative particles. While there traces of movement in the prehistory of Old Irish (such as with possessive pronouns' movement into the DP), such flexibility has disappeared by the stage of Irish of our prose sources.

4.5 CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

Across the seven members of the language set, striking similarities have been found in the behaviour of their noun phrases, and in the abstract structures that respectively generate that behaviour. On the basis of frequency, antiquity and the principles in Chapter 1, these commonalities substantiate the following reconstruction for the abstract syntax of the Proto-Indo-European noun phrase:



(133) is not committed to any configuration of the lower “NP”, which contains the head noun, adjectives, genitives and other adjunct constituents; for one reason, it is outside the ‘left periphery’ and therefore the scope of the thesis. For another, the comparative evidence does not readily indicate one particular ancestral arrangement of these constituents; pre- and post-nominal adjectives and genitives are well attested. To reconstruct their basic order, they would certainly require their own dedicated project, and even then the NP would in all likelihood remain opaque.

Across all seven languages, there is consistent evidence for the DP layer in (133), in which determiners are base-generated as D heads. In some of the languages, there is a definite article; for others, there are demonstratives, of which definiteness is one inherent feature. Furthermore, the DP offers a structural position for other lexical material that

4.5. CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

make their noun phrase definite, such as certain genitive phrases and possessive pronouns; these we can derive through movement from a lower genitive position up to Spec,DP. All in all, for definite noun phrases, the evidence satisfies both the majority rule and the elder rule of reconstruction, and so we can claim that DP was a part of Proto-Indo-European syntax too.

As with the CP in Chapter 2, by reconstructing a DP for the proto-syntax, we gain a common origin for the attested behaviour of definite determiners, and a starting point for the uniform grammaticalisation across early Indo-European of different demonstratives into pre-nominal definite articles. For the proto-language, we can reconstruct at least two definite determiners of the category D, which served ostensive and anaphoric functions: **só* ‘this, that’ and **éy* ‘that’ (Ringe, 2006, 54-7). As D heads within a left-headed syntax, they stood higher than the head noun, adjectives and numbers, which produced a basic determiner-initial linear order.

The picture becomes murky though when we consider indefinite noun phrases. Clearly a bare (determiner-less) noun phrase was typical for indefinite referents; not one of the branches of Indo-European considered has developed an indefinite article by the time of our historical sources. Although we can reconstruct an indefinite function for the stem **k^w*- (Beekes and de Vaan, 2011, 227-31), the use of this element as either an indefinite or interrogative determiner within a noun phrase is scarcely attested across early Indo-European, as mentioned in §2.3.1 and §2.3.5.³¹ It is therefore difficult to reconstruct it

³¹There is an established idea that the stem **k^w*- had two derived forms: an *i*-stem **k^wi*- and an *o*-stem **k^wo*-. The traditional rule, formulated on the basis of Latin, is that the *i*-stem formed pronouns, while

4.5. CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

as an indefinite determiner for Proto-Indo-European. This is to say, we have no secure indefinite lexical items that could be indefinite counterparts to **só* and **éy*. For this reason, it is reasonable to ask whether a D head and DP structure are only reconstructable for definite noun phrases with a lexical determiner.

Paraphrasing Norris (2018, 3) on Estonian, a noun phrase without an article or determiner is not equivalent to a noun phrase without a DP. Many generalisations and diagnostics for a DP component, or absence thereof, have been suggested, notably by Bošković (2008, 2012); these diagnostics include left-branch extraction, adjunct extraction, word-order scrambling, multiple *wh*-fronting and multiple adnominal genitives. These might allow us to diagnose a DP structure in indefinite noun phrases in early Indo-European (and from there in Proto-Indo-European), but the problem is that many of them rely on intuitions of grammaticality, for which we lack the native speakers. Others can be tested; Latin is a language without any articles, and its noun phrases are only NPs according to Bošković (2012, 7). The evidence for this comes partly from a ban on two lexical genitives within one noun phrase, although this claim does not seem to be borne out by the available data (cf. Pinkster, 2015, 1040). All this is to say that we lack secure lexical and syntactic evidence means to claim with confidence that the Proto-Indo-European indefinite noun phrase was also a DP in its syntax.

Our response is therefore likely to be motivated by theory-internal reasons. It may de-

the *o*-stem had an “adnominal function” (Ringe, 2006, 56). This would offer syntactic and morphological support to the view that **k^wo-* could function as an interrogative/indefinite determiner in a noun phrase, if only this distribution of the two stems were robust; Ram-Prasad (2022, 60-1) demonstrates that **k^wo-* can be both adnominal and pronominal, with exceptions to the rule in both Latin and Avestan.

pend heavily on our assumptions about the stability and optionality of syntactic structures: whether they remain the same in all contexts, regardless of whether their components are lexically realised or not. We could claim, if only for the sake of uniformity, that all Proto-Indo-European noun phrases were underlyingly DPs, and that indefinite noun phrases had a lexically null D. This is an established analysis for Modern English, in which the syntax of indefinite and article-less plural noun phrases (e.g. *good cats are playful*) has a null D head, and so is structurally the same as noun phrases with articles (cf. Le Bruyn et al., 2017). One potential lexical expression though for a Proto-Indo-European indefinite D is the aforementioned indefinite pronoun. Such a lexical item, built on the stem $*k^w-$, can be reconstructed for the proto-language. It conveyed an interrogative or an indefinite sense, depending on its accent. In the view of Abney (1987, 181-2) and later Norris (2018, 21-7), such elements are in fact full DPs. Although the evidence is lacking for the use of $*k^w-$ as a determiner, it does offer lexical evidence for an indefinite element of the category D.

Additionally, we might claim that all Proto-Indo-European noun phrases were DPs because the function of D is greater than simply marking definiteness; the view of Borer (2005, 64-7) is that noun phrases can only be arguments of a verb if they are syntactically DPs. Since Proto-Indo-European undoubtedly combined verbs with nominal arguments, Borer's semantic and syntactic arguments could be applied to the proto-language too. This line of argument is theory-internally motivated, but it need not remain so. There is possible evidence that a kind of non-argumental noun phrases, namely vocatives, had a reduced syntactic structure in Proto-Indo-European.

For Longobardi (1994), vocative phrases in Italian are not arguments of the verb and

4.5. CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

are not DPs. This he supports with the observation that vocative singular common nouns do not require the article that their non-vocative counterparts do need (e.g. *caro amico!* ‘dear friend!’ vs. *il caro amico* ‘the dear friend’). As noted in their respective sections, Ancient Greek and Old Irish likewise do not use their definite articles with vocatives, despite the implicit definiteness of their addressee. Old English does display the definite element *se* when its vocative noun phrases contain an adjective, but this is a later (intra-Northwest-Germanic) innovation, motivated by the syntactic requirements of the adjectival weak declension.

Furthermore, there is morphological evidence that early Indo-European vocatives have a lesser complexity, and perhaps a reduced syntactic structure too. Across early Indo-European, vocative singular forms are marked by the absence of a final suffix.³² We can reconstruct this for Proto-Indo-European; while the nominative singular animate was expressed through the suffix *-s, its vocative counterpart was the bare stem (Sihler, 1995, 250). We therefore have comparative evidence from word order and morphology for the view that the early Indo-European vocative noun phrase was structurally less complex than its argumental equivalents. On these grounds, we further have a means to claim tentatively that all early Indo-European noun phrases that were part of the propositional content of a clause were syntactically complex and were DPs, regardless of whether they were definite or indefinite, and that this common syntactic structure can be reconstructed back to the proto-language. Therefore, (133) applied to all (non-vocative) noun phrases

³²For example, the nominative~vocative pairs: Sanskrit *deváh~déva* ‘god’, Ancient Greek *Zeús~Zeû* ‘Zeus’, Latin *dominus~domine* ‘master’, which shows ablaut variation of the final stem vowel.

4.5. CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

in Proto-Indo-European. If nothing else, this evidence from vocatives opens up interesting avenues to explore the interaction between syntax, nominal morphology and the DP in early Indo-European languages.

Returning now to the reconstruction in (133), the noun-phrase left periphery across the language set is not derivable through positing a DP alone. (133) also includes a NumP, to account for the consistent post-determiner but pre-lexical position of numbers across early Indo-European. Furthermore, to account for orders in which the determiner is not initial in its phrase, this thesis proposes an EmphasisP for the synchronic syntax of Latin and Ancient Greek, and tentatively also for Sanskrit, Old English, Old Norse and Old Church Slavonic. By reconstructing EmphasisP, we gain a syntactic and diachronic reason for the considerable word-order variation especially characteristic of the three earlier Indo-European languages. The label of ‘Emphasis’ for its lexically null head is deliberately non-committal, trying to consolidate the descriptions in previous literature of what fronting within the noun phrase achieves. It avoids specific characteristics like ‘Topic’ or ‘Focus’, for which there seems to be little or no firm support. Instead, ‘EmphasisP’ offers constituents of the noun phrase a syntactically marked position, which may or may not have been accompanied with emphatic prosody. This position was connected to the wider discourse context of the noun phrase. In Spec,EmphasisP, a moved constituent could refer back to a known topic, or contrast with something else.³³

Lexico-phonological reconstruction has not identified any lexical items that could be the exponents of the Emphasis head, leaving us to conclude that Emphasis was lexically

³³‘ProminenceP’ or ‘ContrastP’ would be appropriate alternative names.

4.5. CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

null. It made its presence felt though through movement; the evidence for EmphasisP comes from various fronted lexical items, namely nouns, adjectives and genitives, that precede a definite determiner in Latin, Ancient Greek and Sanskrit. These are constituents considered salient in the view of the speaker, linking them to entities beyond the borders of their noun phrase. Moreover, there are instances of fronting the same elements to precede numbers, as in Old Norse; both the unmarked and marked order can be derived if we adopt a syntax with their movement past a high NumP and into an even higher EmphasisP. Furthermore, there is evidence in Classical Greek, pre-Old Norse and pre-Old Church Slavonic of movement to EmphasisP as a way to ‘weaken’ demonstratives. In the absence of definite articles, these languages exploited the word-order variation offered by EmphasisP to produce orders with the demonstrative later or last in the noun phrase. This created a contrast between initial strong demonstratives, which were ostensive towards to objects in the immediate environment, and weak demonstratives, which were anaphoric and referred back to or anticipated entities in the discourse.

Further evidence for a syntactic structure beyond DP might come from quantifiers. These have a consistently pre-determiner position across early Indo-European languages. Having proposed a proto-DP from the common behaviour of determiners, it may seem only right to reconstruct a higher proto-QP too. However, the evidence for this reconstruction would be weaker. In Latin and Ancient Greek, and marginally in Sanskrit and Old Norse, we find that quantifiers (e.g. Ancient Greek *pâs* ‘all’) can stand later in the surface order, and function as attributes of the head noun. For these respective languages, it has been proposed that the initial quantifiers are in fact base-generated low in the structure, and

4.5. CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

then undergo movement higher.³⁴ Where they then stand is unclear. Their apparent complementary distribution with fronted constituents suggests that they target EmphasisP, but the DP seems more appropriate; like determiners, quantifiers help to specify the reference of the overall noun phrase. This can remain unresolved for the individual languages and for the proto-syntax. What matters is that the comparative evidence from quantifiers does not require us to reconstruct for an additional functional projection above DP. (133) can accommodate quantifiers too.

Depending on what lexical items are used, the reconstructed syntax in (133) would have generated linear sequences with the following orders:

- Noun
- Demonstrative Noun
- Number Noun
- Demonstrative Number Noun
- Noun Demonstrative (with noun fronting)
- Noun Demonstrative Number (with noun fronting)
- Number Demonstrative Noun (with number fronting)

³⁴The arguments from morphology given in Latin, Sanskrit and Ancient Greek may pertain also to Proto-Indo-European; lexico-phonological reconstruction posits a distinct declension for its demonstratives, e.g. **só*. What quantifiers we may be able to reconstruct for the proto-language seem to be thematic and presumably adjectival in declension, e.g. **wiso-* ‘all’ (cf. Matasovic, 2009, 540)

4.5. CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

- etc.

For example, following the lexical reconstructions of Ringe (2006), a demonstrative (e.g. **só* ‘that’), a number (e.g. **pénk^we* ‘five’) and a noun (e.g. **kwó* ‘dog’) would be arranged in Proto-Indo-European into an unmarked DemNumN order.

- (134) **tóy* *pénk^we kwónes*
 that.M.NOM.PL five dog.M.NOM.PL
 ‘Those five dogs’

(Proto-Indo-European. Artificial example)

To emphasize the dogs, perhaps to contrast them with **h₂óweyes* ‘sheep’, the syntax permitted the noun to be fronted, producing a NDemNum order.

- (135) **kwónes* *tóy* *pénk^we*
 dog.M.NOM.PL that.M.NOM.PL five
 ‘Those five *dogs*’

(Proto-Indo-European. Artificial example)

(133) can also produce orders with fronted, contrastive adjectives.

- (136) **[h₁rud^hróes* *tóy* *pénk^we kwónes]* [*krsnéh₂es* =*k^we*
 red.M.NOM.PL that.M.NOM.PL five dog.M.NOM.PL black.F.NOM.PL and
téh₂es *swéks h₂óweyes]*
 that.F.NOM.PL six sheep.F.NOM.PL
 ‘Those five *red* dogs and those six *black* sheep’

(Proto-Indo-European. Artificial example)

4.5. CONCLUSIONS: RECONSTRUCTING A PROTO-NOUN PHRASE

As a final note, (133) is also intentionally minimal, consolidating a wide range of word-order evidence to substantiate its two key components. The claim is specifically that the Proto-Indo-European noun was generated by this syntax *at least*; other scholars might wish to enlarge it. It is also intended as only a common origin and starting point for the different noun-phrase behaviours that we find in the later daughter languages. Over time, through grammaticalisation, it may have changed; one example is the diachrony, proposed in §4.2.3, that the DP split into two in between the Homeric and Classical eras of Greek, a consequence of the development of a definite article.

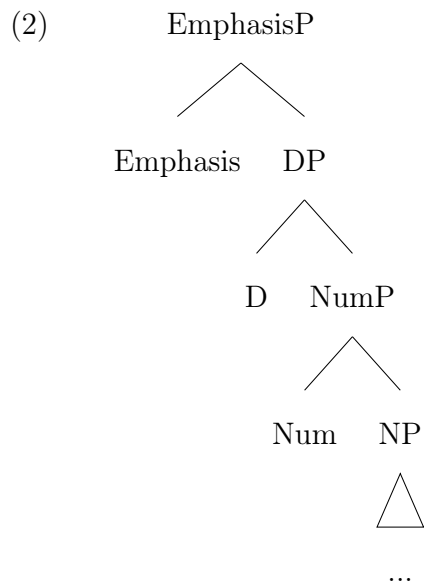
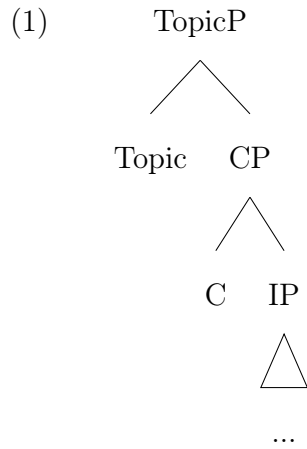
THESIS CONCLUSIONS

5.1 THE RECONSTRUCTED LEFT PERIPHERIES

The preceding three chapters have been an extensive exercise in syntactic reconstruction. On the basis of a wide range of comparative evidence, involving various groups of lexical items and sundry instances of syntactic movement, this thesis reconstructs the following two syntactic structures for the left periphery of the clause (1) and noun phrase (2) in Proto-Indo-European¹:

¹To be precise, late Proto-Indo-European or Proto-*Core*-Indo-European.

5.1. THE RECONSTRUCTED LEFT PERIPHERIES



The clausal syntax in (1) contains two key components: heads of the category C, which encodes clause type and status, and a Topic head, which hosts or triggers the movement of topical material in its specifier position. The noun phrase in (2) meanwhile is proposed to

5.1. THE RECONSTRUCTED LEFT PERIPHERIES

have a DP layer, which is responsible for the noun phrase's (in)definiteness, and an EmphasisP, which is a deliberately non-committal label for a syntactic component associated with fronting and discourse-salience within the noun phrase. These reconstructions are a point of origin for a broad array of documented word-order phenomena across early Indo-European languages, including the considerable variation in older languages like Sanskrit. This flexibility is the product of discourse-functional components, such as Topic, Focus and Emphasis, which modify an unclear fundamental configuration. Yet they also account for the most common word orders that we find, such as the observed typical pre-nominal and pre-adjectival position of definite determiners across Indo-European, and the usual clause-initial position of subordinators, preceded only by topics. These are understood to have been their unmarked positions in the proto-language. It also gives a source for the behaviour of enclitic pronouns as part of Wackernagel's law, since this is partly produced through syntactic movement to the C head.

However, (1) and (2) are intended to represent a minimum of syntactic configuration in Proto-Indo-European. There is evidence for more components, but these did not reach the same level of consistency to be reconstructed so cogently. One possible addition, discussed in §4.5, would be based on the initial position of quantifiers in Indo-European noun phrases. If we do not wish to derive these through movement to EmphasisP or DP, then the most obvious alternative is to reconstruct another functional projection (a 'QP') above DP.²

Being the ancestral proto-syntax, (1) and (2) are not understood to have been maintained down the generations unchanged. There is evidence for modifications in particular

²Alternatively, we could propose that the Proto-Indo-European DP was capable of self-iteration.

5.1. THE RECONSTRUCTED LEFT PERIPHERIES

daughter languages; it is suggested in §4.2.3 that the earlier DP ‘split’ in two with the development of a definite article. For some languages, philological complications obscure potential reflexes of the proto-syntax; there is little to no direct support for a TopicP in Old Church Slavonic, for instance, although the behaviour of the wider Slavic family seems to support a high-Topic syntax. The proposed reconstructions in (1) and (2) are intended to be diachronic starting points for a host of syntactic behaviour, not the final word on Indo-European syntax in general.

In contrast to previous reconstructions (notably Kiparsky’s (1995)), focus is not reconstructed as a feature of this high clausal syntax. The analysis of the language set in §3.3.2 is that the evidence for a high Focus component is negligible and insufficient for reconstruction, and that focal behaviour is better derived through a lower syntactic location.

A final, more general conclusion is that this thesis demonstrates that it is better to think of Proto-Indo-European as a language with configurational syntax. This is to say, we can explain so much of the attested data if we posit that the parent language was configurational, with syntactic constraints on its word order. Its syntax must have incorporated the capacity for modification and variation, understood primarily as the workings of discourse information, but this picture of the proto-language is surely better than the counterclaim that Proto-Indo-European was non-configurational, with no structure beyond the individual word. For one thing, we do not have to explain how configurationality comes to develop seemingly *ex nihilo* across the later Indo-European family.

5.2 DISCUSSION OF THE RECONSTRUCTIONS AND WIDER CONTEXT

One point to note is that there are parallels between the two domains of syntax reconstructed in (1) and (2). While the labels and key characteristics of the topmost two components differ, C corresponds to D, and Topic to Emphasis, in that their functions pertain to the overall domain. The function of C is to encode the type and status of its clause into the syntax. It takes the basic proposition formed through the constituents of the lower structures and relates it to the world, truth and the intentions of the clause's producer. For example, a C head may have the feature of [Declarative]; this encodes that a specific relationship between the producer and the proposition, namely that they believe it to be true; a C with the feature [Interrogative] instead encodes an absence of a belief in the truth of the proposition, and triggers word order associated with requests for information. Through this, C plays a part in the illocutionary force of a clause (Austin, 1975).

D meanwhile has a somewhat similar function, in that it specifies and instantiates the reference of the lower syntax of the noun phrase; "the noun provides a predicate and the determiner picks out a particular member of that predicate's extension" (Abney, 1987, 50). Clause type and definiteness are separate phenomena, and it is not claimed that the noun phrase has comparable 'types' or its own truth value or illocutionary force. Instead, the parallel is that C and D make their respective domains of syntax 'complete'. In taking an IP as its complement, C creates a proper, usable clause (cf. Collins, 2024, 139); a clause's

5.2. DISCUSSION OF THE RECONSTRUCTIONS AND WIDER CONTEXT

“ability to be true or false, and so to make a distinctively sentential move in the language game, supervenes on the fully formed structure” (Gaskin, 2020, 13). In similar way, D combines with the lexical constituents of the NP to create a fully referential noun phrase.

Moreover, Topic and Emphasis are alike, in that they look beyond the borders of their own domain, and help to situate it in its wider context. Topic offered the Proto-Indo-European clause a syntactic position for constituents that are somehow topical and link the new clause to the discourse, in anticipation of new information (part of which is the clausal focus). Constituents in EmphasisP do not seem to have shared this topical quality, but rather were either emphatic or contrastive with some other entity or quality. The comparative evidence therefore does not seem to allow clausal Topic and noun-phrase Emphasis to be reduced into one, but the picture has emerged of a high, left-peripheral syntax with two key components: one that looks inward, the other outward. That there are parallels between the two syntactic domains is noteworthy, but not surprising; the idea that they mirror each other is well established (cf. Abney (1987) on English and Szabolcsi (1987) on Hungarian).

A further similarity between C and D noted by Szabolcsi (1987, 180) is that they make their clause or noun phrase an acceptable argument of a verb; a verb can select a DP or a subordinate CP as its object.

- (3) a. *I see* [_{DP} *the man*]
b. *I see* [_{CP} *that the man is here*]

5.2. DISCUSSION OF THE RECONSTRUCTIONS AND WIDER CONTEXT

This could be claimed as a function of the reconstructed C in (1) as well, were it not for the tricky issue of subordination in Proto-Indo-European. This is a topic that this thesis has avoided, concentrating instead on the interior of the clause. In brief, there is a long-standing idea, supported by comparative evidence from Sanskrit, Hittite, Ancient Greek and Latin, that all finite clauses in the proto-language were syntactically independent (cf. Clackson, 2007, 171-6). Finite clauses were only associated through parataxis; it was out of this original situation that clausal embedding (hypotaxis) later developed. Since this thesis is not concerned with multi-clausal constructions, a position on this diachrony is not necessary. However, another observation of the language set is that this paratactic view of Proto-Indo-European need not reconstruct a language that is radically different in syntax from its descendants, capable of no utterances more complex than strings of simple clauses; the evidence shows that embedding is consistently attested with non-finite clauses. All seven languages exhibit non-finite verbal forms that can have their own arguments, and four (Latin, Ancient Greek, Sanskrit, Old Church Slavonic) display a productive ‘absolute’ construction, which could be reconstructed back to the proto-language (cf. Comrie 1998, 94-5; Ruppel 2012, 227). On the basis of non-finite morphology and syntax, we therefore have grounds to reconstruct embedded clauses for Proto-Indo-European, leaving the Uniformitarian Principle undisturbed (Roberts, 2007, 274).³

A second point is that it seems prudent to distinguish different kinds of subordination

³The syntactic picture of Proto-Indo-European that emerges from the literature and the language set is reminiscent of another Eurasian proto-language, Proto-Uralic, which likewise has been reconstructed with productive embedding of non-finite constructions, but not finite clauses (Janhunen, 1982, 39). Some modern Uralic languages, such as Mari, maintain this grammar (Saarinen, 2022, 466-8).

5.2. DISCUSSION OF THE RECONSTRUCTIONS AND WIDER CONTEXT

in the discussion of the early Indo-European clause. The argument that finite clauses originally had no connection stronger than adjunction⁴ is strong, but even if we accept this absence of embedding and syntactic subordination, Proto-Indo-European may have been capable of *lexical* subordination. This is to say, the behaviour of subordinators across the language set is so consistent that we can reconstruct a situation in which specific lexical items are marking their clause as subordinate to another, even if it can only stand at the edges of that superordinate clause. One such element was surely the reconstructed **yó-*stem⁵, which, as a marker of relativity, occupied the CP layer in (1).

On the subject of subordinators, an idea noted in the conclusions in §2.4 is that syntactic comparison and reconstruction may even succeed in areas where lexico-phonological reconstruction fails; it is true that there is little evidence for specific lexical subordinators in Proto-Indo-European, yet their comparative syntactic behaviour across early Indo-European may be the only trace of the existence of older lexical items now otherwise lost to us.

A related point made briefly in §3.2.4 is that the syntax in (1), reconstructed for all finite clauses, may have changed with the development of finite-clausal embedding. Assuming the traditional ‘parataxis to hypotaxis’ diachrony, a further claim is that TopicP was gradually lost in new syntactically subordinate clauses. While the original Topic-Subordinator order is common still in Sanskrit, a language arguably without finite-clausal embedding, the subordinator-second orders in (4a) and (4b) may represent rare archaisms in languages in

⁴As in the common correlative construction, discussed in §3.2.1

⁵The etymon of various subordinators, including Sanskrit *yá-* and Old Irish =Ø^{L/N}.

5.3. FURTHER WORK

(Halpern, 1992, 1995). This, done to satisfy prosodic constraints, is a key piece of the analysis of word order in Sanskrit, Ancient Greek and Old Irish.⁶ It is especially important in explaining second-position conjunctions and pronouns in clauses in which CP and TopicP are not lexically realised. In such cases, pronouns move to the C head as usual, but then trigger prosodic inversion with the next available host in the linear order. This inversion seems so common across the language set that it too is best explained as an inherited prosodic remedy for host-less enclitics.

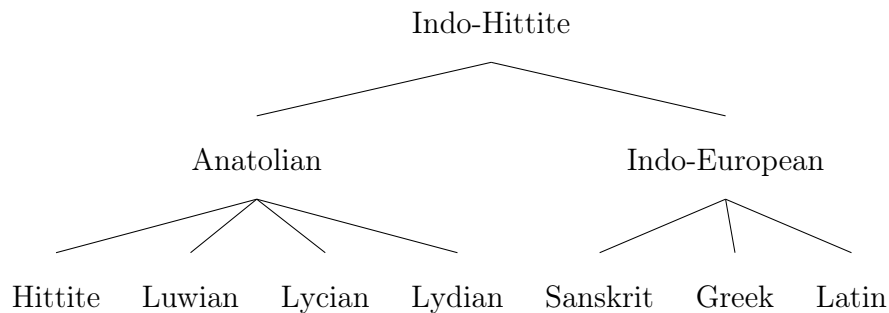
5.3 FURTHER WORK

This thesis opens up so many avenues for further research into Indo-European syntax. The most obvious is to compare and test the reconstructions in (1) and (2) against more Indo-European languages. With a bigger comparative set, the reconstructions can be corroborated, modified, improved or even rejected. The most important candidate for comparison is Hittite. At present, the seven languages considered can only lead us to their common ancestor, which is *not* the point of departure of the whole Indo-European family. This later ancestor has herein been referred to as Proto-Core-Indo-European, but an alternative name is simply Proto-Indo-European, in contrast with the older stage of Proto-Indo-*Hittite* (cf. Sturtevant, 1933) or Proto-Indo-*Anatolian* (cf. Kloekhorst and Pronk, 2019).⁷

⁶With additional support from Latin and Gothic, cf. §2.3.5.

⁷Meanwhile, for Ringe (2006, 5), the topmost node of the family tree is Proto-Indo-European, while what remains after the Anatolian split is “North Indo-European”. After parting ways with the speakers of

(5)



(From Sturtevant (1962, 109))

This relationship of the Anatolian languages to (the rest of) Indo-European means that differences in their syntax need not doom (1) and (2); they may instead offer evidence for innovations within either Proto-Anatolian or Proto-Core-Indo-European. Since Hittite does not, it has been claimed, show the same *wh*-movement⁸, it may be possible to build a case that the clause-typing C in (1) was a Proto-Core-Indo-European innovation. This case would consequently elaborate our understanding of the diachrony of the proto-syntax. Anatolian does seem to show the same Wackernagelian behaviour of enclitic pronouns, but the Tocharian languages do not (see Onishi, 2022). Both of these branches need to be studied in detail and integrated into the discussion, if we are to get back to the ultimate origins of this famous word order.

An interesting point, raised in a recent conversation about the conclusions of this thesis, is that it paints a picture of considerable syntactic stability. The issue of dating Proto-Indo-European is complex, not least because a proto-language is not the same thing as a future Proto-Tocharian, the next node down is “West Indo-European”.

⁸See the conclusions of §2.3.1.

5.3. FURTHER WORK

‘real’ language spoken in one place and one time (cf. Clackson, 2007, 16). One reckoning (Atkinson et al., 2005) dates the post-Anatolian stage of Proto-Core-Indo-European to between seven and six thousand years before the present day. If so, we have a time gap of at least two thousand years before the earliest documented member of the language set (Vedic Sanskrit) and five thousand before the latest (Old Norse). This prompts the question of whether the syntactic configurations in (1) and (2) could really have remained the same over those millennia. Yet there does not seem to be an obvious reason for why they could not have; syntax is regarded by many to be slow to change in general (cf. Roberts, 2007, 444-9), perhaps even inert (Longobardi, 2001a; Keenan, 2002).

“Inertia: Things stay as they are unless acted on by an outside force or decay.”

(Keenan, 2002, 327)

If the causes of change were simply not there, there is nothing in the syntax that forbids its endurance over the centuries and into the emerging branches of Indo-European. What is more interesting is why it then does seem to change in later languages. (1) and (2) are not intended to derive the word order of the descendant languages of the set, such as modern-day Germanic languages, for which Topic, C, Emphasis and D may not be necessary or sufficient analyses. The general impression is that there is a divide in the set between the ancient trio (Latin, Ancient Greek, Sanskrit) and three of the post-classical languages (Old English, Old Norse, Old Irish) in the arrangement of their clauses. ‘Syntacticization’ (Walkden, 2014, 91) and the rise of new syntactic constraints (such as Germanic V2 or Celtic VSO) attest to shifts away from an older configuration that had endured for

5.3. *FURTHER WORK*

centuries. How and why a discourse-driven, more flexible word order of, say, Latin becomes the syntactically constrained, less flexible word order of, say, Italian is a fascinating question. We can speculate about historical factors, arising from a more interconnected linguistic landscape in late antique and early medieval Europe, but perhaps these are simply unrelated changes across an ever-diversifying language family. What is certain is that these are matters for another project; this thesis has navigated its intended course, from the syntax of Proto-Indo-European to its early attested descendants.

APPENDICES

A

APPENDICES

A.1 ABBREVIATIONS

Abbreviation	Meaning	Abbreviation	Meaning
ABL	ablative	POSS	possessive
ACC	accusative	PPRF	pluperfect
ACT	active	PRF	perfect
AOR	aorist	PRS	present
CND	conditional	PRT	particle
CP	Complementizer Phrase	PST	past
DAT	dative	PV	preverb
DEF	definite	Q	interrogative particle
DEP	deponent	REL	relative
DP	Determiner Phrase	RES	resultative
DU	dual	SBJ	subjunctive
F	feminine	SG	singular
GEN	genitive	ST	strong
IMP	imperative	VN	verbal noun
IP	Inflectional Phrase	VOC	vocative
IPRF	imperfect	VP	Verb Phrase
IND	indicative	WK	weak
INF	infinitive		
INJ	injunctive		
INS	instrumental		
LOC	locative		
M	masculine		
MED	mediopassive		
N	neuter		
NOM	nominative		
NP	Noun Phrase		
OPT	optative		
PASS	passive		
-P	participle		
PL	plural		

A.2 SOURCES INFORMATION

This appendix provides information about the sources and resources used throughout this thesis, along with clickable links. Examples taken from specific pieces of work are acknowledged as such, yet the majority are taken instead from primary sources, which also owe their availability to the hard work of the scholars who have preserved, collated, digitised and parsed them. This thesis stands on their shoulders.

A.2.1 LATIN

The early Latin texts used as examples (Caesar, Cicero, Plautus, etc.) are mostly taken from the versions on the Perseus Digital Library, others from the Loeb Classical Library.

A.2.2 ANCIENT GREEK

Classical texts are likewise taken from the Perseus Digital Library. Biblical quotations in Koine Greek come from Bible Hub.

A.2.3 SANSKRIT

Being almost entirely taken from the language of the Rigveda, Sanskrit examples were gathered from the University of Cologne's VedaWeb.

A.2. SOURCES INFORMATION

A.2.4 OLD CHURCH SLAVONIC

The examples for early Slavic were taken from the TITUS database and the Syntacticus treebank, as well as a digitised version of the *Codex Zographensis*.

A.2.5 OLD ENGLISH

The resources for Old English primary sources were mixed; examples from the *Martyrology* come from the edition of Rauer (2013), examples from the Old English translation of Bede's *Historia ecclesiastica* come from Miller's (1890) edition.

A.2.6 OLD NORSE

The examples from the *Heimskringla* were drawn from the University of Oslo's Library of Old Norse, those from the *Prose Edda* from the publicly available edition of Guðni Jónsson.

A.2.7 OLD IRISH

Examples of the St Gall glosses come from the online version of Bauer et al. (2017); those of the Würzburg glosses from the the edition of Doyle (2018); and those of the Milan glosses from the database of Griffith (2013) and from TITUS.

REFERENCES

- Abney, S. P. (1987). The English Noun Phrase in its Sentential Aspect. PhD Thesis, MIT.
- Adams, J. (1976). A typological approach to Latin word order. Indogermanische Forschungen, 81:70–99.
- Adger, D. (2006). Post-Syntactic Movement and the Old Irish Verb. Natural Language & Linguistic Theory, 24(3):605–654.
- Agbayani, B. and Golston, C. (2010). Second-position is first-position: Wackernagel’s Law and the role of clausal conjunction. Indogermanische Forschungen, 115:1–21.
- Ahlqvist, A. (1985). The Relative Endings of the Old Irish Simple Verb. Ériu, 36:137–142.
- Alexiadou, A. (2002). Word order patterns in Greek nominals: aspects of diachronic change. ZAS Papers in Linguistics, 27:91–107.
- Alexiadou, A. (2014). Multiple determiners and the structure of DPs. Number 211 in *Linguistics Today*. John Benjamins Publishing Company, Amsterdam.
- Allan, R. J. (2012). Clause intertwining and word order in Ancient Greek. Journal of Greek Linguistics, 12(1):5–28.
- Allan, R. J. (2014). Changing the topic: topic position in Ancient Greek word order. Mnemosyne, 67(2):181–213.

REFERENCES

- Allen, C. (2007). Variation in the NP/DP in Old English: Determiner and possessive combinations. In Architectures, Rules, and Preferences; variations on themes by Joan W. Bresnan, pages 3–20. Center for the Study of Language and Information (CSLI), Stanford, CA.
- Allen, C. L. (2008). Genitives in Early English: Typology and Evidence. Oxford University Press, Oxford.
- Allen, J. H. and Greenough, J. B. (1903). Allen and Greenough's New Latin grammar for schools and colleges: Founded on comparative grammar. Ginn, Boston, MA.
- Andersen, H. (1973). Abductive and Deductive Change. Language, 49(4):765–793.
- Antinucci, F. and Cinque, G. (1977). Sull'ordine delle parole in italiano: l'emarginazione. Studi di grammatica italiana VI, pp. 121-146.
- Arad, M. and Roussou, A. (1997). Particles and C-positions in Classical Greek. Ms. UCL and Bangor.
- Atkinson, Q., Nicholls, G., Welch, D., and Gray, R. (2005). From Words to Dates: Water into Wine, Mathemagic or Phylogenetic Inference? Transactions of the Philological Society, 103(2):193–219.
- Austin, J. L. (1975). How to Do Things with Words. Harvard University Press, Cambridge, MA, second edition.
- Bailey, L. (2010). Latin word order and the EPP. Handout from CamLing, Cambridge University.
- Baker, M. C. (2008). The Macroparameter in a Microparametric World. In The Limits of Syntactic Variation, number 132 in Linguistics Today, pages 351–374. John Benjamins Publishing Company, Amsterdam.
- Bakker, S. J. (2009). The noun phrase in Ancient Greek: A functional analysis of the order and articulation of NP constituents in Herodotus. Brill, Leiden, Boston.
- Barnes, M. (2008). A New Introduction to Old Norse: Part I: Grammar. Viking Society for Northern Research, London, third edition.

REFERENCES

- Barðdal, J., Bjarnadóttir, V., Danesi, S., Dewey, T. K., Eythórsson, T., Fedriani, C., and Smitherman, T. (2013). The story of ‘woe’. Journal of Indo-European Studies, 41(3/4):321–377.
- Barðdal, J., Gildea, S., and Luján, E. R. (2020). Reconstructing Syntax. Brill, Leiden, Boston.
- Bate, D. L. (2024). Deriving the Old Irish Clause. Journal of Historical Syntax, 8(1):1–48.
- Bauer, B., Hofman, R., and Moran, P. (2017). St Gall Priscian Glosses, version 2.0. www.stgallpriscian.ie/resource.
- Bech, K. (1998). Pragmatic factors in language change: XVS and XSV clauses in Old and Middle English. Folia Linguistica Historica, 19:79–102.
- Bech, K. (2001). Word order patterns in Old and Middle English: A syntactic and pragmatic study. PhD Thesis, University of Bergen.
- Bech, K. (2017). Old English and Old Norwegian noun phrases with two attributive adjectives. Bergen Language and Linguistics Studies, 8(1):1–18.
- Beekes, R. (2009). Etymological Dictionary of Greek (2 vols.). Brill, Leiden, Boston.
- Beekes, R. S. and de Vaan, M. (2011). Comparative Indo-European Linguistics: An Introduction. John Benjamins Publishing Company, Amsterdam, second edition.
- Belletti, A. (2001). Inversion as Focalization. In Subject Inversion in Romance and the Theory of Universal Grammar, Oxford Studies in Comparative Syntax, pages 60–90. Oxford University Press, Oxford.
- Belletti, A. (2004). Aspects of the low IP area. The structure of CP and IP. The cartography of syntactic structures, 2:16–51.
- Belletti, A. and Shlonsky, U. (1995). The order of verbal complements: A comparative study. Natural Language & Linguistic Theory, 13(3):489–526.

REFERENCES

- Bennett, C. E. (1910). Early Latin Syntax, volume 1. Georg Olms Verlag, Hildesheim.
- Bernstein, J. B. (1997). Demonstratives and reinforcers in Romance and Germanic languages. Lingua, 102(2):87–113.
- Bertrand, N. and Faure, R. (2022). Wh-Interrogatives In Ancient Greek: Disentangling Focus- And Wh-Movement. Studia Linguistica, 76(3):735–771.
- Bianchi, V. (2003). On finiteness as logophoric anchoring. Temps Et Point De Vue/Tense and Point of View, pages 213–246.
- Bianchi, V. (2013). On ‘focus movement’ in Italian. Information Structure and Agreement, pages 193–216.
- Borer, H. (1984). Parametric syntax: case studies in Semitic and Romance languages. Foris Publications, Dordrecht.
- Borer, H. (2005). Structuring Sense: Volume I: In Name Only. Oxford University Press, Oxford.
- Borer, H. (2014). The Category of Roots. In The Syntax of Roots and the Roots of Syntax, Oxford Studies in Theoretical Linguistics, pages 112–148. Oxford University Press, Oxford.
- Bošković, Z. (2008). What will you have, DP or NP? In Proceedings of the North East Linguistic Society 37, volume 37, pages 101–114. GLSA Publications. Issue: 1.
- Bošković, Z. (2012). On NPs and clauses. In Discourse and Grammar: From Sentence Types to Lexical Categories, volume 112 of Studies in Generative Grammar, pages 179–242. De Gruyter Mouton, Berlin, New York.
- Breatnach, L. (1984). Canon law and secular law in early Ireland: the significance of Bretha Nemed. Peritia, 3:439–459.
- Bresnan, J. W. (1970). On Complementizers: Toward a Syntactic Theory of Complement Types. Foundations of language, 6(3):297–321.

REFERENCES

- Brunel, J. (1964). La construction de l'adjectif dans les groupes nominaux du grec. Presses universitaires de France, Paris.
- Cardinaletti, A. (2018). On different types of postverbal subjects in Italian. Italian Journal of Linguistics, 30(2):79–106.
- Cardona, G. (2017). The documentation of Indic. In Handbook of Comparative and Historical Indo-European Linguistics, volume 1, pages 309–325. De Gruyter Mouton, Berlin, New York.
- Carlson, A. M. (1978). A diachronic treatment of English quantifiers. Lingua, 46(4):295–328.
- Carlton, C. M. (1970). Descriptive Syntax of the Old English Charters. Number 111 in *Janua Linguarum*. Series Practica. De Gruyter, Berlin, Boston.
- Carnie, A., Pyatt, E., and Harley, H. (1994). The resurrection: raising to COMP? Some evidence from Old Irish. Studies in the Linguistic Sciences, 24(2):85–100.
- Cecchetto, C. and Oniga, R. (2002). Consequences of the analysis of Latin infinitival clauses for the theory of case and control. Lingue e Linguaggio, 1(1):151–189.
- Chomsky, N. (1970). Remarks on nominalization. In Jacobs, R. and Rosenbaum, P., editors, Readings in English Transformational Grammar, pages 184–221. Ginn, Waltham, MA.
- Chomsky, N. (1977). On Wh-Movement. Formal Syntax, pages 71–132.
- Chomsky, N. (1981). Lectures on Government and Binding. Foris Publications, Dordrecht.
- Chomsky, N. (1982). Some Concepts and Consequences of the Theory of Government and Binding. MIT press, Cambridge, MA.
- Chomsky, N. (1995). The Minimalist Program. MIT Press, Cambridge, MA.

REFERENCES

- Chomsky, N. (1999). Derivation by Phase. MIT Occasional Papers in Linguistics. MIT, Department of Linguistics, Cambridge, MA.
- Chomsky, N. and Lasnik, H. (1993). The Theory of Principles and Parameters. In Syntax: Ein internationales Handbuch zeitgenössischer Forschung / An International Handbook of Contemporary Research, volume 1. De Gruyter Mouton, Berlin, New York.
- Cichosz, A. (2010). Influence of Text Type on Word Order of Old Germanic Languages: A Corpus-based Contrastive Study of Old English and Old High German, volume 27. Peter Lang, Frankfurt am Main.
- Cichosz, A. (2017). Verb-initial main clauses in Bede: A translation effect? In Current Developments in English Historical Linguistics: Studies in Honour of Rafał Molencki, pages 68–89. Wydawnictwo Uniwersytetu Śląskiego Katowice, Katowice.
- Cinque, G. (1977). The Movement Nature of Left Dislocation. Linguistic Inquiry, 8(2):397–412.
- Cinque, G. (1993). A Null Theory of Phrase and Compound Stress. Linguistic Inquiry, 24(2):239–297.
- Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In Paths towards universal grammar. Studies in honor of Richard S. Kayne, pages 85–110. Georgetown University Press, Washington, DC.
- Cinque, G. and Rizzi, L. (2015). The Cartography of Syntactic Structures. In Heine, B. and Narrog, H., editors, The Oxford Handbook of Linguistic Analysis, pages 65–78. Oxford University Press, Oxford.
- Clackson, J. (2007). Indo-European Linguistics: An Introduction. Cambridge University Press, Cambridge.
- Clackson, J. (2017). The Comparative Method and Comparative Reconstruction. In Ledgeway, A. and Roberts, I., editors, The Cambridge Handbook of Historical Syntax, Cambridge Handbooks in Language and Linguistics, pages 189–206. Cambridge University Press, Cambridge.

REFERENCES

- Collins, D. E. (2011). The Pragmatics of "Unruly" Dative Absolutes in Early Slavic. Oslo Studies in Language: Indo-European Syntax and Pragmatics: Contrastive Approaches, 3(3).
- Collins, J. (2024). Syntax, Truth, and the Fate of Sentences. The Philosophical Quarterly, 74(1):125–144.
- Comrie, B. (1998). The Indo-European Linguistic Family: Genetic and Typological Perspectives. In The Indo-European languages, pages 74–97. Routledge, London, New York.
- Corthals, J. (1996). Early Irish retoirics and their late antique background. Cambrian Medieval Celtic Studies, 31:17–36.
- Cowgill, W. (1975). The Origins of the Insular Celtic Conjunct and Absolute Verbal Endings. Reichert Verlag, Wiesbaden.
- Dal Lago, N. (2010). Fenomeni di prolessi (pro) nominale e struttura della periferia sinistra nel greco di Senofonte. PhD Thesis, Università degli studi di Padova.
- Danckaert, L. (2011). On the left periphery of Latin embedded clauses. PhD Thesis, Ghent University.
- Danylenko, A. (2019). Do the parallels meet? On the origin of the accusative with infinitive construction in Slavic. Belgian Journal of Linguistics, 33(1):150–182.
- de Jonge, C. C. (2007). Chapter Eleven. From Demetrius To Dik. Ancient And Modern Views On Greek And Latin Word Order. In The Language of Literature: Linguistic Approaches to Classical Texts, pages 211 – 232. Brill, Leiden, Boston.
- de Vaan, M. and García, J. M. (2014). Introduction to Avestan, volume 1. Brill, Leiden, Boston.
- Delbrück, B. (1878). Die altindische Wortfolge aus dem Çatapathabrāhmaṇa, volume 3 of Syntaktische Forschungen. Verlag der Buchhandlung des Waisenhauses, Halle.
- Delbrück, B. (1888). Altindische Syntax. Verlag der Buchhandlung des Waisenhauses, Halle.

REFERENCES

- Delbrück, B. (1893). Vergleichende syntax der indogermanischen sprachen. Verlag Karl J. Trübner, Strasbourg.
- den Besten, H. (1977). On the interaction of root transformations and lexical deletive verbs. University of Amsterdam.
- Devine, A. M. and Stephens, L. D. (2000). Discontinuous Syntax: Hyperbaton in Greek. Oxford University Press, New York.
- Devine, A. M. and Stephens, L. D. (2006). Latin Word Order: Structured Meaning and Information. Oxford University Press, New York.
- Devine, A. M. and Stephens, L. D. (2019). Pragmatics for Latin: From Syntax to Information Structure. Oxford University Press, New York.
- Dezső, L. (1978). Towards a typology of theme and rheme: SOV languages. In Akten des 12. Linguistischen Kolloquiums, Pavia 1977, Bd. 1, pages 3–12. Max Niemeyer Verlag, Halle.
- DiGirolamo, C. M. (2018). Word order and information structure in the Würzburg Glosses. In Martins, A. M. and Cardoso, A., editors, Word Order Change, Oxford Studies in Diachronic and Historical Linguistics, pages 143–160. Oxford University Press, Oxford.
- Dik, H. (1995). Word order in ancient Greek: a pragmatic account of word order variation in Herodotus. In Word Order in Ancient Greek. Brill, Leiden, Boston.
- Dik, S. C. (1997). The Theory of Functional Grammar: Part 2: Complex and Derived Structures, volume 21 of Functional Grammar Series. Mouton de Gruyter, Berlin, New York.
- Dimitrova, T. (2008). The Old Bulgarian noun phrase: Towards an annotation specification. PhD Thesis, Norwegian University of Science and Technology.

REFERENCES

- Disterheft, D. (1981). Remarks on the history of the Indo-European infinitive. Folia Linguistica Historica, 15:3–34.
- Doyle, A. (2018). Würzburg Irish Glosses. <https://wuerzburg.ie>.
- Drinka, B. (2011). The sacral stamp of Greek: periphrastic constructions in new testament translations of Latin, Gothic, and Old Church Slavonic. Oslo Studies in Language: Indo-European Syntax and Pragmatics: Contrastive Approaches, 3(3):41–73.
- Dryer, M. S. (2013a). Order of Adverbial Subordinator and Clause (v2020.3). In Dryer, M. S. and Haspelmath, M., editors, The World Atlas of Language Structures Online. Zenodo.
- Dryer, M. S. (2013b). Polar Questions (v2020.3). In Dryer, M. S. and Haspelmath, M., editors, The World Atlas of Language Structures Online. Zenodo.
- Du Toit, H. (1987). Wh–movement in Ancient Greek and Latin. South African Journal of Linguistics, 5(2):61–74.
- Dunn, G. (1988). Syntactic word order in Herodotean Greek. Glotta, 66(1/2):63–79.
- Embick, D. and Noyer, R. (2001). Movement Operations after Syntax. Linguistic Inquiry, 32(4):555–595.
- Erschler, D. (2012). From preverbal focus to preverbal “left periphery”: The Ossetic clause architecture in areal and diachronic perspective. Lingua, 122(6):673–699.
- Eska, J. F. (2012). Absolute and conjunct, Cowgill and apocope. In The Indo-European Verb: Proceedings of the Conference of the Society for Indo-European Studies, Los Angeles 13–15 September 2010, pages 51–59. Reichert Verlag, Wiesbaden.
- Evans, E. (2019). The origin, functions, and histories of Germanic adjective endings. PhD Thesis, Indiana University.

REFERENCES

- Faarlund, J. T. (1990). Syntactic Change: Toward a Theory of Historical Syntax. Number 50 in Trends in Linguistics. Studies and Monographs. De Gruyter Mouton, Berlin, New York.
- Faarlund, J. T. (2004). The Syntax of Old Norse: With a survey of the inflectional morphology and a complete bibliography. Oxford University Press, Oxford.
- Faarlund, J. T. (2009). On the history of definiteness marking in Scandinavian. Journal of Linguistics, 45(3):617–639.
- Fauconnier, S. (2014). Internal and external relative clauses in Ancient Greek. Journal of Greek Linguistics, 14(2):141–162.
- Ferraresi, G. and Goldbach, M. (2003). Some reflections on inertia: Infinitive complements in Latin. Zeitschrift für interkulturellen Fremdsprachenunterricht, 8(2).
- Ferraresi, G. and Goldbach, M. (2008). Syntactic Reconstruction: Methods and New Insights. In Principles of Syntactic Reconstruction, number 302 in Current Issues in Linguistic Theory, pages 1–26. John Benjamins Publishing Company, Amsterdam.
- Ferrari-Bridgers, F. (2010). The Ripano dialect: towards the end of a mysterious linguistic island in the heart of Italy. In Marginal Dialects: Scotland, Ireland and Beyond, pages 106–130. Forum for Research on the Languages of Scotland and Ireland, Aberdeen.
- Fischer, O. (2000). The position of the adjective in Old English. In Generative Theory and Corpus Studies: A Dialogue from 10 ICEHL, number 31 in Topics in English Linguistics, pages 153–182. Mouton de Gruyter, Berlin, New York.
- Fischer, O. (2012). The status of the postponed ‘and-adjective’ construction in Old English: attributive or predicative. In Analysing Older English, Studies in English Language, pages 251–284. Cambridge University Press, Cambridge.

REFERENCES

- Fortson, B. (2006). Proto-Indo-European Syntax. In Encyclopedia of Language and Linguistics. Elsevier.
- Frascarelli, M. and Hinterhölzl, R. (2007). Types of topics in German and Italian. In On Information Structure, Meaning and Form: Generalizations across languages, number 100 in Linguistics Today, pages 87–116. John Benjamins Publishing Company, Amsterdam.
- Fraser, B. L. (1999). Word order, focus, and clause linking in Greek tragic poetry. PhD Thesis, University of Cambridge.
- Frederick, N. (2009). The Syntax of Determiners in Attic Greek: A Theoretical Approach. BA Thesis, University of Michigan.
- Friedrich, P. (1975). Proto-Indo-European Syntax: the Order of Meaningful Elements, volume 1 of Journal of Indo-European Studies. Institute for the Study of Man, Washington, DC.
- Fulk, R. D. (2014). An Introductory Grammar of Old English. ACMRS Press, Tempe, AZ.
- Fykias, I. (2013). Infinitives (Syntax). In Encyclopedia of Ancient Greek Language and Linguistics. Brill, Leiden, Boston.
- García-Castillero, C. (2020). Clause Typing in the Old Irish Verbal Complex. Trends in Linguistics. Studies and Monographs. De Gruyter Mouton, Berlin, New York.
- Garrett, A. (1990). The syntax of Anatolian pronominal clitics. PhD Thesis, Harvard University.
- Garrett, A. (1994). Relative clause syntax in Lycian and Hittite. Die Sprache, 36(1):29–69.
- Gaskin, R. (2020). Language and World: A Defence of Linguistic Idealism. Routledge, London, New York.
- Giannakis, G. K. (2021). The Concept of ‘Emphasis’ in Ancient Greek and Indo-European. In Synchrony and Diachrony of Ancient Greek: Language, Linguistics and Philology, number 112 in Trends in Classics, pages 413–438. De Gruyter, Berlin, Boston. De Gruyter.

REFERENCES

- Gildea, S., Luján, E. R., and Barðdal, J. (2020). The Curious Case of Reconstruction in Syntax. In Reconstructing Syntax, pages 1–44. Brill, Leiden, Boston.
- Giovanni, G. and Celano, A. (2013). Word Order. In Encyclopedia of Ancient Greek Language and Linguistics. Brill, Leiden, Boston.
- Giusti, G. and Iovino, R. (2016). Latin as a split-DP language. Studia linguistica, 70(3):221–249.
- Givón, T. (1971). Historical syntax and synchronic morphology: An archaeologist’s field trip. In Papers from the 7th Meeting of the Chicago Linguistic Society, pages 394–415. Chicago Linguistic Society, Chicago.
- Gneuss, H. (1991). Old English language. In The Cambridge Companion to Old English Literature, pages 23–54. Cambridge University Press, Cambridge.
- Goedegebuure, P. (2009). Focus structure and Q-word questions in Hittite. Linguistics, 47:945–969.
- Goldstein, D. M. (2010). Wackernagel’s law in fifth-century Greek. PhD Thesis, University of California, Berkeley.
- Goldstein, D. M. (2015). Classical Greek syntax: Wackernagel’s law in Herodotus. Brill, Leiden, Boston.
- Goldstein, D. M. (2016). Variation versus change: Clausal clitics between Homer and Herodotus. Indo-European Linguistics, 4(1):53–97.
- Goldstein, D. M. (2022). The Old Irish article. Journal of Celtic Linguistics, 23:1–34.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Universals of Language, pages 40–70. MIT Press, Cambridge, MA.
- Greenberg, J. H. (1978). How does a language acquire gender markers? In Universals of Human Language, volume 3. Stanford University Press, Stanford, CA.

REFERENCES

- Greene, D. (1992). Celtic. In Indo-European Numerals, pages 497–554. De Gruyter Mouton, Berlin, New York.
- Griffith, A. (2013). The Milan Glosses Database. <https://indogermanistik.univie.ac.at/milan-glosses/>.
- Grimshaw, J. (2000). Locality and extended projection. In Lexical Specification and Insertion, number 197 in Current Issues in Linguistic Theory, pages 115–134. John Benjamins Publishing Company, Amsterdam.
- Guardiano, C. (2009). The Greek demonstratives: history and syntax. In 9th International Conference on Greek Linguistics (ICGL 9). University of Chicago.
- Hackstein, O. (2016). Negative interrogatives and whatnot: The conversion of negation in Indo-European. Indo-European Linguistics, 4(1):150–189.
- Haerberli, E. (2002). Inflectional Morphology and the Loss of V2 in English. In Syntactic Effects of Morphological Change, pages 88–106. Oxford University Press, Oxford.
- Hale, M. (1987a). Notes on Wackernagel’s Law in the language of the Rigveda. In Studies in Memory of Warren Cowgill, 1929–1985: Papers from the Fourth East Coast Indo-European Conference, Cornell University, June 6–9, number 3 in Untersuchungen zur indogermanischen Sprach- und Kulturwissenschaft. NF / Studies in Indo-European Language and Culture. New Series, pages 38–50. De Gruyter, Berlin, Boston.
- Hale, M. (1987b). Studies in the comparative syntax of the oldest Indo-Iranian languages. PhD Thesis, Harvard University.
- Hale, M. (1991). Some observations on intersentential pronominalization in the language of the Taittirīya Saḥitā. In Sense and Syntax in Vedic, volume 4 of Panels of the Seventh World Sanskrit Conference, page 218. Brill, Leiden, Boston.

REFERENCES

- Hale, M. (1993). Tmesis and Movement in Avestan. *Indo-Iranian Journal*, 36(1):29–43.
- Hale, M. (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*, pages 165–197. Center for the Study of Language and Information (CSLI), Stanford, CA. Center for the Study of Language and Information.
- Hale, M. (2009). Phonology, Syntax and Discourse in the Rigveda: or, R.I.P. Wackernagel’s Law. In *The International Workshop on Indo-European Syntax and Pragmatics, May 13-15, University of Georgia*.
- Hale, M. (2017). Preliminaries to the investigation of clitic sequencing in Greek and Indo-Iranian. In *On looking into words (and beyond): Structures, Relations, Analyses*, pages 289–310. Language Science Press, Berlin.
- Hale, M. (2018). The syntax of Indo-Iranian. In *Handbook of Comparative and Historical Indo-European Linguistics*, volume 3, pages 1924–1942. De Gruyter Mouton, Berlin, Boston.
- Halla-Aho, H. (2018). *Left-Dislocation in Latin: Topics and Syntax in Republican Texts*, volume 28 of *Amsterdam Studies in Classical Philology*. Brill, Leiden, Boston.
- Halle, M. and Marantz, A. (1993). Distributed Morphology and the pieces of Inflection. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, pages 111–176. MIT Press, Cambridge, MA. MIT Press.
- Halpern, A. L. (1992). *Topics in the Placement and Morphology of Clitics*. PhD, Stanford University.
- Halpern, A. L. (1995). *On the Placement and Morphology of Clitics*. Center for the Study of Language and Information (CSLI), Stanford, CA.
- Harris, A. C. and Campbell, L. (1995). *Historical Syntax in Cross-Linguistic Perspective*. Number 74 in *Cambridge Studies in Linguistics*. Cambridge University Press, Cambridge.

REFERENCES

- Haudry, J. (1973). Parataxe, hypotaxe et corrélation dans la phrase latine. Bulletin de la Société de Linguistique de Paris, 68:147–186.
- Haugan, J. (2000). Old Norse word order and information structure. PhD Thesis, Norwegian University of Science and Technology.
- Havers, W. (1926). Der sog. ‘Nominativus pendens’. Indogermanische Forschungen, 43(1):207–257.
- Hawkins, J. A. (2009). Germanic Languages. In The World’s Major Languages, pages 51–58. Routledge, London, New York.
- Hejná, M. and Walkden, G. (2022). A History of English. Number 9 in Textbooks in Language Sciences. Language Science Press, Berlin.
- Hettrich, H. (1988). Untersuchungen zur Hypotaxe im Vedischen. Untersuchungen zur indogermanischen Sprach- und Kulturwissenschaft. NF / Studies in Indo-European Language and Culture. New Series. De Gruyter, Berlin, Boston.
- Hickey, R. (1996). Sound change and typological shift: Initial mutation in Celtic. In Linguistic Reconstruction and Typology, number 96 in Trends in Linguistics. Studies and Monographs, pages 133–182. De Gruyter Mouton, Berlin, New York.
- Hinterhölzl, R. and Petrova, S. (2009). From V1 to V2 in West Germanic. Lingua, 120(2):315–328.
- Hinterhölzl, R. and van Kemenade, A. (2012). The interaction between syntax, information structure and prosody in word order change. In The Oxford Handbook of the History of English, pages 1154–1182. Oxford University Press, Oxford.
- Hock, H. H. (1982). The Sanskrit quotative: A historical and comparative study in papers on diachronic syntax: Six case studies. Studies in the Linguistic Sciences, 12(2):39–85.

REFERENCES

- Hock, H. H. (1989). Conjoined We Stand: Theoretical Implications of Sanskrit Relative Structures in Papers in General Linguistics. Studies in the Linguistic Sciences, 19(1):93–126.
- Hock, H. H. (1991). Principles of Historical Linguistics. Mouton de Gruyter, Berlin, New York, Amsterdam, second edition.
- Hock, H. H. (1992). Reconstruction and syntactic typology: A plea for a different approach. In Explanation in Historical Linguistics, number 84 in Current Issues in Linguistic Theory, page 105–122. John Benjamins Publishing Company, Amsterdam.
- Hoffner, H. A. (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, pages 87–104. Nederlands Historisch-Archaeologisch Instituut te Istanbul.
- Hogg, R. M. and Fulk, R. D. (2011). A Grammar of Old English, Volume 2: Morphology. John Wiley & Sons, Chichester.
- Horrocks, G. and Stavrou, M. (1987). Bounding theory and Greek syntax: Evidence for wh-movement in NP. Journal of Linguistics, 23(1):79–108.
- Huggard, M. (2011). On Wh-(non)-movement and internal structures of the Hittite preposed relative clause. In Proceedings of the 22nd Annual UCLA Indo-European Conference, pages 83–104. Hemen, Bremen.
- Huntley, D. (2001). Old Church Slavonic. In The Slavonic Languages, pages 125–187. Routledge, London, New York.
- Iovino, R. (2012). La sintassi dei modificatori nominali in latino. Lincom Europa, Munich.
- Janhunen, J. (1982). On the structure of Proto-Uralic. Finnisch-Ugrische Forschungen, 1982(44):23–42.

REFERENCES

- Jeffers, R. J. (1976). Syntactic change and syntactic reconstruction. In Current progress in historical linguistics: Proceedings of the Second International Conference on Historical Linguistics, Tucson, Arizona, 12-16 January 1976, volume 1, pages 1–15. North-Holland Publishing Company, Amsterdam.
- Jones, W. (1798). The third anniversary discourse, delivered 2nd February, 1786: on the Hindus. Asiatick Researches, 1:415–431.
- Kane, F. (2013). Demonstratives in Irish. In Minority Languages, Microvariation, Minimalism and Meaning: Proceedings of the Irish Network in Formal Linguistics, pages 131–148. Cambridge Scholars Publishing, Newcastle upon Tyne.
- Kane, F. (2015). The Fine Structure of the Irish NP. PhD Thesis, Ulster University.
- Kayne, R. S. (1994). The Antisymmetry of Syntax. Number 25 in Linguistic Inquiry Monograph. MIT press, Cambridge, MA.
- Keenan, E. L. (2002). Explaining the creation of reflexive pronouns in English. In Minkova, D. and Stockwell, R., editors, Studies in the History of the English Language: A Millennial Perspective, number 39 in Topics in English Linguistics, pages 325–354. De Gruyter Mouton, Berlin, New York.
- Keydana, G. (2011). Wackernagel in the Language of the Rigveda. A Reassessment. Historische Sprachforschung, 124(1):106–133.
- Kim, A. H.-O. (1988). Preverbal focusing and type XXIII languages. In Studies in Syntactic Typology, number 17 in Typological Studies in Language, pages 147–169. John Benjamins Publishing Company, Amsterdam.
- Kim, R. (2000). The Distribution of the Old Irish Infix Pronouns: Evidence for the Syntactic Evolution of Insular Celtic? Current Work in Linguistics: University of Pennsylvania Working Papers in Linguistics, 6(3):167–187. ERIC.

REFERENCES

- Kiparsky, P. (1995). The Indo-European Origins of Germanic Syntax. In Clause Structure and Language Change, Oxford Studies in Comparative Syntax, pages 140–169. Oxford University Press, Oxford.
- Kirk, A. (2007). A syntactic account of split DPs in Herodotus. PhD Thesis, Concordia University.
- Kiss, K. (1995a). Discourse Configurational Languages. Oxford Studies in Comparative Syntax. Oxford University Press, Oxford.
- Kiss, K. (1995b). Introduction. In Discourse Configurational Languages, Oxford Studies in Comparative Syntax. Oxford University Press, Oxford.
- Kloekhorst, A. (2022). Anatolian. In Olander, T., editor, The Indo-European Language Family: A Phylogenetic Perspective, pages 63–82. Cambridge University Press, Cambridge.
- Kloekhorst, A. and Pronk, T. (2019). Introduction: Reconstructing Proto-Indo-Anatolian and Proto-Indo-Uralic. In The Precursors of Proto-Indo-European: The Indo-Anatolian and Indo-Uralic Hypotheses, pages 1 – 14. Brill, Leiden, Boston.
- Kortlandt, F. H. H. (2007). Italo-Celtic Origins and Prehistoric Development of the Irish Language, volume 14 of Leiden studies in Indo-European. Rodopi, Amsterdam.
- Krifka, M. (1993). Focus and Presupposition in Dynamic Interpretation. Journal of Semantics, 10(4):269–300.
- Langston, K. (2018). The documentation of Slavic. In Handbook of Comparative and Historical Indo-European Linguistics, volume 3, pages 1397–1413. De Gruyter Mouton, Berlin, Boston.
- Lash, E. (2020). Transitivity and Subject Positions in Old Irish. Transactions of the Philological Society, 118(1):94–140.
- Le Bruyn, B., de Swart, H., and Zwarts, J. (2017). Bare nominals. In Oxford Research Encyclopedia of Linguistics. Oxford University Press, Oxford.

REFERENCES

- Lehmann, W. P. (1974). Proto-Indo-European Syntax. University of Texas Press, Austin, TX.
- Leu, T. (2008). The Internal Syntax of Determiners. PhD Thesis, University of New York.
- Lightfoot, D. W. (1979). Principles of Diachronic Syntax. Number 23 in Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.
- Lightfoot, D. W. (2002). Myths and the prehistory of grammars. Journal of Linguistics, 38(1):113–136.
- Lindberg, R. H. (2013). Definiteness in Old Church Slavonic: A study of how long and short form in adjectives reflect information status. Master's thesis, University of Oslo.
- Links, M. (2019). Expressing conditionality in earlier English. English Language and Linguistics, 23(1):155–182.
- Longobardi, G. (1994). Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form. Linguistic Inquiry, 25(4):609–665.
- Longobardi, G. (2001a). Formal syntax, diachronic minimalism, and etymology: the history of French, *chez*. Linguistic Inquiry, 32(2):275–302.
- Longobardi, G. (2001b). The structure of DPs: Some principles, parameters, and problems. In The Handbook of Contemporary Syntactic Theory, pages 562–603. Blackwell Publishing, Oxford.
- Longobardi, G. (2008). Foreword. In Principles of Syntactic Reconstruction, number 302 in Current Issues in Linguistic Theory. John Benjamins Publishing Company, Amsterdam.
- Lord, A. B. (1987). Characteristics of Orality. Oral Tradition, 2(1):54–72.
- Los, B. (2005). The Rise of the 'To'-Infinitive. Oxford University Press, Oxford.
- Louagie, D. (2023). The Noun Phrase. In Bower, C., editor, The Oxford Guide to Australian Languages, pages 196–204. Oxford University Press, Oxford.

REFERENCES

- Lowe, J. J. (2014). Accented Clitics in the Rgveda. Transactions of the Philological Society, 112(1):5–43.
- Lowe, J. J. (2015). Participles in Rigvedic Sanskrit: The Syntax and Semantics of Adjectival Verb Forms. Oxford University Press, Oxford.
- Lunt, H. G. (2001). Old Church Slavonic Grammar. De Gruyter Mouton, Berlin, New York, seventh edition.
- Luraghi, S. (2004). Null objects in Latin and Greek and the relevance of linguistic typology for language reconstruction. In Proceedings of the Fifteenth Annual UCLA Indo-European Conference: Los Angeles, November 7-8, 2003, number 49 in Journal of Indo-European Studies, pages 234–256. Institute for the Study of Man, Washington, DC.
- Lühr, R. (2015). Traces of discourse configurationality in older Indo-European languages? In Perspectives on Historical Syntax, number 169 in Studies in Language Companion Series, pages 203–232. John Benjamins Publishing Company, Amsterdam.
- Macdonell, A. A. (1916). A Vedic grammar for students including a chapter on syntax and three appendixes. Clarendon Press, Oxford.
- Magni, E. (2009). The Evolution of Latin Word (Dis)order. In Universals of Language Today, number 76 in Studies in Natural Language and Linguistic Theory, pages 225–251. Springer, New York.
- Manolessou, I. and Horrocks, G. (2007). The development of the definite article in Greek. Studies in Greek linguistics, 27(1):224–236.
- Masica, C. (1991). The Indo-Aryan Languages. Cambridge Language Surveys. Cambridge University Press, Cambridge.
- Matasovic, R. (2009). Etymological Dictionary of Proto-Celtic. Brill, Leiden, Boston.

REFERENCES

- Matić, D. (2003). Topic, focus, and discourse structure: Ancient Greek word order. Studies in Language, 27(3):573–633.
- McCone, K. (1980). The Nasalizing Relative Clause with Object Antecedent in the Glosses. Ériu, 31:10–27.
- McCone, K. (1997). The Early Irish Verb. Maynooth Monographs. An Sagart, Maynooth.
- Melazzo, L. (2005). Latin object and subject infinitive clauses. In Universal Grammar in the Reconstruction of Ancient Languages, volume 83 of Studies in Generative Grammar, pages 339–372. De Gruyter Mouton, Berlin, New York.
- Melchert, H. C. and Hoffner, H. A. (2008). A Grammar of the Hittite Language: Part 1: Reference Grammar. Eisenbrauns, Winona Lake, IN.
- Metzger, B. M. (1977). The Old Church Slavonic Version. In The Early Versions of the New Testament: Their Origin, Transmission, and Limitations, pages 394–442. Clarendon Press, Oxford.
- Miller, T., editor (1890). The Old English Version of Bede's Ecclesiastical History of the English People. Early English Text Society, Oxford.
- Minard, A. (1936). La subordination dans la prose védique. Les Belles Lettres, Paris.
- Molencik, R. (2012). Old English: Syntax. In English Historical Linguistics: An International Handbook, number 34 (1) in Handbooks of Linguistics and Communication Science, pages 294–312. De Gruyter Mouton, Berlin, Boston.
- Monier-Williams, M. (1899). A Sanskrit-English Dictionary. The Clarendon Press, Oxford.
- Montague, R. (1973). The proper treatment of quantification in ordinary English. In Approaches to Natural Language: Proceedings of the 1970 Stanford workshop on grammar and semantics, number 49 in Synthese Library, pages 221–242. Springer, Dordrecht.

REFERENCES

- Norris, M. (2018). Nominal structure in a language without articles: The case of Estonian. Glossa, 3(1)(41):1–39.
- Nygaard, M. (1905). Norrøn syntax. H. Aschehoug & Company, Oslo.
- Oberlies, T. (2017). The evolution of Indic. In Handbook of Comparative and Historical Indo-European Linguistics, volume 1, pages 447–470. De Gruyter Mouton, Berlin, New York.
- Olander, T. (2022a). Introduction. In The Indo-European Language Family: A Phylogenetic Perspective, pages 1–17. Cambridge University Press, Cambridge.
- Olander, T. (2022b). To *b or not to *b: Proto-Indo-European *b in a phylogenetic perspective. Historische Sprachforschung, 133(1):182–208.
- Oniga, R. (2014). Latin: A Linguistic Introduction. Oxford University Press, Oxford.
- Onishi, T. (2022). Pronominal Clitics in Tocharian: A Study in the Morphology-Syntax Interface. PhD Thesis, University of California, Los Angeles.
- Osthoff, H. and Brugmann, K. (1878). Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen, volume 1. S. Hirzel Verlag, Leipzig.
- Pancheva, R. (2008). Head-Directionality of TP in Old Church Slavonic. In Formal Approaches to Slavic Linguistics 16: The Stony Brook Meeting 2007, pages 313–332. Michigan Slavic Publications.
- Paul, H. (1886). Principien der Sprachgeschichte. Halle, second edition.
- Petrocchi, A. (2022). The morphosyntax and semantics of cardinal numerals in classical sanskrit. Transactions of the Philological Society, 120(2):276–298.
- Pfaff, A. and Walkden, G. (2024). Adjectival articles in early Germanic. In Noun phrases in early Germanic languages, pages 323–364. Language Science Press, Berlin.

REFERENCES

- Pinkster, H. (1991). Evidence for SVO in Latin? In Latin and the Romance Languages in the Early Middle Ages, pages 69–82. Routledge, London.
- Pinkster, H. (2015). The Oxford Latin Syntax: Volume 1: The Simple Clause. Oxford University Press, Oxford.
- Pinkster, H. (2021). The Oxford Latin Syntax: Volume 2: the Complex Sentence and Discourse. Oxford University Press, Oxford.
- Pintzuk, S. (1993). Verb seconding in Old English: verb movement to Infl. Linguistic Review, 10:5–35.
- Pintzuk, S. and Haeberli, E. (2008). Structural variation in Old English root clauses. Language Variation and Change, 20(3):367–407.
- Pires, A. and Thomason, S. G. (2008). How much syntactic reconstruction is possible? In Principles of Syntactic Reconstruction, number 302 in Current Issues in Linguistic Theory, pages 27–72. John Benjamins Publishing Company, Amsterdam.
- Pokorny, J. (1959). Indogermanisches etymologisches Wörterbuch. A. Francke Verlag, Bern.
- Polo, C. (2004). Word Order between Morphology and Syntax. UniPress, Padova.
- Porck, T. (2020). “Men þa leofestan!” Manuscript variations of an Old English formula. <https://thijsporck.com/2020/10/26/men-tha-leofestan-manuscript-variations-of-an-old-english-formula/>.
- Probert, P. (2015). Early Greek Relative Clauses. Oxford University Press, Oxford.
- Probert, P. and Dickey, E. (2016). Six notes on Latin correlatives. In Early and Late Latin: Continuity or Change?, pages 390–419. Cambridge University Press, Cambridge.
- Punske, J. (2014). Functional structure inside nominal phrases. In Carnie, A., Sato, Y., and Siddiqi, D., editors, The Routledge Handbook of Syntax, pages 65–88. Routledge, London. Routledge.

REFERENCES

- Pysz, A. (2009). The syntax of prenominal and postnominal adjectives in Old English. Cambridge Scholars Publishing, Newcastle upon Tyne.
- Ram-Prasad, K. J. (2022). The Syntax of Relative Clauses and Related Phenomena in Proto-Indo-European. PhD Thesis, University of Cambridge.
- Ramat, A. G. and Ramat, P., editors (1998). The Indo-European Languages. Routledge, London, New York, first edition.
- Ratkus, A. (2017). Features of the Gothic adjective, with special reference to determination. In Lingua Gotica: Novye Issledovanija, volume 3. Institute of Linguistics, Academy of Sciences of the Russian Federation, Moscow.
- Rauer, C. (2013). The Old English Martyrology: Edition, Translation and Commentary. Anglo-Saxon Texts. Boydell & Brewer, Cambridge.
- Recht, T. (2015). Verb-Initial Clauses in Ancient Greek Prose: A Discourse-Pragmatic Study. PhD Thesis, University of California, Berkeley.
- Reinöhl, U. (2016). Grammaticalization and the Rise of Configurationality in Indo-Aryan. Oxford Studies in Diachronic and Historical Linguistics. Oxford University Press, Oxford.
- Ringe, D. (2006). From Proto-Indo-European to Proto-Germanic, volume 1 of A Linguistic History of English. Oxford University Press, New York.
- Ringe, D. and Taylor, A. (2014). The Development of Old English. Oxford University Press, Oxford.
- Risselada, R. (1993). Imperatives and Other Directive Expressions in Latin: A Study in the Pragmatics of a Dead Language, volume 2 of Amsterdam Studies in Classical Philology. Brill, Leiden, Boston.
- Rizzi, L. (1997). The Fine Structure of the Left Periphery. In Elements of Grammar: Handbook in Generative Syntax, pages 281–337. Kluwer, Dordrecht.

REFERENCES

- Roberts, I. (1998). "Historical Syntax in Cross-Linguistic Perspective", by Alice C. Harris and Lyle Campbell (Book Review). Romance Philology, 51(3):363–370.
- Roberts, I. (2007). Diachronic Syntax. Oxford University Press, Oxford.
- Roberts, I. (2010). Agreement and Head Movement: Clitics, Incorporation, and Defective Goals. MIT Press, Cambridge, MA.
- Rooth, M. (1992). A Theory of Focus Interpretation. Natural Language Semantics, 1(1):75–116.
- Ross, J. R. (1967). Constraints on variables in syntax. PhD Thesis, MIT.
- Ruppel, A. (2012). Absolute Constructions in Early Indo-European. Cambridge Classical Studies. Cambridge University Press, Cambridge.
- Russell, P. (2014). An Introduction to the Celtic Languages. Routledge, London, New York.
- Rögnvaldsson, E. (1995). Old Icelandic: A non-configurational language. North-Western European Language Evolution, 26(1):3–29.
- Saarinen, S. (2022). Mari. In Bakró-Nagy, M., Laakso, J., and Skribnik, E., editors, The Oxford Guide to the Uralic Languages, pages 432–470. Oxford University Press, Oxford.
- Salvesen, C. M. and Walkden, G. (2017). Diagnosing embedded V2 in Old French and Old English. In From Micro-change to Macro-change in Diachronic Syntax, Oxford Studies in Diachronic and Historical Linguistics. Oxford University Press, Oxford.
- Salvi, G. (2005). Some firm points on Latin word order: The left periphery. In Universal Grammar in the Reconstruction of Ancient Languages, number 83 in Studies in Generative Grammar, pages 429–456. De Gruyter Mouton, Berlin, New York.
- Sansone, D. (1993). Towards a new doctrine of the article in Greek: some observations on the definite article in Plato. Classical Philology, 88(3):191–205.

REFERENCES

- Schrijver, P. (1994). The Celtic Adverbs for ‘Against’ and ‘With’ and the Early Apocope of -I. Ériu, 45:151–189.
- Schumacher, S., Schulze-Thulin, B., and aan de Wiel, C. (2004). Die keltischen Primärverben: ein vergleichendes etymologisches und morphologisches Lexikon, volume 110 of Innsbrucker Beiträge zur Sprachwissenschaft. Institut für Sprachen und Literaturen der Universität Innsbruck, Innsbruck.
- Schäufele, S. W. (1991a). Free word-order syntax: The challenge from Vedic Sanskrit to contemporary formal syntactic theory. PhD Thesis, University of Illinois at Urbana-Champaign.
- Schäufele, S. W. (1991b). Verb-medial clauses in Vedic: Some theoretical implications. In Studies in Sanskrit syntax: a volume in honour of the centennial of Speijer’s Sanskrit syntax (1886-1986), pages 177–196. Motilal Banarsidass Publishers, Delhi.
- Sevdali, C. (2007). Infinitival clauses in Ancient Greek: Overt and full subjects, the role of Case and Focus. PhD Thesis, University of Cambridge.
- Sideltsev, A. V. (2014). Wh-in-situ in Hittite. In Typology of Morphosyntactic Parameters. Proceedings of the International Conference Typology of Morphosyntactic Parameters 2014, pages 199–222. Sholokhov Moscow State University for Humanities, Moscow.
- Sideltsev, A. V. (2015a). Hittite Clause Architecture. Revue d’Assyriologie et d’archéologie orientale, 109(1):79–112.
- Sideltsev, A. V. (2015b). The riddles of Hittite indefinite pronouns. Altorientalische Forschungen, 42(2):199–275.
- Sihler, A. L. (1995). New Comparative Grammar of Greek and Latin. Oxford University Press, New York.
- Sleeman, A. P. and Perridon, H. (2011). The Noun Phrase in Romance and Germanic: Structure, Variation, and Change. Number 171 in Linguistics Today. John Benjamins Publishing Company, Amsterdam.

REFERENCES

- Smyth, H. W. (1920). A Greek Grammar for Colleges. American Book Company, New York.
- Sommerer, L. (2018). Article Emergence in Old English: A Constructionalist Perspective. Number 99 in Topics in English Linguistics. De Gruyter Mouton, Berlin, New York.
- Speyer, J. S. (1886). Sanskrit Syntax. E. J. Brill, Leiden.
- Spieker, E. H. (1885). On the so-called Genitive Absolute and its Use especially in the Attic Orators. The American Journal of Philology, 6(3):310–343.
- Stalnaker, R. (2002). Common Ground. Linguistics and Philosophy, 25(5/6):701–721.
- Stifter, D. (2009). Early Irish. In The Celtic Languages, pages 55–116. Routledge, London, New York, second edition.
- Strang, B. M. H. (1970). A History of English. Routledge, London, New York.
- Strawson, P. F. (1964). Identifying reference and truth-values. Theoria, 30(2):96–118.
- Sturtevant, E. H. (1913). The Genitive and Dative Singular of the Latin Pronominal Declension. Transactions and Proceedings of the American Philological Association, 44:99–105.
- Sturtevant, E. H. (1933). A Comparative Grammar of the Hittite Language. Linguistic Society of America, Philadelphia, PA.
- Sturtevant, E. H. (1962). The Indo-Hittite Hypothesis. Language, 38(2):105–110.
- Sussex, R. and Cubberley, P. (2006). The Slavic Languages. Cambridge Language Surveys. Cambridge University Press, Cambridge.
- Swan, T. (1994). A note on Old English and Old Norse initial adverbials and word order with special reference to sentence adverbials. In Language Change and Language Structure, number 73 in Trends in Linguistics. Studies and Monographs, pages 233–270. De Gruyter Mouton, Berlin, New York.

REFERENCES

- Szabolcsi, A. (1984). The possessor that ran away from home. The Linguistic Review, 3:89–102.
- Szabolcsi, A. (1987). Functional categories in the noun phrase. In Approaches to Hungarian, volume 2, pages 167–191. József Attila University, Szeged.
- Taylor, A. (1990). Clitics and Configurationality in Ancient Greek. PhD Thesis, University of Pennsylvania.
- Thurneysen, R. (1946). A Grammar of Old Irish. Dublin Institute for Advanced Studies, Dublin.
- Torrego, M. E. (2017). Res Gestae Divi Augusti: Word order and pragmatics of the Latin original. In Pragmatic Approaches to Latin and Ancient Greek, number 190 in Studies in Language Companion Series, pages 159–179. John Benjamins Publishing Company, Amsterdam.
- Travis, L. d. (1984). Parameters and Effects of Word Order Variation. PhD Thesis, MIT.
- Uriagereka, J. (1995a). Aspects of the Syntax of Clitic Placement in Western Romance. Linguistic Inquiry, 26(1):79–123.
- Uriagereka, J. (1995b). An F Position in Western Romance. In Discourse Configurational Languages, volume 26 of Oxford Studies in Comparative Syntax, pages 79–123. Oxford University Press, Oxford.
- Vai, M. (2017). Struttura informativa della frase in greco omerico: periferia alta, periferia bassa; collocazione delle relative nella periferia sinistra. In Ancient Greek Linguistics: New Approaches, Insights, Perspectives, pages 625–640. De Gruyter Mouton, Berlin, Boston.
- Vaillant, A. (1942). L'article en vieux slave. Revue des Études Slaves, 20(1/4):5–12.
- van Gelderen, E. (2004). Economy, innovation, and prescriptivism: From Spec to Head and Head to Head. The Journal of Comparative Germanic Linguistics, 7:59–98.
- van Gelderen, E. (2011). The Linguistic Cycle: Language Change and the Language Faculty. Oxford University Press, Oxford.

REFERENCES

- van Gelderen, E. (2016). Features and affix-hop. Acta Linguistica Hungarica, 63(1):1–22.
- van Kemenade, A. (1987). Syntactic Case and Morphological Case in the History of English. Foris Publications, Dordrecht.
- van Kemenade, A., Hinterhölzl, R., and Struik, T. (2023). Word order change, architecture, and interfaces: Evidence from the development of V to C movement in the history of English. Journal of Historical Syntax, 7(25):1–55.
- Večerka, R. (1989). Altkirchenslavische (altbulgarische) Syntax: Bd. I, Die lineare Satzorganisation. U. W. Weihe, Freiburg.
- Viti, C. (2007). Strategies of Subordination in Vedic. FrancoAngeli, Milan.
- Viti, C. (2008). Genitive word order in Ancient Greek: A functional analysis of word order freedom in the noun phrase. Glotta, 84(1-4):203–238.
- Wackernagel, J. (1892). Über ein Gesetz der indogermanischen Wortstellung. Indogermanische Forschungen, 1:333–436.
- Walkden, G. (2013). The correspondence problem in syntactic reconstruction. Diachronica, 30(1):95–122.
- Walkden, G. (2014). Syntactic Reconstruction and Proto-Germanic. Oxford Studies in Diachronic and Historical Linguistics. Oxford University Press, Oxford.
- Walkden, G. (2017). Language contact and V3 in Germanic varieties new and old. The Journal of Comparative Germanic Linguistics, 20:49–81.
- Watkins, C. (1963). Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb. Celtica, 6:1–49.
- Watkins, C. (1964). Preliminaries to the reconstruction of Indo-European sentence structure. In Proceedings of the Ninth International Congress of Linguists, pages 1035–1045. Mouton, The Hague.

REFERENCES

- Watkins, C. (1969). On the Pre-History of Celtic Verb Inflexion. *Ériu*, 21:1–22.
- Watkins, C. (1976). Towards Proto-Indo-European syntax: problems and pseudo-problems. In Papers from the Parasession on Diachronic Syntax, April 22, 1976, pages 306–326. Chicago Linguistic Society, Chicago.
- Wenthe, M. R. (2012). Issues in the placement of enclitic personal pronouns in the Rigveda. PhD Thesis, University of Georgia.
- West, M. (2011). Old Avestan Syntax and Stylistics: with an edition of the texts. Number 13 in *Abhandlungen der Akademie der Wissenschaften zu Göttingen. Neue Folge*. De Gruyter, Berlin, Boston.
- Willis, D. (2000). Verb movement in Slavonic conditionals. In Diachronic Syntax: Models and Mechanisms, pages 322–348. Oxford University Press, Oxford.
- Willis, D. (2011). Reconstructing last week's weather: Syntactic reconstruction and Brythonic free relatives. *Journal of Linguistics*, 47(2):407–446.
- Windhearn, R. D. (2020). Evidence from Innovation: Reconstructing Disharmonic Headedness for Proto-Indo-European. PhD Thesis, Cornell University.
- Winter, W. (1984). Reconstructing comparative linguistics and the reconstruction of the syntax of undocumented stages in the development of languages and language families. In Fisiak, J., editor, Historical Syntax, number 23 in *Trends in Linguistics. Studies and Monographs*, pages 613–626. De Gruyter Mouton, Berlin, New York.
- Yoon, H.-C. (2002). Word order and structure of Old English: with special reference to Ælfric's Catholic Homilies. PhD Thesis, University of Edinburgh.
- Ó hUiginn, R. (1998). Complementation in Early Irish: The "Verba Dicendi". *Ériu*, 49:121–148.

REFERENCES

Porgeirsson, H. (2012). Late placement of the finite verb in Old Norse fornyrðislag meter. Journal of Germanic linguistics, 24(3):233–269.

Šereikaitė, M. (2019). Strong vs. weak definites: Evidence from Lithuanian adjectives. In Definiteness Across Languages, volume 25 of Studies in Diversity Linguistics. Language Science Press, Berlin.