

**Leonid Kulikov
and Nikolaos Lavidas (eds.)**

Proto-Indo-European Syntax and its Development

BENJAMINS CURRENT TOPICS

75

Proto-Indo-European Syntax and its Development

Benjamins Current Topics

ISSN 1874-0081

Special issues of established journals tend to circulate within the orbit of the subscribers of those journals. For the Benjamins Current Topics series a number of special issues of various journals have been selected containing salient topics of research with the aim of finding new audiences for topically interesting material, bringing such material to a wider readership in book format.

For an overview of all books published in this series, please see
<http://benjamins.com/catalog/bct>

Volume 75

Proto-Indo-European Syntax and its Development

Edited by Leonid Kulikov and Nikolaos Lavidas

These materials were previously published in *Journal of Historical Linguistics* 3:1 (2013).

Proto-Indo-European Syntax and its Development

Edited by

Leonid Kulikov

Ghent University

Nikolaos Lavidas

Aristotle University of Thessaloniki

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/bct.75

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2015014279 (PRINT) / 2015017629 (E-BOOK)

ISBN 978 90 272 4263 1 (HB)
ISBN 978 90 272 6828 0 (E-BOOK)

© 2015 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Preface	1
Reconstructing Proto-Indo-European categories: The reflexive and the middle in Hittite and in the proto-language <i>Paola Cotticelli Kurras and Alfredo Rizza</i>	7
The rise of ‘subordination features’ in the history of Greek and their decline: The ‘Indirect Speech Traits Cycle’ <i>Ioannis Fykias and Christina Katsikadeli</i>	29
Proto-Indo-European verb-finality: Reconstruction, typology, validation <i>Hans Henrich Hock</i>	51
Hittite <i>pai-</i> ‘go’ and <i>uwa-</i> ‘come’ as Restructuring Verbs <i>Bernhard Koller</i>	79
Reconstructing passive and voice in Proto-Indo-European <i>Leonid Kulikov and Nikolaos Lavidas</i>	101
Toward a syntactic phylogeny of modern Indo-European languages <i>Giuseppe Longobardi, Cristina Guardiano, Giuseppina Silvestri, Alessio Boattini and Andrea Ceolin</i>	125
Subject index	157

Preface

The last decades have been marked by an increasing interest in the study of the archaic syntax of Indo-European languages and, gradually, in the reconstruction of the main features of Proto-Indo-European (PIE) syntax. Suffice it to mention, among many others (in chronological order) Watkins (1964, 1976), Lehmann (1974), Kortlandt (1983), Hettrich (1990), Giannakis (1997), Bauer (2000), Luraghi (2004, 2012), Boley (2004), Barðdal & Smitherman (2009), Barðdal & Eythórsson (2012), and Barðdal (2013), adding much to our knowledge based on such seminal works on ancient Indo-European syntax as Delbrück (1893–1897) and Hirt (1934–1937). Although for some scholars the very possibility of syntactic reconstruction remains dubious, numerous studies have appeared reconstructing a variety of basic elements of Proto-Indo-European syntax based on evidence available particularly from ancient and/or archaic Indo-European languages. Such aspects of the proto-language as ergative/active alignment, basic word order, and subject and object marking have given rise to lively discussions among Indo-Europeanists, typologists and syntacticians. Furthermore, the possibility of an efficient lexically blind system of syntactic comparison, the parametric comparison method, was first suggested in Longobardi (2003), Guardiano & Longobardi (2005).

These and other related subjects constitute the range of topics that are addressed in the chapters collected in the present volume. These chapters were first published in the special issue of the *Journal of Historical Linguistics* 3:1 (2013): “*Proto-Indo-European Syntax and its development*”. The special issue and, accordingly, the present book volume originate from the Workshop “PIE Syntax and its Development” held April 1–2, 2011 at the Aristotle University of Thessaloniki (Greece) and arranged as a part of the 20th International Symposium on Theoretical and Applied Linguistics.

The idea of this workshop was to bring together scholars interested in these and related problems and to shine new light on current research into ancient Indo-European syntax. Special attention was paid to the development of the hypothetical reconstructed features within the documented history of Indo-European languages. Accordingly, the issues addressed in the papers presented at the workshop and now collected within the present book include the following:

- Is syntactic reconstruction possible?
- Which syntactic features can be reconstructed?
- Modern approaches to the analysis of the archaic Indo-European syntax and to syntactic reconstruction
- Proto-Indo-European syntactic reconstruction and its contribution to linguistic theory
- Transitivity, voice, middle, stative and related categories in early/late Proto-Indo-European and ancient Indo-European languages
- Word order and its evolution in Proto-Indo-European and Indo-European
- Relative clauses and other types of subordinate clauses in Proto-Indo-European and their evolution
- Syntax of non-finite forms (infinitives, converbs, etc.)

The contributors offer a systematic presentation of many aspects of Proto-Indo-European syntax and its development, such as reflexivity and middle voice, in particular, in Hittite and Baltic languages (Cotticelli Kurras & Rizza), indirect speech in Greek and the dichotomy between main and subordinate clauses in Indo-European (Fykias & Katsikadeli), word order and verb-finality in Proto-Indo-European (Hock), phraseological verbs in Hittite and a restructuring approach (Koller), encoding the passive pattern in Indo-European languages and reconstruction of voice in Proto-Indo-European (Kulikov & Lavidas), and the historical classification of Indo-European languages based on syntactic structures (Longobardi et al.).

Cotticelli Kurras & Rizza explore the typological parallelism between the constructions employed to express the category of the reflexive in Hittite and the development of a verbal strategy to mark reflexivity in the Baltic languages. They claim that the set of endings of the Anatolian middle displays the reflexes of the reconstructed endings of the ‘stative’, while the endings of the middle in other Indo-European languages represent an innovation.

Fykias & Katsikadeli analyze the emergence and diachronic development of subordination phenomena in the domain of finite and infinitival clauses that serve as complements of verbs of speaking and knowing. They focus in particular on the development and distribution of subordinating conjunctions, person shift, mood shift, and tense shift in the history of Greek. The authors interpret these developments in light of contemporary syntactic theory.

Hock shows that there is strong evidence in favor of the hypothesis that the unmarked word order of Proto-Indo-European was verb final (*pace* Pires & Thomason 2008, who questioned the fruitfulness of Indo-European syntactic reconstruction): the earliest attestation of Indo-European languages overwhelmingly have a verb-final order; the Subject-Object-Verb (SOV) reconstruction receives

support from phonological and prosodic changes in finite verbs. Hock argues, however, that Proto-Indo-European was not a “strict SOV” language: SOV languages only have the order Relative Clause (RC) — Correlative Clause (CC) but for Proto-Indo-European we must reconstruct both RC–CC and CC–RC orders; all of the early Indo-European languages can have a verb-initial word order.

Koller provides a new syntactic analysis of the Hittite phraseological construction involving the verbs *pai-* ‘go’ or *uwa-* ‘come’ and a second finite verb. He proposes that a restructuring approach to the phraseological construction in Hittite can account for most of its major features: the position of the enclitics, the restriction to a single subject, and the lack of an overt complementizer introducing the embedded clause.

Kulikov & Lavidas examine various aspects of the reconstruction of the passive in Proto-Indo-European, foremost on the basis of evidence from Indo-Aryan (Early Vedic) and Greek branches. The authors analyze the contrast between non-specialized and specialized markers of the passive in Early Vedic and Greek. They furthermore show that the two branches, Indo-Aryan and Greek, instantiate two basic types of the development of transitivity oppositions and system of voices: the syncretic type found in many Western branches, including Greek, and the anti-syncretic type attested in some Eastern branches, in particular in Indo-Aryan.

Longobardi et al. apply various quantitative tools to parametric data, providing additional evidence for traditional genealogical groupings and confirming the unstable character of taxonomies that previous comparative methods have challenged. Furthermore, the authors focus on and account for two peculiar instances within the (traditional) genealogical classification of Indo-European languages: the outlier position of Bulgarian within Slavic and the peculiar position of Farsi within Iranian.

A number of chapters not only discuss evidence from ancient and modern Indo-European languages, but also offer valuable observations and generalizations related to the historical linguistics and language change (see Lavidas 2009 on the relation between modern syntactic approaches and Indo-European historical linguistics). The editors are very confident that a diachronic typological approach can be successfully applied to a number of linguistic categories (see Kulikov 2010 for details of such an approach, guidelines, and a tentative typological questionnaire that can be used for a coherent study of transitivity and valency-changing categories in a historical perspective). This approach, which is still in its infancy, will undoubtedly reveal new features and aspects of linguistic phenomena that were not systematically studied in earlier scholarship. As one of the guest-editors of this issue has argued (Kulikov 2010), further research in this domain should be based on evidence from languages or language families that are textually well-documented over a sufficiently long period of time (around 1000 years or more).

Consistently and coherently presented diachronic evidence can further be used for outlining some sort of family (group) ‘portrait’, or profile, of a linguistic category in this language group or family (for instance, a portrait of voice or case in Indo-European, Germanic, or Semitic etc.), tracing it from the earliest attested ancient languages or reconstructable proto-languages (Proto-Indo-European, Proto-Germanic, Proto-Romance / Vulgar Latin, or Proto-Semitic, etc.) onward to its reflexes in the daughter languages.

Altogether, diachronic typology will open new perspectives for research in the field of historical linguistics, linguistic typology and syntax, thereby providing new insights for Indo-European reconstruction and the study of Proto-Indo-European and other proto-languages.

We would like to thank John Benjamins Publishing Company, which selected the special issue on *Proto-Indo-European Syntax and Its Development* for inclusion in the series “Benjamins Current Topics”. We sincerely thank the editors of the *Journal of Historical Linguistics*, Jóhanna Barðdal and Silvia Luraghi, as well as Anke de Looper, an acquisition editor of John Benjamins Publishing Company, for their valuable help in the preparation of the special issue. We also thank Gabriel Bakkum, Jóhanna Barðdal, Joseph Emonds, Thórhallur Eythórsson, Alice Harris, Dag Haug, Brian Joseph, Götz Keydana, Jared Klein, Thomas Krisch, Silvia Luraghi, Gerhard Meiser, Johanna Nichols, Elisabeth Rieken, Anna Roussou, Andrej Shatskov, Ioanna Sitaridou, and four reviewers who wish to remain anonymous, for their contribution to reviewing papers submitted to the special issue of *JHL*.

April 2013 – March 2015

Leonid Kulikov and Nikolaos Lavidas

Ghent University, Aristotle University of Thessaloniki

References

- Barðdal, Jóhanna. 2013. Construction-Based Historical-Comparative Reconstruction. *Oxford Handbook of Construction Grammar* ed. by Thomas Hoffmann & Graeme Trousdale, 438–457. Oxford: Oxford University Press.
- Barðdal, Jóhanna & Thórhallur Eythórsson. 2012. Reconstructing Syntax: Construction Grammar and the Comparative Method. *Sign-Based Construction Grammar* ed. by Hans C. Boas & Ivan A. Sag, 257–312. Stanford: CSLI Publications.
- Barðdal, Jóhanna & Thomas Smitherman. 2009. Typological Changes in the Evolution of Indo-European Syntax? *Diachronica* 26:2.253–263. DOI: 10.1075/dia.26.2.04sam
- Bauer, Brigitte L.M. 2000. *Archaic Syntax in Indo-European: The Spread of Transitivity in Latin and French*. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110825992

- Boley, Jacqueline. 2004. *Tmesis and Proto-Indo-European Syntax*. Innsbruck: Institut für Sprachen und Literaturen.
- Delbrück, Berthold. 1893–1897. *Vergleichende Syntax der indogermanischen Sprachen*. Strassburg: Trübner.
- Giannakis, Giorgios K. 1997. *Studies in the Syntax and Semantics of the Reduplicated Presents of Homeric Greek and Indo-European*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Guardiano, Cristina & Giuseppe Longobardi. 2005. Parametric Comparison and Language Taxonomy. *Grammaticalization and Parametric Variation* ed. by Montserrat Batllori, Maria-Lluïsa Hernanz, Carme Picallo & Francesc Roca, 149–174. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199272129.003.0010
- Hettrich, Heinrich. 1990 [1991]. *Der Agens in passivischen Sätzen altindogermanischer Sprachen*. Göttingen: Vandenhoeck & Ruprecht.
- Hirt, Hermann. 1934–1937. *Indogermanische Grammatik*. Vols. VI–VII: *Syntax*. Heidelberg: Winter.
- Kortlandt, Frederick H.H. 1983. Proto-Indo-European Verbal Syntax. *Journal of Indo-European Studies* 11.307–324.
- Kulikov, Leonid. 2010. Bridging Typology and Diachrony: A Preliminary Questionnaire for a Diachronic Typological Study of Voice and Valency-Changing Categories. *Problemy grammatiki i tipologii: sbornik statej pamjati Vladimira Petroviča Nedjalkova (1928–2009)* [Issues in Grammar and Typology: A Memorial Volume for Vladimir Nedjalkov] ed. by Valentin Vydrin, Sergej Dmitrenko, Natalja Zaika, Sergej Saj, Nina Sumbatova & Viktor Xrakovskij, 139–163. Moscow: Znack.
- Lavidas, Nikolaos. 2009. *Transitivity Alternations in Diachrony: Changes in Argument Structure and Voice Morphology*. Newcastle: Cambridge Scholars Publishing.
- Lehmann, Winfred P. 1974. *Proto-Indo-European Syntax*. Austin: University of Texas Press.
- Longobardi, Giuseppe. 2003. Methods in Parametric Linguistics and Cognitive History. *Linguistic Variation Yearbook* 3.101–138. DOI: 10.1075/livy.3.06lon
- Luraghi, Silvia. 2004. Null Objects in Latin and Greek and the Relevance of Linguistic Typology for Language Reconstruction. *Proceedings of the 15th Annual UCLA Indo-European Conference* (= JIES Monograph, 49) ed. by Karlene Jones-Bley, Martin Huld, Angela Della Volpe & Miriam Robbins Dexter, 234–256. Washington, DC: Institute for the Study of Man.
- Luraghi, Silvia. 2012. Basic Valency Orientation and the Middle Voice in Hittite. *Studies in Language* 36.1.1–32. DOI: 10.1075/sl.36.1.01lur
- Pires, Acrisio & Sarah G. Thomason. 2008. How Much Syntactic Reconstruction is Possible? *Principles of Syntactic Reconstruction* ed. by Gisella Ferraresi & Maria Goldbach, 27–72. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.04pir
- Watkins, Calvert. 1964. Preliminaries to the Reconstruction of Indo-European Sentence Structure. *Proceedings of the 9th International Congress of Linguists, Cambridge MA* ed. by Horace G. Lunt, 1035–1045. The Hague: Mouton.
- Watkins, Calvert. 1976. Towards Proto-Indo-European Syntax: Problems and Pseudo-problems. *Papers from the Parasession on Diachronic Syntax* ed. by Sanford B. Steever, Carol A. Walker, & Salikoko S. Mufwene, 305–326. Chicago: Chicago Linguistic Society.

Reconstructing Proto-Indo-European categories

The reflexive and the middle in Hittite and in the Proto-language*

Paola Cotticelli Kurras and Alfredo Rizza

University of Verona & University of Würzburg

Starting from the analysis of constructions employed to express the category of reflexive in Hittite, encoded both by the verbal ending set of the middle and by the pronominal marker *-za* with both active and middle verbal forms, we present a typological parallelism with the Baltic languages that has consistently developed, from a pronominal, a verbal strategy to mark reflexivity. It is also shown that a development regarding the ways of encoding reflexivity involve other Indo-European languages as well.

The Anatolian languages attest the reflexes of the original set of endings referring to the semantic categories of Reflexive, Middle and “Resultative”, while the other Indo-European languages attest an innovated “mixed morphology” for the category of Middle and Reflexive as opposed to the proper endings of the historical perfect. Within such a theoretical framework, the development of alternative strategies, using pronominal devices or particles, aims to disambiguate a wide polysemous ending set. A ‘Wackernagel’ (2P) particle in Hittite, namely *-z*, is particularly active in disambiguating reflexivity. Lithuanian *-si*, an original pronoun that developed at first into a 2P particle and subsequently into a verbal suffix, extends its functional field and takes over the place of the original middle, as in other Baltic and Slavonic languages.

1. Introduction

Recent studies consider similar semantic extensions for the categories of middle voice (Kemmer 1993) and reflexive (Geniušienė 1987). Thus the so-called ‘middle

* Although the authors prepared this article together, discussing all topics treated, the sections of this article can be assigned as follows: P.C.K.: Sections 1 and 5; A.R.: Sections 2 and 3; P.C.K. and A.R.: Sections 4 and 6.

voice' and the 'reflexive' may be different labels referring to morphological means coding a totally or partially coinciding functional field. Speaking of the functional field that is typical of the reflexive and/or the middle voice, we can see that reflexivity can be explained as one semantic realization within the Middle, others being, e.g., passive, impersonal, decausative (Kemmer 1993). The Middle is, in this present article, a metalinguistic expression used in contrast to 'middle voice'. We will in fact refer, with 'Middle', to a wide field of semantic categories not necessarily coded by a specific morphological *facies*.

From the morphological perspective, Hittite and Lithuanian present typologically similar solutions, i.e. the particles *-z* and *-si* respectively, in order to express the semantic category 'reflexivity', which can be understood as a sub-component of the wider semantic field 'Middle'. Nevertheless, the Middle, as a semantic field, can be encoded in Indo-European languages through different morphological patterns of endings that do not go back to a single original set in the proto-language. Our present analysis shows a situation in which the semantic extensions of the morphological encodings of the Middle do not match in the different languages. Hence we attempt to reconstruct a previous linguistic stage within Proto-Indo-European on the basis of the attested material in the historical languages. The process of assignment of the semantic functions can only be achieved through the analysis of the mapping of the arguments onto the grammatical functions in the original proto-inflection, which appears to be shared by the historically attested morphological categories of perfect, middle voice and reflexive.

The structure of this article is as follows: in Section 2 we treat the problem of the relationship between reflexive and middle morphology; Section 3 is devoted to the Hittite particle *-z* in comparison to Lithuanian *-si*. Once we have outlined the semantic extension of the two and where they are comparable, we analyze, in Section 4, the extension of similar pronominal devices in other Indo-European languages, adding the data from Hittite to the work by Haspelmath (2003). In Section 5 we turn to the problem of the origin of the middle in Proto-Indo-European and we summarize our results in Section 6.

2. Reflexivity and the Middle in IE languages

The tradition in Indo-European studies recognizes a reconstructed specific set of pronouns simply called the 'reflexive pronoun', but reflexivity was also typically expressed by the set of verbal endings traditionally labeled 'middle voice'.¹

1. Hackstein (2003) reconstructs a reflexive accented pronoun **s-wé* (accusative) for PIE, which is homophonous with the personal unaccented pronoun **se* (accusative). Some inflectional

Other strategies for encoding reflexivity are also available in IE languages as, for example, in Hittite, which developed the use of a specific particle. As in Celtic (see TDIR; especially for Middle Welsh see Parina 2007; on the typology of IE middle voice see Meiser 2009) and other IE languages (Old/Middle English), in Hittite the reflexive participant could be marked by an anaphoric pronoun (see König & Gast 2002).

- (1) Clitic pronouns as reflexive markers in Hittite
 KUB 16, 16 i 28 (NH, SU (Liver)-Oracle II)
 warpanzi=ma=wa=šmaš UL
 wash.ACT.PRS.3PL=but=QUOT=them not
 ‘they don’t wash them(selves)’

Metaphorical periphrases are also attested in Hittite, but it is the particle *-z* that is commonly associated with a reflexive event, and resorting to middle voice endings eventually became not even the primary strategy.

- (2) Nominal periphrases as reflexive markers in Hittite
 KBo 4, 4 ii 49 (NH, Muršili II annals, 9th year)
 nu=za mahḫan kun memian
 CONN=REFL when this.ACC.SG word.ACC.SG
 ZI-ni EGIR-pa kiššan AQ-BI
 soul.DAT.SG back.ADV in this way.ADV re-called
 ‘and when I had recalled this word thus myself in this way’ (literally: in my
 soul, mind; by heart)

A recent study (Puddu 2005) hypothesizes that PIE actually lacked a reflexive pronominal marker: the absence of a reflexive pronoun in peripheral Celtic/Old English and in the oldest attested language (Hittite) may point to a reconstruction without a proper reflexive marker. This would explain why the different groups developed different solutions beginning from a set of potentially adequate morphemes. We will discuss here some problematic data in attempting to provide a new interpretative model.

The following functions have been reconstructed for the IE middle voice in the IE languages (cf. Clackson 2007: 142–143):

- Personal involvement (Old Indic *yájate* ‘(s)he performs the sacrifice’, “said of person for whose benefit the sacrifice is made”, Clackson 2007: 143)

forms like **-soi* are analyzed both as 3.sg. anaphoric personal and reflexive pronouns. The reflexive pronoun is lost in many IE languages and is replaced by pronouns that specify person and number, like personal pronouns or other suppletive markers, such as Sanskrit *ātman* “self”, Greek *autós* or German *selbst* (Hock 2006, Hackstein 2003: 73f.). Regarding Hittite *-z* see Boley (1993), Hoffner & Melchert (2008).

- Reflexivity (Greek *louomai* ‘I wash myself’)
- Reciprocity (Hittite *ŠU-za appantat* ‘they took each other by the hand’)
- Passive as a later development (e.g. in Latin, Gothic, but also Greek and Anatolian)

Traditionally, we can define the Middle as the situation in which the subject is in some way affected by the event (Lazzeroni 1990, Benedetti 2005). The middle voice was only preserved in Anatolian, Greek, Latin, Tocharian, Indo-Iranian, Germanic (Gothic only), and Celtic, but in the individual languages peculiar functions arose, as is, e.g., the case for Latin and Gothic middle forms, which encode passives, except for a few deponent verbs. In contrast, Ancient Greek, Tocharian, Hittite and Indo-Iranian preserve a functional difference or productive opposition between the active and middle voices.

Reflexivity can be achieved by either a pronominal or a verbal strategy (Kazenin 2001). Pronominal strategies in IE languages may be the following (adapted from Petit 1999 and Puddu 2005; RM = ‘reflexive marker’).

Table 1. Pronominal strategies for reflexivity in the ancient IE languages

RM [+person]	a.	full personal, distinguishing 1-2-3 persons, e.g. Ancient Greek
	b.	3rd reflexive, contrasting the 3 person alone with a specific marker, e.g. Latin <i>se</i> , Homeric <i>ἐ</i>
RM [-person]	a.	generic pronoun for all persons, e.g. OCS <i>sebe</i> , <i>svoj</i> , Lith. <i>savė</i> , <i>sāvas</i>
	b.	particle, e.g. Hittite <i>-z</i>

Verbal strategies, on the other hand, are typically represented by middle voice morphology, like Greek *louomai* ‘I wash (myself)’. The two strategies are not mutually exclusive. As a matter of fact, we sometimes observe the overlapping of the two, as in the case of morphologically middle verbs with (functionally) reflexive pronouns or (functionally) reflexive particles. See the Hittite examples in (7, 10, 16–17, 19), all with verbs with middle morphology co-occurring with the particle *-z*.

So, in order to express reflexivity (in a broader sense) we may find three strategies in a single language.

Table 2. Combinations of pronominal and verbal strategies

RM	middle voice
–	+
+	+
+	–

It seems reasonable to believe that the overlapping of the two strategies captures a diachronic process that we tentatively “reconstruct” for the PIE diachrony as follows:

Table 3. The diachronic development of the combinations of pronominal and verbal strategies

	Reflexivity		
	Late PIE	oldest attested IE lgs.	later attested IE lgs.
middle voice	+	+/-	–
RM	–/+	+	+

The stages of the process would be as follows:

- Late PIE: The middle voice covers the entire spectrum of reflexivity, but some reflexive functions receive extra marking.
- Oldest attested IE: The new, originally additional, markers undergo processes of grammaticalization extending their spectrum and starting to appear without the middle voice (Anatolian, Ancient Greek). The single languages develop individually, but they follow similar paths.
- Later attested IE: The middle voice is lost in most languages.²

One of the individual paths of development is characterized by the use of a second position, or ‘Wackernagel’ particle (i.e., a [–person] RM). We find a reflexive particle without person features in the following languages: Hittite (Wackernagel clitic); OCS and, later, Serbian, Croatian, Russian and other Slavic languages; Lithuanian and Latvian.³ We find the situation of the Baltic languages to be of particular interest, especially Lithuanian and Latvian, where the historical attestations of the reflexive marker may suggest a development from a Wackernagel clitic to a ‘Wackernagel’ verbal suffix and, finally, to a typical verbal suffix (see Nevis & Joseph 1992; on clitics in Baltic, see Petit 2010: 261–307). This development may be of interest if compared with the situation of the Hittite Wackernagel particle *-z*.

In this article we will introduce a classification of the use of Hittite *-z* according to the typology offered in Geniušienė (1987), where Lithuanian and Latvian

2. For Albanian cf. Orel (2000: 213), “In the present of the mediopassive, the 1–3sg. and 3pl. directly and indirectly continue the Indo-European middle. The Albanian phonology does not allow us to find out whether the Proto-Albanian paradigm used primary or secondary endings of the middle.”

3. Luvian also has second position elements, but these are pronouns. They are, however, neither ‘full personal’ nor generic: the 2nd and 3rd persons fall together. For a diachronic explanation and for some possible connections to Hittite *-z*, see Yakubovich (2010: 161–195).

are conveniently used to exemplify the various functions of a morphological verbal strategy that codes a typically reflexive and middle semantics. Geniušienė (1987) uses ‘reflexivity’ as a term that covers functions that are typologically established as the typical functions of the middle voice in Kemmer (1993). The Hittite particle *-z* and the Lithuanian and Latvian *-si* may describe a path of linguistic development which we believe is attested in divergent ways in IE languages. The Hittite data should describe the beginning of the process while the Lithuanian and Latvian show its final stage. We will mainly refer to Lithuanian for practical reasons, using the examples already presented by Geniušienė (1987) to illustrate the subdivisions of reflexivity. Etymologically and functionally similar devices are also present in the other Baltic and Slavic languages, but we leave them aside, since the examples taken from the Baltic languages in Geniušienė (1987) suffice to justify our proposal to categorize the Hittite evidence.

3. Focusing on Hittite *-z*

In Cotticelli Kurras & Rizza (2011), we attempted to explain the function of the Hittite particle *-z*. On the one hand, we described the various functions of *-z* through the scalar interpretation of the different semantic nuances of the Middle (Lazzeroni 1990).⁴ On the other hand, we defined the particle as a function of dependency of identity roughly along the lines of Safir (2004). Here we want to compare the functional range of *-z* with that of similar morphemes in other IE languages. In particular we will consider the Baltic IE languages for two main reasons: first the Lithuanian and Latvian morphemes *-si* have many features in common with *-z* (most importantly the lack of person distinction, [–person]), and, second, there is a rich body of typological work about reflexives, mainly based on Baltic languages, providing a ready-to-use and organized collection of data (Geniušienė 1987).

3.1 Functions of the middle/reflexive *-z*

In the following sections we organize the Hittite data along similar lines to the ones provided in Geniušienė (1987).

4. For a typological approach to the particle *-z*, cf. Josephson (2003); here we try to capture other aspects using different tools from a different perspective.

3.2 The data

3.2.1 First group: Transitive reflexive

In “transitive reflexive” constructions, subject and object have different referential identities. The subject, however, is also somehow involved in the effects of the event. A typical example may be a situation in which the action of the subject bears positive or negative effects on the subject itself.

When a reflexive marker is used to convey this involvement we represent it as follows:

S	O	IO
NP _i	NP _{ii}	RM _i

A Hittite example is the predicate *-z huittiya-* ‘to let go’ with active voice.⁵ The predicate *huittiya-* does not usually appear with *-z*. When it does, *-z* marks the pertinence of the subject in a transitive construction with an expressed direct object, as in (3).⁶

- (3) KBo 5, 8 ii 29–31 (NH, Muršili II annals, 20th year)
 antuḥṣatar=(r)a=z=kán kuit tēpawaz (30) anda
 humanity=CONN=REFL=PART REL ADV ADV
 huittiyan ḥarta nu ŠA KUR palā
 let go.PTCP.N AUX.ACT.PRET.3SG CONN of land Pala
 (31) ANA [^LUKÚR pa]rā ŪL kuitki pešta
 to enemy ADV NEG anything.N give.ACT.PRET.3SG
 ‘And (regarding) the people that he [Ḫutupiyanz] had let go in few groups he delivered nobody to the enemy from Pala.’

The predicate *huittiya-* also appears in an intransitive construction with *-z* as a dative *commodi*.

- (4) KBo 5, 3 ii 28–29 (NH, Ḫuqqana treaty)
 nu=mu=ššan măn ÉRIN^{MEŠ}-it ANŠE.KUR.RA^{HI.A}
 CONN=1SG.me=PART if troops-INS horses
 lammar UL ārti (29) nu=z=(š)an
 hour/time NEG come (to).ACT.PRS.2SG CONN=REFL=PART
 parā imma huittiyaši
 ADV ADV delay.ACT.PRS.2SG
 ‘if you do not immediately come to my [aid] with troops (and) horses, but you rather prolong [your coming to help] in your own interest’⁷

5. Cf. HW², s.v. *huittiya-* (III/2, Lfg. 19).

6. Capital OH, MH or NH refer respectively to Old, Middle or New Hittite.

7. CHD P, s.v. *parā*, p. 119b.

The predicate *-z ḥuttiya-* also has a reflexive auto-causative meaning.

- (5) KUB 19, 23 Vs. 9 (NH, letter)
 nu=z ANA PANI EN-YA kuit šara
 CONN=REFL to presence lord-my SUB.CONN ADV
 ḥuittiyan ḥarkun
 pull.PTCP.SG.N AUX.PRET.1SG
 ‘given the fact that I had pulled up myself before my lord’⁸

For a comparison with Lithuanian we give the following example.

- (6) Geniušienė (1987: 135) ex. 117b
 Jonas at-si-vedė vaiką į mokyklą
 Jonas-NOM pref-REFL-brought child to school
 ‘Jonas brought the child with him to the school’

Geniušienė (1987: 135) specifies that with the reflexive *-si* the sentence “implies that the action is performed by the Agent for his own sake”.

Transitive reflexives of the kind “possessive/body action” may also be listed in this category. Hittite examples are: *-z ḥatta-* ‘hit’, with the middle voice (or *-z arra-* ‘wash’, with both the middle and the active voice). Hittite *ḥatta-*, ‘hit’ is attested with the particle *-z* functioning either as direct object reflexive or a possessive reflexive.⁹

- (7) KUB 1, 14 ii 8–12 (NH copy of a festival fragment CTH 744)
 EGIR-ŠU=ma LÚ^{MES} zilipuriyatallaš (9) uwanzi
 ADV=CONN men zilipuriyatallaš come.ACT.PRES.3PL
 nu=za 6-ŠU (10) walḥanzi 1-aš=za=kan (11)
 CONN=REFL six-times beat.PRES.3PL one=REFL=PART
 ŠU-ZU ḥatta 1-aš=ma=za=kan (12)
 hand-his strike.MID.PRES.3SG one=CONN=REFL=PART
 GİR-ŠU ḥatta
 foot-his strike.MID.PRES.3SG
 ‘afterward the two priests of the God Zilipuri come; they beat themselves six times: one beats his own hand, the other one his own foot’¹⁰

Lithuanian can have a valence reduction or a possessive-reflexive.

8. CHD Š, s.v. *šarā*, p. 214a.

9. Cf. HED, s.v. *hat(t)-* and HW², s.v. *ḥatta(i)-* for forms and other meanings.

10. A different translation is found in Neu (1968: 52).

- (8) Geniušienė (1987: 140, ex. 125a–b)
 On-a prausia-si
 Ann-NOM washes-REFL
 ‘Ann washes her face’ / ‘Ann washes herself’
- (9) Geniušienė (1987: 140, ex. 125c)
 On-a prausia-si veid-ą
 Ann-NOM washes-REFL face-ACC
 ‘Ann washes her face’

Geniušienė (1987: 140) explains the semantic ambiguity of her examples (125a–b) by claiming that ‘Ann washes herself’ derives via reduction from sentences like *Ona prausia-si veidą* ‘Ann washes her face’ (125c) and takes on the meaning ‘Ann washes her face’, i.e. ‘Ann washes (something of) herself (most typically the face)’.

3.2.2 Second group: Subjective reflexive

In “subjective reflexive” constructions, the subject and the object either share the same referential identity, or the identity of the object is cancelled, achieving a reduction in the valence of the verb. We recognize four subtypes: “true reflexive”, “absolute”, “autocausative”, and “reciprocal”.

As the first subtype we treat the “true reflexives”, Geniušienė’s (1987) “semantic reflexives”, where subject and object share the same referential identity.

S	O
NP _i	RM _i

Hittite *arra-* ‘to wash’ can receive a reflexive interpretation in the middle voice, with or without *-z*. With *-z* it is also attested with the active voice.

- (10) KUB 59, 66 iii 7–8 (NH, Elkunirša and Ašertu)
 nu=za kēdanta w[etenit] (8) arrattaru
 CONN=REFL this.INS water.INS wash.MID.IMP.3SG
 ‘and with this water he has to wash himself’

For Lithuanian cf. (8) above, i.e. Geniušienė (1987: 140, ex. 125b) *Ona prausiasi* ‘Ann washes herself’.

We consider now the second subtype of subjective reflexive, the “absolutes”, where the object is cancelled and a reflexive marks the construction.

S	O	–
NP	–	RM

of an antipassive constructions”.

present article.¹¹

(=z) *haššik*- ‘to satiate oneself’ that may be rather ‘true’ reflexives.

- (11) KUB 33, 24+33, 29 i 17–18 (NH, disappearance of the Storm God)¹²

pertinent concept here.

12. With exemplar CTH 325.C, KUB 33.29+ (cf. hethiter.net/: CTH 325 [3/2011]).

In Geniušienė (1987), “absolute” reflexives refer to the usual transitive construction with an animate object without any possessive relationship with the subject, e.g. in Lithuanian as the typical counterpart of a transitive with animate direct object.

- (12) Geniušienė (1987:84, ex. 28b)

šu-o kandžioja-si
dog-NOM bites-REFL
‘the dog bites’

- (13) Geniušienė (1987:84, ex. 28a)

šu-o kandžioja vaik-us
dog-NOM bites child-ACC.PL
‘the dog bites children’

In example (11) with =*z ed-*, one could also consider a sort of intensifying nuance to -*z*: ‘they did eat and they were not satiated’, but we believe that, at this stage of research, it is better to understand -*z* as marking the action as performed for the subject’s sake, coherently with the other functions of the particle.

The third subtype of “subjective reflexives” is the “auto-causative reflexive”. The most interesting example of the “auto-causative reflexive” in Hittite is certainly the verb *eš-* ‘sit, sit down, place’. Hittite texts show four different constructions:

1. Intransitive active voice, ‘to sit’, equivalent to German *sitzen*, from OH
2. Intransitive middle voice, ‘to sit down’, equivalent to German *sich setzen*, from OH
3. Transitive middle/active, ‘to place’ (generally with -*z*), from the beginning of NH¹³
4. Intransitive middle with -*z*, equivalent to German *sich setzen*
5. Transitive active with -*z*, equivalent to German *sich setzen*, *Platz nehmen*

Regarding the use with -*z*, it is possible to trace a diachronic development from the middle voice without -*z* in older Hittite times down to the active voice with -*z*. Consider the following examples.

- (14) KUB 2, 2 iii 58 (NH copy of OH Hattic-Hittite bilingual, CTH 725)

DINGIRŠulinkattiš=šan LUGAL-uš anda ešzi
Šulinkattiš=PART king-SG.NOM ADV sit.ACT.PRES.3SG
‘Šulinkatte the king will sit (stay) there’

13. HW², s.v. *eš-/aš-*(2), p. 109: “Eindeutig datierbar ab Muršili II.”

- (15) KBo 3, 1+ ii 16 (NH, copy of OH Telipinu Edict)
 mān=šan ^mTelipinuš INA ^{GIŠ}GU.ZA ABI=YA ēšhat
 when=PART Telipinuš on throne father=my sit.MID.PRET.1SG
 ‘When I, Telipinu, sat down on the throne of my father’
- (16) KBo 3, 4+ iii 57–58 (NH, Mursili II annals, 5th year, Neu 1968: 28)
 nu=za ANA PANI ABI ABI=YA (58) kuiš URU
 CONN=REFL to presence grandfather=my REL city
 Gašgaš ^{HUR.SAG}Tarikarimun GĒŠPU-az ešat
 Kaskean (mountain) Tarikarimu forcefully-ABL settle.MID.PRET.3SG
 ‘[...] the Kaskean city that at the presence of my grandfather forcefully
 occupied (the mountain) Tarikarimu’
- (17) KBo 3, 4+ i 5 (*ibid*, throne accession, Neu 1968: 27)
^mArnuantaš=ma=za=kan ŠEŠ-YA ANA
 Arnuwantaš=CONN=REFL=PART brother=my on
^{GIŠ}GU.ZA ABI-ŠU ešat
 throne father-his sit.MID.PRET.3SG
 ‘My brother Arnuwanda sat down on the throne of his father’

As an example of the autocausative in Baltic, we give here one from Latvian.

- (18) Geniušienē (1987: 86, ex. 31b)
 Es spiežo-s pie sien-an
 I.NOM press-REFL to wall-GEN
 ‘I flatten myself against the wall’

The final subtype within “subjective reflexive” is the “reciprocal”.

- (19) KUB 12, 26 ii 1–3 (NH, ritual of the river CTH 441)¹⁴
^{DINGIR}UTU-uš=wa ^{DINGIR}Kamrušipaš=ša UDU^{HLA}
 (god)Sun-NOM.SG=QUOT (god)Kamrušepa=CONN sheep
 kiškanzi (2) nu=war=at=za=kan (3)
 comb.ACT.PRES.3PL CONN=PART=PRON.3PL.S=REFL=PART
 ḥanna(n)tati nu=war=at=za=kan
 contend.MID.PRET.3 CONN=PART=PRON.3PL.S=REFL=PART
 šāntati
 enrage.MID.PRET.3PL
 ‘The Sun-God and Kamrušepa are combing the sheep. They contended and
 raged with each other.’

14. Cf. Archi (1993: 406–407). The text is interesting because it has two instances of reciprocals. However, the semantics of the two verbs here is not totally clear, especially in connection with what follows. We lack the preceding context. Cf. Benedetti (1981), but see Friedrich (1930: 71–72, n. 2).

The verbs *hanna-* ‘to contend’ and *ša-* ‘to be(come) angry’ are here third plural present middle forms with a typical *-ti* (/di/) extension.

For Lithuanian, see the following example.

- (20) Geniušienė (1987:89, ex. 32c)
 Petr-as ir On-a (jiedu) bučiuoja-si
 Peter-NOM and Ann-ACC (they two) kiss-REFL
 ‘Peter and Ann (they) kiss (are kissing).’

As far as the marking of “subjective reflexives” with a particle lacking person features is concerned, we are able to provide examples both from Hittite and from Baltic. From now on, as we will see, the situation changes.

3.2.3 *Third group: Objective reflexive*

Different from the subjective reflexive, the subject in the objective reflexives has a transitive counterpart in the direct object. Objective reflexives thus share certain features with the passive derivation, but with objective reflexives the agent is not recoverable, as in the case of the passive. A reflexive marker marks the construction. The prototypical direct object of a transitive construction appears in the reflexive construction as the intransitive subject, while the original agent is either unknown or irrelevant and it can be stated only through some indirect object construction generally not identical with the expression of the agent in passive constructions.

While in the Baltic languages the objective reflexive is obviously well represented, in Hittite it is definitely not easy to find comparable examples involving *-z* as a proper marker. One typical Lithuanian example of the causative–decausative alternation can be taken from Geniušienė (1987).

- (21) Geniušienė (1987:98, ex. 46a)
 Petr-as / vėj-as at-idarė dur-is
 Peter-NOM / wind-NOM PREV-opened door-ACC.PL
 ‘Peter opened the door’
- (22) Geniušienė (1987:98, ex. 46b)
 Dur-ys at-si-darė
 door-NOM.PL PREV-REFL-opened
 ‘The door opened’

In Hittite, the causative–decausative (or anticausative) alternation has recently been studied in Luraghi (2010: 149): “In Hittite the semantic properties of middle voice are distributed between morphological middle and reflexive constructions”. For the particle *-z*, Luraghi (2010: 150) concludes: “[...] the Hittite data seems to point toward a division in the domain of middle voice: in particular, subject

involvement is connected with the particle *-z(a)*, while spontaneous events are connected with grammatical middle”.

So far we have seen that a similar element, namely a reflexive particle unmarked for person, has similar functional ranges within the framework of Geniušienė (1987). Hittite and Baltic, however, are only comparable as far subjective reflexives are concerned; they do not seem to travel identical paths with the objective reflexives. In the next section we will devote more attention to the semantic extensions of the reflexive marker in Hittite, in comparison to other IE languages.

4. Functional extension of reflexive constructions in IE languages

Below, we present a graphic representation of the functional extension that can be safely defined as typical, in Hittite, for the particle *-z(a)* and the middle morphology. The lines show that the middle voice covers presumably all the different types of reflexives, from ‘transitive reflexive’, through ‘subject reflexive’ down to ‘object reflexive’, i.e. the typical realm of the semantic Middle. Hittite *-z* appears to be more typically represented in the higher or active, section, stressing subject involvement. In the figures, TrR = ‘transitive reflexive’, SR = ‘subjective reflexive’, and OR = ‘objective reflexive’.

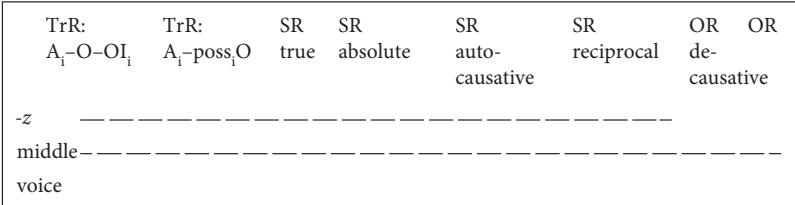


Figure 1. Comparative extension of *-z* and middle voice in Hittite

Complementing Haspelmath (2003), we can add the Hittite data collected here and hopefully improve the figure as follows:

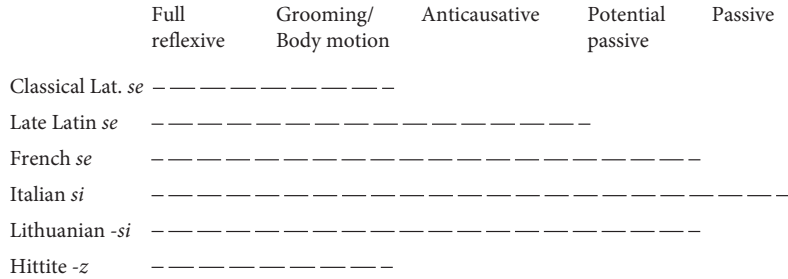


Figure 2. Comparative chart of functional extensions from reflexive to passive

Summing up the two preceding figures, we may observe that Classical Latin and Hittite show consistent use of the middle voice, which retains distinguishing function in the lower section of semantic field, which we call here ‘Reflexivity/Middle’. In Late Latin the middle voice is recessive, and in French and Italian it is lost.

We do not have certain knowledge about the etymological origin of Hittite *-z*, which may be different from that of Baltic *-si*, i.e., from the original third person reflexive. The two, however, seem to follow a very similar path of development compared with that of Classical Latin and the Romance languages. In particular, Hittite captures the initial development from a Wackernagel particle producing a competitor to the middle voice for the realization of a specific area of the realm of Reflexivity. Hittite does not show, however, the loss of the middle voice and the grammaticalization of the particle to a verbal suffix, i.e. the situation that we find in Lithuanian and Latvian. Putting together the evidence from the two languages, we may hypothesize a general tendency to develop from a pronominal strategy (both Hittite *-z* and, originally, Lithuanian *-si*) to a new verbal strategy (Lithuanian *-si*) in substitution for an older one, which have been lost (the original middle). The relevant steps of the process can be summarized as follows:

1. Starting point: The original Late PIE state of affairs encodes the Reflexive categories only through middle endings.
2. First stage of development: A ‘reflexive’ particle in Wackernagel position appears to an increasing degree in the last stage of PIE and in some of the attested IE languages, in sentences with a verb exhibiting middle voice morphology.¹⁵
3. Second stage of development: A reduction in the use of middle endings happens in individual languages, replaced by active morphology with the reflexive particle. Later, the particle undergoes a development from Wackernagel clitic to verbal affix.
4. Grammaticalization of the particle: middle endings are lost and the Wackernagel particle is reanalyzed as the marker of a specific new middle/reflexive voice (as in Lithuanian and Latvian).

5. Reconstructing the PIE category ‘Middle’

So far, in the present article we have investigated the relationship between the usage of the Hittite particle *-z* and the middle voice in the light of a typological and contrastive comparison with the Baltic ‘reflexive’ *-si*. This has allowed us to assess

15. These reflexive particles need not be etymologically identical in the individual languages and most probably the syntactic position is not relevant to the function, but may depend on a more general PIE rule.

the question of the development of linguistic strategies to mark the semantic field 'Middle'. As a consequence we found ourselves facing the question of the origin of the category 'Middle' in PIE, both from a functional and from a morphological point of view.

We assume, for our purposes, the "*Zeit-Raum*" model of Neu (1976, 1985) and Meid (1975). The origin of the middle voice, in this model, is directly placed, for its semantic values, in the realm of the original stative diathesis. Jasanoff's (2003) model, while sharing the hypothesis of the genesis of the Hittite "-*hi* conjugation" from the "**h*₂ endings set", argues differently with regard to the formal data.¹⁶

The theory of the Active/Stative opposition, at the base of the *Zeit-Raum-Model*, provides the PIE verbal system with two sets of original endings. An opposition in the basic semantics (inherent or primitive semantics) of the verbal roots is a good starting point for describing a possible formation of the category or diathesis of the Middle to account for all aspects formally attested in the historical languages. At this stage it is possible that the inherent meaning of a lexical root selects a dedicated set of endings for an event or a state in a broad meaning. We refer to this set as the "*-*h*₂ endings set" as opposed to the "*-*m* endings set".

With 'active' and 'stative', however, we do not refer, in this article, to the primitive inherent semantic of the predicates, but to two conjugations within the verbal inflexional system. The 'active', i.e. the '-*m* conjugation', while having 'actions' and 'causatives' as semantic prototypes, is by no means restricted to these two realizations. Rather it represents the unmarked set of endings that does not explicitly mark the grammatical subject as the place or the direction of the effects of the described event. The 'stative', i.e. the '-*h*₂ conjugation', on the contrary, explicitly marks the subject as being at different levels involved in the effects of the event. This set, in turn, is by no means restricted to pure 'states', but also encode 'decausatives' and 'spontaneous events', and even 'causatives', whenever the subject is in some way affected by the event.

On a comparative basis, there is no direct possibility of projecting the category of the middle voice back into the earliest phase of IE morphologically. Indeed, only the Anatolian languages preserve the core of the "*-*h*₂ set" of endings of the first and most ancient phase of PIE (PIE I). In the Anatolian languages, those endings designate the category of the middle, which may show, in addition, an extension of this *-*h*₂ set by an *-*r* morpheme only found in a certain linguistic area of other IE languages, also as a marker of the middle voice.

The evidence from the other IE languages with regard to the category Middle is varied, as this semantic category is represented by other morphemes. The "*-*h*₂ set"

16. See also Clackson (2007: 142ff.) for discussion of the two main positions on the possibility of reconstruction of the category Middle, both morphological and semantic.

of endings has a direct continuant in Greek and Indo-Aryan, but there it designates the category of the Perfect. The other languages, which did in fact have a category Middle, were forced to go through a systemic change and received other markers for the Middle, perhaps because they lost the laryngeals. This implies that the middle endings of the other IE languages had a different morphological origin from those in the Anatolian languages.

In Table 4 we attempt to present a structure which takes into consideration the morphological as well as the semantic components of the verbal categories.

Table 4. An outline of the formation of the Middle through the phases of PIE

I PIE	ACTIVE “*-m set”		STATIVE “*-h ₂ set” (stative, resultative)				
Ila PIE	Pres *-m-i	Past *-m	present/perfect		past/middle		
			*-h ₂ -e		*-h ₂ -o		
Iib PIE			*-h ₂ -e > Anat. -ha		*-h ₂ -o > Anat. -ha		
		(except Anat.): formation of the middle endings through mixed morphology de- riving from both the “*-m set” and the “*-h ₂ set”	(prob- ably except Anat.): Formation of the perf. (*-h ₂ e)	Anat. *-h ₂ -e +-i → Hitt. -hi conjug.	Anat. Hitt. middle -ha(ri)/ -ha(t)(i)	Anat. -m > Hitt. past -hun	Anat. Luv. past -ha
III PIE	Late PIE active	Late PIE middle (-passive)	Late PIE perfect				
	Pres 'Active', reflexive (with pron.)	Past 'Middle', reflexive, stative	Pres Past Perfect	stative, resultative			

We reconstruct a first phase of PIE which is characterized by two sets of endings designating a semantic opposition of the root meaning.

This was followed by an extension of the system through the introduction of the tense category of the present, which was marked by the agglutination of the morpheme *-i to the “*-m set” endings. Parallel to this, a split occurred in the “*-h₂ set”, which added a new opposition through an ablauting element *-e/o. This opposition may have denoted a differentiation in the semantics of the “stative”, and perhaps a tense distinction at the same time. Thus PIE Ila had a verbal system with a “tense” opposition that was characterized differently in the “*-m” and in the “*-h₂” set: *-m-i vs. *-m and *-h₂e vs. *-h₂o, respectively (Neu 1985).

In the following phase (PIE IIB) we can locate the origin of the new diathesis of the Middle and subsequently the branching-off of Anatolian languages; this phase is morphologically represented in the Anatolian verbal system only. The Anatolian languages indeed do not show any other tense–aspect–stem opposition, which is firmly entrenched in the other IE languages.

It is only the evidence of the Anatolian languages that lets us clearly understand the relationship between the middle voice and the original series “ $*-h_2$ ”, since in Hittite we find a core set of endings for the 1st, 2nd and 3rd person singular that unequivocally find their best *comparanda* in the endings of the Late PIE perfect, as seen in Greek and Vedic (Neu 1968); the laryngeal in Anatolian remained in the endings of the 1st *-ha-(ri)* of the middle and generally in the *-hi* conjugation. In other IE languages, the phonological changes concerning the laryngeals may have made it necessary to remodel the morphological means of encoding the semantics of the ‘Middle’.

Even in the Anatolian languages the endings containing the sequence $/*h_2e/$ and $/*h_2o/$ (i.e., 1sg. $*-h_2e$, $*-h_2o$ and 2sg. $*-th_2e$ and $*-th_2o$) phonetically produced the same result */ha/*. An ending *-ha* alone only survives in Luvian for the denotation of the past, whereas in Hittite the endings of the past as well as those of the middle were extended by further markers (such as *-n* or *-t(i)*, *-ri* respectively).

After phase II of PIE it is conceivable that the verbal system of the other IE languages underwent a restructuring process not shared by Anatolian. This also included the restructuring of the original “ $*-h_2$ set” through a contamination with the “ $*-m$ set” for the designation of the (new) category of the middle, which is documented in particular in Greek and Indo-Aryan.

The semantic content of the new category Middle shows a specialization of the original category “stative”, which was partly integrated into the perfect category and partly entered into the reflexive category through further grammaticalization processes.

6. Conclusions

We think that a functionally wide category ‘middle voice’ was formed in the two diverging groups (i.e., Late PIE and Anatolian) rather independently. From a morphological point of view, the endings set of the Anatolian middle displays the reflexes of the reconstructed endings of the ‘stative’, also attested in the Greco-Aryan perfect conjugation, while the endings of the middle in the other Indo-European languages represent an innovation.

This category included both Geniušienė’s (1987) “objective” and “subjective” reflexives. In other words, it spread from the realm of statives, decausatives and

other events with a non-agentive subject to reflexives and even few simple transitive agent-patient constructions (in this sense we do not disagree, in the general outline, with Luraghi 2012).

Because the verbal morphology had become opaque in delimiting the subdivisions between reflexive (or ‘subjective reflexive’ as per Geniušienė 1987) and middle semantics (or ‘objective reflexive’ in the sense of Geniušienė 1987), additional markers were introduced, such as particles and pronouns, in different ways in the individual historical branches, to better define some of the subfunctions within the wide semantic field that includes both reflexivity and middle semantics.

This explanation is congruent with our considerations regarding the development of the category of the reflexive (Sections 2–4). It is evident that two opposing strategies were working in the IE languages to mark reflexivity. A reflexive marker with active endings [+RM –middle], i.e. a pronominal strategy, against its opposite [–RM +middle], i.e. a verbal strategy. They were sometimes co-existing and overlapping [+RM +middle]. Where a ‘reflexive’ particle, underspecified for person, was operative, we believe we can see a dynamic of distributional rules among a pronominal, a mixed, and verbal strategy. Considering Hittite and Lithuanian, the first showing a 2P particle, the second a suffix that historically is derived by a 2P particle, we have distinguished a possible path of historical development documented in a split fashion. Where the particle and middle voice co-existed, as in Hittite, the former marked the “higher” or more “active” subset of functions typologically recognized in the wide semantic category ‘Middle’; the latter, instead, marked more typically the “lower”, non-agentive, “decausative” and “stative” one. Where the middle voice was lost, the particle could develop into a marker strictly connected with the active verbal endings, thus producing, *de facto*, a new verbal strategy for the Middle, as seen in the Baltic languages.

References

- Archi, Alfonso. 1993. Kamrušepa and the Sheep of the Sun-God. *Orientalia Nova Series* 62.404–409.
- Benedetti, Benedetto. 1981. Sul simbolismo del “legare” e del “tessere”. *Studi e Ricerche* 1.7–18.
- Benedetti, Marina. 2005. Dispersioni formali del medio indoeuropeo. Acquisizione e mutamento di categorie linguistiche. *Atti del Convegno della Società Italiana di Glottologia (Perugia, 23–25 ottobre 2003)* ed. by Lidia Costamagna & Stefania Giannini, 95–119. Roma: Il Calamo.
- Boley, Jacqueline. 1993. *The Hittite Particle -z/-za*. (= Innsbrucker Beiträge zur Sprachwissenschaft, 79.) Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- CHD. Güterbock, et al., eds., (1980–2005).

- Clackson, James. 2007. *Indo-European Linguistics: An Introduction*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511808616
- Cotticelli Kurras, Paola & Alfredo Rizza. 2011. Die hethitische Partikel -z(a) im Licht neuer theoretischer Ansätze. *Indogermanistik und Linguistik in Dialog, XIII: Akten der Fachtagung der Indogermanischen Gesellschaft, vom 21. bis 28. September 2008 in Salzburg* ed. by Thomas Krisch & Thomas Lindner, 120–130. Wiesbaden: Reichert.
- CTH. Laroche (1971).
- Dixon, Robert M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511611896
- Friedrich, Johannes. 1930. Die hethitischen Bruchstücke des Gilgameš-Epos. *Zeitschrift für Assyriologie und verwandte Gebiete N.F.* 39.1–82. DOI: 10.1515/zava.1929.39.1-3.1
- Friedrich, Johannes, Annelies Kammenhuber, Paola Cotticelli-Kurras, & Albertine Hagenbuchner-Dresel, et al. 1975ff. *Hethitisches Wörterbuch: Zweite, völlig neubearbeitete Auflage auf der Grundlage der edierten hethitischen Texte*. Heidelberg: Carl Winter Verlag.
- Geniušienė, Emma. 1987. *The Typology of Reflexives*. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110859119
- Güterbock, Hans G., Harry A. Hoffner & Theo P.J. van den Hout, eds. 1980–2005. *Hittite Dictionary of the Oriental Institute of the University of Chicago*. Chicago: The Oriental Institute.
- Hackstein, Olav. 2003. Reflexivpronomina, Präverbien und Lokalpartikel in indogermanischen Sprachen. *Tocharian and Indo-European Studies* 10.69–95.
- Haspelmath, Martin. 2003. The Geometry of Grammatical Meaning: Semantic Maps and Cross-Linguistic Comparison. *The New Psychology of Language* Vol. 2 ed. by Michael Tomasello, 211–242. Mahwah, NJ: Erlbaum.
- HED. Puhvel (1984ff.).
- Hock, Hans H. 2006. Reflexivization in the Rig-Veda (and beyond). *Themes and Tasks in Old and Middle Indo-Aryan Linguistics, Papers of the 2004 World Sanskrit Conference*, Vol. 5 ed. by Bertil Tikkani & Heinrich Hettrich, 19–44. Delhi: Motilal Banarsidass.
- Hoffner, Harry A. Jr., & Harold C. Melchert. 2008. *A Grammar of the Hittite Language. Part I: Reference Grammar*. Winona Lake: Eisenbrauns.
- HW². Friedrich, et al, eds., (1975ff.)
- Jasanoff, Jay. 2003. *Hittite and the Indo-European Verb*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249053.001.0001
- Josephson, Folke. 2003. The Hittite Reflexive Construction in a Typological Perspective. *Language in Time and Space: A Festschrift for Werner Winter on the Occasion of his 80th Birthday* ed. by Brigitte L.M. Bauer & Georges-Jean Pinault, 211–232. Berlin: de Gruyter.
- Kazenin, Konstantin I. 2001. Verbal Reflexives and the Middle Voice. *Language Typology and Language Universals: An International Handbook* ed. by Martin Haspelmath, 916–927. Berlin: de Gruyter.
- Kemmer, Susanne. 1993. *The Middle Voice*. Amsterdam: John Benjamins. DOI: 10.1075/tsl.23
- König, Eckard & Volker Gast, eds. 2002. Reflexives and Intensifiers: The Use of Self-Forms in English. Special Issue of *Zeitschrift für Anglistik und Amerikanistik* 50:3.
- Laroche, Emmanuel. 1971. *Catalogue des textes hittites*. Paris: Klincksieck.
- Lazzeroni, Romano. 1990. La diatesi come categoria linguistica: studio sul medio indoeuropeo. *Studi e saggi linguistici* 30.1–22.
- Luraghi, Silvia. 2010. Transitivity, Intransitivity and Diathesis in Hittite. *Indo-evropejskoe jazykoznanie i klassičeskaja filologija* 14.133–154.

- Luraghi, Silvia. 2012. Basic Valency Orientation and the Middle Voice in Hittite. *Studies in Language* 36:1.1–32. DOI: 10.1075/sl.36.1.01lur
- Meid, Wolfgang. 1975. Probleme der räumlichen und zeitlichen Gliederung des Indogermanischen. *Flexion und Wortbildung: Akten der V. Fachtagung der Indogermanischen Gesellschaft, Regensburg, 9–14.9.1973* ed. by Helmut Rix, 204–219. Wiesbaden: Reichert.
- Meiser, Gerhard. 2009. Zur Typologie des urindogermanischen Mediums. *Protolanguage and Prehistory: Akten der Fachtagung der Indogermanischen Gesellschaft, vom 11. bis 15. Oktober 2004 in Krakau* ed. by Rosemarie Lühr & Sabine Ziegler, 318–334. Wiesbaden: Reichert.
- Neu, Erich. 1968. *Interpretation der hethitischen mediopassiven Verbalformen.* (= Studien zu den Boğazköy-Texten, 5.) Wiesbaden: Harrassowitz.
- Neu, Erich. 1976. Zur Rekonstruktion des indogermanischen Verbalsystems. *Studies in Greek, Italic, and Indo-European Linguistics Offered to L.R. Palmer* ed. by Anna Morpurgo Davies & Wolfgang Meid, 239–254. Wiesbaden: Harrassowitz.
- Neu, Erich. 1985. Das frühindogermanische Diathesensystem. Funktion und Geschichte. *Grammatische Kategorien: Funktion und Geschichte. Akten der VII. Fachtagung der Indogermanischen Gesellschaft, Berlin, 20–25. Februar 1983* ed. by Bernfried Schlerath & Veronica Rittner, 275–296. Wiesbaden: Reichert.
- Nevis, Joel A. & Brian D. Joseph. 1992. Wackernagel Affixes: Evidence from Balto-Slavic. *Yearbook of Morphology* 1992.93–111.
- Orel, Vladimir. 2000. *A Concise Historical Grammar of the Albanian Language: Reconstruction of Proto-Albanian.* Leiden: Brill.
- Parina, Elena. 2007. Reflexivpronomina im Mittelkymrischen. *Kelten-Einfälle an der Donau: Akten des Vierten Symposiums deutschsprachiger Keltologinnen und Keltologen; philologische, historische, archäologische Evidenzen. Konrad Spindler (1939–2005) zum Gedenken (Linz/Donau, 17–21. Juli 2005)* (= Philosophisch-Historische Klasse, 345) ed. by Helmut Birkhan, 390–396. Wien: Akademie der Wissenschaften in Wien.
- Petit, Daniel. 1999. **su e- en grec ancien: la famille du pronom réfléchi.* Leuven: Peeters.
- Petit, Daniel. 2010. *Untersuchungen zu den baltischen Sprachen.* Leiden: Brill.
DOI: 10.1163/9789047440710
- Puddu, Nicoletta. 2005. *Riflessivi e intensificatori: greco, latino e le altre lingue indoeuropee.* Pisa: ETS.
- Puhvel, Jaan. 1984ff. *Hittite Etymological Dictionary.* Berlin: de Gruyter.
- Safir, Ken. 2004. *The Syntax of Anaphora.* Oxford: Oxford University Press.
DOI: 10.1093/acprof:oso/9780195166132.001.0001
- TDIR = *Typological Database of Intensifiers and Reflexives.* Available at <http://noam.philologie.fu.berlin.de/~gast/tdir>.
- Yakubovich, Ilya. 2010. *Sociolinguistics of the Luvian Language.* Leiden: Brill.

The rise of ‘subordination features’ in the history of Greek and their decline

The ‘Indirect Speech Traits Cycle’*

Ioannis Fykias and Christina Katsikadeli

University of Salzburg

This study is a contribution based on Greek material to a field of inquiry that deals with the diachronic development of formal syntactic devices and their interrelationship with the dichotomy between main and subordinate clauses in Indo-European (Kiparsky 1995, Lühr 2008). First, we focus on some devices signaling indirect speech that emerged in Pre-Classical and Classical Greek, such as the development of a system of complementizers (*hóti* ‘that.COMP’, *hōs* ‘that.COMP’) and some characteristic usages of moods (the optative of indirect speech). In Post-Classical Greek, this system of traits that had been employed to code indirect speech collapsed, as evidenced by the disappearance of *hōs* ‘that.COMP’ and the optative of indirect speech as well as the high frequency of pleonastic *hóti* ‘that.COMP’. Later in the history of Greek a new subordination system arises. We interpret these developments in the light of contemporary syntactic theory (Emonds 2004, 2012), and try to formulate a hypothesis regarding the cycle-like regularities and recurrent patterns that are followed by (clusters of) traits, that is, the “Indirect Speech Traits Cycle”.

* We would like to express our gratitude to Leonid Kulikov and Nikos Lavidas, the organizers of the workshop “Proto-Indo-European Syntax and its Development” (as part of ISTAL 2011) in Thessaloniki, where an earlier version of this paper was presented. We thank all participants for stimulating comments and discussions and especially Joseph Emonds for his encouragement and for his kind offer to send us an earlier draft of his 2012 paper. In the same connection we also want to thank Lobke Aelbrecht, Liliane Haegeman and Rachel Nye, the editors of the volume *Main Clause Phenomena*, for the permission they granted us to cite from Emonds’ contribution for the purposes of this study. Last but not least, we wish to thank Silvia Luraghi and Jóhanna Barðdal, as well as two anonymous reviewers for providing us with detailed comments on the contents of this paper, which have proven to be particularly helpful, as they have contributed to a better understanding and analysis of the issues discussed in the current article.

1. Introduction

In the current study, we attempt an analysis of the emergence and the diachronic development of some subordination phenomena primarily in the domain of finite and infinitival clauses that serve as complements of verbs of speaking and knowing, by following the evolution, progress and decline of a number of formal subordination features in various historical phases all the way from Homeric Greek to the Post-Koine period. We focus on the distribution and development of subordinating conjunctions, person shift, mood shift, and tense shift in the course of the major historical stages of Greek.

In Section 2, it is demonstrated that the aforementioned subordination features belong to a system of traits that serve for coding indirect speech. It is argued that variable sets of traits may be exploited by different languages and even by stages of a single language. Those traits interact with each other and form clusters, and the employment of one trait may have consequences regarding the exploitation or abandonment of another trait. In Section 3, we examine the predictions of contemporary syntactic theory with respect to some exceptional properties of a class of subordinate clauses, and we try to explain the fact that subordinate clauses of Ancient Greek exhibit main clause properties in a way different from their counterparts in Germanic languages. In Sections 4 through 7, the profiles of each historical stage are characterized in terms of traits and trait clusters.

2. The problem

In the context of Indo-European studies, the issue of the diachronic development of word order and other formal syntactic principles as well as their interrelationship with the dichotomy of main and subordinate clauses (MCs and SCs) was first systematically addressed as early as 1894, when Eduard Hermann submitted his doctoral thesis in Jena on the topic “Did Indo-European have subordinate clauses?”. Some years ago, Lühr (2008) claimed in response to Hermann:

The answer is undoubtedly ‘Yes’, but are these clauses embedded or can the reconstructed object clause also support Kiparsky’s (1995:155) opinion that such clauses are only adjoined? Modern generative syntax will help us to find an appropriate answer. (Lühr 2008: 122)

In 1895 Hermann listed 12 formal traits of subordination (Hermann 1895: 481–535). Below, we reproduce the list of traits in their English translation (adopted from Lehmann 1980: 141) along with our abbreviations T_1 – T_{12} (T = trait), which we use in the present article. Of course, it has to be taken into consideration that

these ideas were put forward in a pre-theoretical holistic framework, in which there was no rigid division between levels of linguistic analysis. Therefore, the list includes traits (like T_5 , T_6 , and T_7) that can hardly be a part of a contemporary syntactic analysis, but are rather phonetic or phonological in nature.

- T_1 : Special clause-connecting words (special pronouns, conjunctions)
- T_2 : Shift in person (indirect discourse)
- T_3 : Shift in mood (indirect discourse)
- T_4 : Shift in tense (indirect discourse)
- T_5 : Clause-accent of the clause (SCs enclitic with respect to MCs)
- T_6 : Tempo (SCs spoken faster)
- T_7 : Duration of the pause between the clauses (shorter)
- T_8 : Clause arrangement (narrow connection)
- T_9 : Position of words (especially of the verb: final in SCs, medial in MCs)
- T_{10} : Clause-accent of the verb ("enclitic" in MCs, "orthotonic" in SCs)
- T_{11} : Method of compounding of the verb (tmesis only in MCs)
- T_{12} : Words which are not clause-connectives that occur only in SCs (Greek *spheîs* REFL.NOM.3PL, *sphôn* REFL.GEN.3PL) (cf. also Gippert 2004: 48)

More than 100 years after Hermann's initial discussion, the issue of subordination in Indo-European was taken up within the generative framework. On the basis of Vedic, Greek, and Hittite evidence, Kiparsky (1995) contends that the Indo-European proto-language lacked the category of complementizer and had no syntactically embedded sentences. According to his view, finite SCs, including relative clauses and sentential complements, were syntactically **adjoined** to the MC, exhibiting 'main-clause properties', such as topicalization of constituents to clause-initial position. In most daughter languages, subordinate clauses became syntactically embedded, taking up argument or modifier positions within the main clause; at the same time they lost their main-clause properties and became headed by C^0 . The introduction of complementizers is a consequence of the shift from adjoined to embedded subordination. This restructuring led to several major syntactic characteristics of Germanic, including the rise of V-to- C^0 movement (Kiparsky 1995: 141). Kiparsky's analysis contains two major leading ideas, which play an important role in the current study. First of all, he draws attention to the systematic way in which one formal device may influence another: verb placement and subordinating conjunctions (phenomena corresponding to the traits T_1 and T_9 in Hermann's terminology) are shown to interact with each other. Second, a further hypothesis that is put forth in this article is the idea that a shift from adjoined to embedded subordination can be postulated for the early history of Indo-European languages.

Up to now, the bulk of historical linguistic investigation has focused on T_9 , i.e. the position of the finite verb, final in SCs and medial in MCs. This is not

surprising, since the Germanic languages that have been most systematically studied in this context, cf. e.g. Vikner (1995), are the ideal candidates for the exploration of this phenomenon. This line of research has undoubtedly enriched our understanding of the evolution of clause types in Indo-European.

The emphasis on the Germanic languages, however, has also led to a somewhat one-sided account, in the sense that the rest of the traits have been more or less neglected. In our investigation of the historical stages of Greek, we utilize (some of) Hermann's traits, as heuristics, in a more balanced fashion by taking into account the distribution of more than one trait in a single historical stage, and by analyzing the interaction within a T-cluster as well as between T-clusters.

In what follows, we briefly outline the characteristic features and aims of our approach that differentiate it from other accounts. While our analysis is more restricted as far as width of coverage is concerned in the sense that we explore — in some detail — only one Indo-European language, namely Greek, it still is more extensive, since different historical stages of Greek are systematically investigated.

The fact that we pursue an analysis of subordination traits as a whole set, instead of studying individual traits in isolation, forces us to look for what is shared by all Ts. As is demonstrated below, the common denominator boils down to the distinction direct vs. indirect discourse. It is noteworthy, that in Greek we do not observe the phenomena that are characteristic for Germanic languages and are due to the interplay of the traits T_9 (verb placement) and T_1 (subordinating conjunctions). Nevertheless, there are clear effects of the dichotomy direct vs. indirect discourse (henceforth DD vs. ID) in other domains as well, crucially involving interaction of Ts and formation of T-clusters.

We focus on the distribution and development of subordinating conjunctions (T_1), person shift (T_2), mood shift (T_3) and tense shift (T_4) in the course of the major historical stages of Greek. The stages we postulate (somewhat arbitrarily) for the purposes of our analysis are as follows:

- STAGE A: Pre-Classical Greek (the period before the 5th century BCE)
- STAGE B: Classical Greek (primarily the Attic dialect of the 5th and 4th centuries BCE)
- STAGE C: Non Literary Koine (3rd century BCE to 1st/2nd century AD)
- STAGE D: Post-Koine Greek.

These stages will be discussed in the Sections 4 through 7 below.

What is characteristic for our exploration of the interplay of Ts in a given period of time is the attempt to demonstrate that changes affecting one trait almost automatically affect other Ts that are part of the cluster as well. In this way, we make evident that the traits interact in a fashion reminiscent of the way in which parameters are conceived of and interact in the Principles and Parameters model.

For instance, it has been shown, in connection with certain parameters, which are associated with invariant principles of UG and specify a range of options, that if variant properties cluster, then it may be that one aspect of a grammar is acquired 'for free' once another is acquired, e.g. AuxV order as a consequence of VO order, cf. Roberts (2007: 21, 24). This means that the emergence, or alternatively the decline, of one T has consequences regarding the presence or absence of other Ts. In this article, we will confine our study to establishing this fact at the level of descriptive generalizations.

In agreement with Lühr (2008), and in the spirit of investigating competing structures, we also examine the place that infinitival complements occupy in each historical stage of Greek, although, of course, infinitival clauses are arguably not Proto-Indo-European, which explains why they do not form part of Hermann's system. But for the purposes of our analysis, which is based predominantly on Greek material, we add infinitival complement clauses to the list of traits under the label T_0 . Due to space limitations, participial complements are not examined in this study.

One of the major aims of our account is to highlight the fact that the diachronic development of those traits, either taken individually or as clusters, may constitute evidence as to several kinds of cycles which are at play. A cycle is conceived of as a time span during which certain events take place and come to a conclusion. Towards the end of the cycle, similar events start again, but they are at least slightly different and happen at a different pace. The changes are therefore unidirectional (cf. van Gelderen 2011: 1). One kind of cycle can be identified by following the emergence, by means of reinterpretation and/or grammaticalization of an item, development and decline of an individual trait. A more complex type of cycle involves a cluster of several traits, e.g. a combination of subordinating conjunctions, mood shift, and person shift, that characterizes a specific historical stage of a linguistic system, but is no longer available for a succeeding stage and, in the ideal case, it recurs at an even later stage, cf. also Heine et al. (1991: 243ff) on linguistic cycles.

As already mentioned above, a fundamental component of our working hypothesis is the consideration of the distinction DD vs. ID which, as a consequence, creates an immediate link to formal theoretical approaches, in the context of contemporary syntax, that appeal to similar notions and categories. What we have specifically in mind is the seminal work by Emonds (1970, 1976, 2004, 2012), who has explored a number of syntactic phenomena that highlight the relevance of the distinction DD vs. ID for formal syntactic principles and patterns. He has investigated asymmetries between root and embedded clauses and his central thesis can be summed up as follows:

The crux of this issue is how to specify the distribution of indirect discourse, a puzzle for which an elegant formal solution has eluded grammarians for centuries. Resolution of this problem will perhaps be found in terms of formal semantic representations ... In any case, one thing is certain: these RIDE [= root-like indirect discourse embedding] projections are not simply a subcase of selected clausal complements. (Emonds 2012: 29)

We start with some significant insights of Emonds' theory, which are important, because they enable us to plausibly hypothesize what kind of principles might have been at work as part of the leading force behind some significant developments in the diachrony of Greek, in other words, which principles or tendencies might have contributed to the development of individual Ts and T-clusters.

3. The distinction root vs. non-root clauses

In his earlier work, Emonds (1970, 1976) introduced the distinction between 'structure-preserving' and 'root' transformational operations. The former are 'structure preserving' in the sense that a constituent of a category A is moved to a position hosting constituents of the same category A. Root transformations, on the other hand, create new positions at the root of the sentence. Operations of the latter type are also known as Main Clause Phenomena, because they are usually restricted to non-embedded domains. Hooper & Thompson (1973), however, showed that a number of root transformations are permitted in a set of embedded clauses. This point was acknowledged by Emonds, who, in subsequent work (Emonds 2004, 2012) attempted to offer an explanation for those phenomena by drawing attention to the fact that dependent clause contexts like *warn someone that* ____ and *make a promise that* ____ imitate the freedom of root structures in what is traditionally called **indirect discourse**. Such root-like indirect discourse embedding ("RIDE"), however, is incompatible with most dependent clause positions. In English and German, RIDEs are always (a) finite, (b) complements rather than adjuncts, and (c) tend to be governed by V or A rather than by lexical N or P. Moreover, (d) some argument of these governing Vs and As must be animate. When embedded clauses do not fulfill all four of these conditions on RIDE positions, root transformations produce clear unacceptability (Emonds 2004: 76). A subset of the crucial contrasting data in Emonds (2004: 77–78) that support his position are presented below as (1–10).

Finite IP:

(1) Topicalization:

Bill warned us that [_{RIDE} *flights to Chicago* we should try to avoid].

- (2) Negative preposing:

I made a promise that [_{RIDE} *only until five* would we work].

Non-finite IP:

- (3) *Bill warned us [flights to Chicago to try to avoid].

- (4) *We will propose [only until five working] to the management.

Adjunct IP:

- (5) *Mary used another company since [flights to Chicago they could avoid].

- (6) *I ignored the boss who was so angry that [only until five did we work].

Complements to N:

- (7) *A warning that [flights to Chicago they should avoid] will soon be posted.

- (8) *Their promise that [only until five will they work] will be hard to keep.

No animate argument of governing V:

- (9) *The weather meant that [flights to Chicago travellers carefully avoided].

- (10) *No experiment showed that [such material this metal would react with].

To account for this asymmetry, Emonds proposes a few general principles to specify the structural “tops” of clausal projections. These include (i) the Tensed S Constraint of Chomsky (1973), and (ii) a simplified Structure Preserving Constraint (“SPC”, Emonds 1970, 1976). These principles can account for the above and many other paradigms in unembedded (“root”) clauses in, e.g., English and German. Perhaps the most important proposal here is that a construct of unlabeled Discourse Shells is the best way to analyze the “left periphery” of root clauses. Certain unselected finite clause IPs called Discourse Projections may be immediately dominated by a series of categorically unspecified XPs. ‘Unselected’ means a clause that is not an argument or adjunct of an underlying lexical item in X^0 . The category-less XPs that can dominate unselected Discourse Projections are called “Discourse Shells”. Only SPECs of Discourse Shells serve as categorically unrestricted landing sites for movements, the property that characterizes ‘root’ phenomena. (Emonds 2012: 25). Discourse Shells are a-categorial counterparts to Rizzi’s (1997) FocP and TopP.

According to this framework, unembedded, unselected finite clause IPs are Discourse Projections in all languages. As far as the rest of the clauses are concerned, there are language specific systems that interact with Discourse Projections in special ways. In accordance with the Discourse Projection Parameter, particular languages may specify larger classes of finite clauses as Discourse Projections.

Many languages, including English, allow Discourse Projections in the positions of certain types of embedded clauses. The main-embedded contrast is stronger in Dutch, where few if any dependent clauses qualify as root-like. In Icelandic and Yiddish, a wider range of clauses permit V-second than in English or German (Vikner 1995: 80–83). Some analyses have claimed that these languages use SPEC (IP) as a landing site for topicalization, rather than SPEC (CP) as in German and Dutch. Emonds' model suggests that Icelandic and Yiddish have larger sets of embedded RIDEs than does English. If so, embedded clauses, beyond those that serve to convey indirect discourse, may additionally qualify as Discourse Projections in Icelandic and Yiddish. As a result, a finite V in these languages moves to second position in more types of clauses (Emonds 2004: 88).

In the context of the current study, one of the most challenging questions concerning historical Greek syntax is how the phenomena of different stages may fit into this variation spectrum allowed within this theoretical framework. From the point of view of Emonds' system, the most remarkable feature of STAGE A and STAGE B syntax is the existence of infinitival clauses that serve for conveying ID and exhibit root-like properties. This state of affairs is unexpected, since it stands in apparent contradiction to the requirement for Discourse Projections to conform to the Tensed S Constraint. In other words, we have clauses which at the same time display root properties and are, at least *prima facie*, not finite. In Section 5.2, we provide an explanation for resolving this apparent contradiction.

An equally remarkable fact in connection with historical Greek syntax that also calls for an explanation is that infinitival clauses disappeared altogether from later Greek. The total absence of the infinitive, and the use of finite clauses instead, has become one of the most salient typological characteristics of Post-Koine Greek (STAGE D). In Emonds' framework this can be interpreted as preference for a more unmarked option of the Discourse Projection Parameter, after the collapse and the subsequent reorganization of the system of ID subordination traits.

In the next sections, we investigate how individual Ts and T-clusters emerge, change, disappear, and in some cases reappear isolated or as a block, in each of the stages of Greek. At the same time we pay close attention to the development of several types of MCs and SCs in association with the capacity to exhibit root-like properties.

4. Pre-Classical Greek

4.1 STAGE A and subordinating conjunctions (T_1): Genesis of complementizers

In diachronic studies on Germanic (Kiparsky 1995, Roberts & Roussou 2003), it has been hypothesized that the innovation of the category 'complementizer' led to the introduction of the CP-structure, a development which in turn triggered the emergence of V-to-C movement in independent clauses. In Greek, there are no spectacular word order effects such as verb second or verb final structures that depend on the presence or absence of a complementizer, respectively. Nevertheless, we can trace the emergence of categories like complementizers as a product of reanalysis of pronouns and adverbial particles, as far as individual words are concerned.

Before we tackle the situation of STAGE A both with respect to the issue of the reanalysis of categories like pronouns as complementizers, and with respect to other traits that are specific to subordinate structures, some important clarifications are in order, since they may help us envisage what the stage preceding STAGE A might have looked like. As Roberts (2007: 174) argues, it cannot be seriously maintained that earlier stages of certain languages may have lacked subordination altogether, since this would be tantamount to rejecting the idea that all languages at all times reflect the same basic UG. In languages like Classical Latin and Turkish, however, finite complementation plays a rather marginal role. So it can be plausibly assumed that finite complementation is just one parametric option. Related to the traditional parataxis-to-hypotaxis hypothesis is the assumption that adjunct clauses of various kinds may be reanalyzed as complement clauses.

We use Homeric data as evidence for STAGE A, with the objective of substantiating a viable hypothesis as to what might have happened in the course of the transition from Indo-European to early Greek, and led to the development of the spectrum of finite and nonfinite clauses that serve as complements of certain classes of verbs.

Let us sketch a plausible scenario of how the grammaticalization of *hóti* might have progressed. The first transition phase in the history of Greek, as far as T_1 is concerned, can be documented by means of the Homeric evidence. According to Monro (1891: 245), the accusative neuter of the relative pronoun in Homer, when used adverbially, yields the three 'conjunctions' *hó*, *hóti*, *hó te* which mean literally 'in respect that', hence usually (a) 'because', or (b) 'that' (after a verb of speaking, knowing, etc). He contends that the transition to the meaning 'that' may be seen in examples like *Iliad* 2.255.

- (Homer, *Iliad* 2.255)

as *Iliad* 8.175, glossed and translated in (12) below, and *Iliad* 1.536 show.

- victory and great glory.' (Homer, *Iliad* 8.175)

with when we study Classical Greek texts.

4.2 STAGE A: Tense shift (T_4) instead of mood shift (T_3)

stance of *hóti* with the optative and just one instance of *hōs* with the optative (Od.

24.237) in Homer. The present tense of DD usually becomes the imperfect of ID, the perfect becomes the pluperfect.¹

- (13) ou gár hoí tis etétumos ángelos
 NEG PTC her.F.DAT.SG INDEF.M.NOM.SG true.M.NOM.SG messenger.M.NOM.SG
 elthòn éngeil' hótí rhá
 come.PTCP.AOR.ACT.M.NOM.SG announce.IND.AOR.ACT.3SG that then.PTC
 hoi pósis éktothi mímne puláōn
 her.F.DAT.SG husband.M.NOM.SG outside stay.IND.IMP.ACT.3SG gate.F.GEN.PL
 'for no sure messenger had come to her and told her how her husband had
 held his ground there outside the gates' (Homer, *Iliad* 22.439)²

Thus, in terms of traits, there is evidence for the use of T_4 (tense shift) rather than T_3 (mood shift) in this phase of Greek. As far as MC vs. SC characteristics are concerned, we can conclude for this stage that there is no significant differentiation between root and non-root clauses in terms of word order, like the one that can be observed in languages like modern English or German. In other words, most STAGE A finite SCs exhibit root-like properties in Emonds' sense.

5. Classical Greek: STAGE B and the system of subordination traits

At this stage we have evidence for the highest degree of development and elaboration as far as individual Ts and T-clusters are concerned.

5.1 STAGE B: Subordinating conjunctions (T_1) pertaining to the complementizer *hóti*

By this time (5th/4th century BCE), the embedded *hóti* pattern had been established in the grammar of the language. We can clearly identify root-like properties in finite *hóti*-complement clauses selected by verbs of speaking. In (16–17), we present representative illustrations of two of the types of constructions that have been adduced by Emonds (2004:79) as supporting evidence for the availability of

1. As Monro (1891:245) remarks: "Every assertion is made from the speaker's own point of view: ... what was present to the person quoted must be treated as now past." It has to be pointed out that we do find some instances of the optative in indirect questions already in Homer, cf. Od. 17.368.

2. We are aware that the etymological analysis and interrelation of the Ancient Greek conjunctions *hótti* and *hóti* are not completely clear, but in this example we simply focus on the presence of tense shift and the absence of mood shift.

“root” operations in embedded clauses, e.g. constructions involving negative constituent fronting (cf. the corresponding English example in 14), and preposing of AP over ‘be’ (cf. the corresponding English example in 15).

- (14) She said that never in her life would she work this hard again.
- (15) Carol said that most embarrassing of all was falling off the stage.
- (16) Xenophōn d’ apekrínato hótī oudèn
 Xenophon.M.NOM.SG but reply.IND.AOR.MID.3SG that nothing.N.ACC.SG
 àn toutōn eípoi eis tèn stratián
 PTC these.GEN.PL say.OPT.AOR.ACT.3SG to ART.F.ACC.SG army.F.ACC.SG
 ‘Xenophon replied that he would say nothing of this to the army’
 (Xenophon, *Anabasis* 5.6.37)
- (17) égnōsan hoi stratiōtai hótī
 realize.IND.AOR.ACT.3PL ART.M.NOM.PL soldiers.ART.M.NOM.PL that
kenòs ho phóbos eíē
 empty.M.NOM.SG ART.M.NOM.SG fear.M.NOM.SG be.PRS.OPT.3SG
 ‘the soldiers recognized that their fear was groundless’ (Xenophon, *Anabasis* 2.2.21)

5.2 STAGE B: Infinitival complement clauses (T_0)

What follows concerns competing structures serving as complements of verbs of speaking, hearing, knowing, etc. It can be argued that at STAGE B, infinitival complements and *hótī* clausal complements can be regarded as equivalent, isomorphic complements of verbs of knowing like *oída* ‘know’.

Both infinitival and finite complement clauses can be shown to be structurally equivalent in this phase of Greek. Both STAGE A and STAGE B infinitives are exceptional in the sense that they can be shown to be tensed, whereas infinitives in familiar modern languages are not. In what follows, we present three arguments in favor of the isomorphism thesis for Greek infinitives at this historical stage. Isomorphism in this connection means that, apart from person and number information and the corresponding morphemes, infinitival constructions employ almost identical morphological means as their finite counterparts to implement the same grammatical functions.

Sevdali (2007: 134) contrasts the possibility of coordinating a finite clause with an infinitival one in Ancient Greek (18) and the ungrammaticality of the corresponding structure in English (19).

- (18) humās mèn pántas eidénai hēgoūmai
 2PL.ACC on.the.one.hand all.M.ACC.PL know.INF.PRES think.IND.PRES.1SG

[hóti egō mèn orthōs légō,
 that 1SG.NOM on.the.one.hand correctly speak.IND.PRS.1SG
 [toūton dè hoútō skaiòn eīnai.]
 him.ACC.SG on.the.other.hand so stupid.ACC.SG be.INF.PRS
 'I think that you have all perceived that I speak correctly, whereas he is
 stupid.' (Lysias, *Against Theomnestus* 1: 15,3)

- (19) *I think that you have all perceived that I, for my part, speak correctly, but
 him to be stupid.

It is noteworthy that in (18) "the two bracketed clauses are linked with the *mèn ... dè* 'on the one ... on the other hand' type of co-ordination" (Sevdali 2007: 134). As Sevdali (2007: 134) concludes, "The fact that AG infinitives are found co-ordinated with finite clauses is suggestive of their status as CPs". Further, Jannaris (1897: 570) supplies a number of examples of declarative clauses, which sometimes begin analytically (with *hóti*, rarely *hōs*) but close with an infinitive, as in (20).

- (20) akoúō hóti kai sunthēreutás tinas
 hear.IND.PRS.ACT.1SG that and cohunters.M.ACC.PL some.M.ACC.PL
 tōn paidōn soi genésthai autoū
 ART.GEN.PL sons.GEN.PL 2SG.DAT become.INF.AOR.MID 3SG.GEN
 'I understand that some of his sons were among your companions in the
 chase' (Xenophon, *Cyropaedia* 2.4.15)

A third argument in favor of the special status of Pre-Classical and Classical Greek infinitives (Miller 2002: 34–35, Sevdali 2007: 44, Markopoulos 2009: 30ff) pertains to the morphological nature of Greek infinitives: (a) They have non-prototypical infinitival-like morphology that includes voice and tense/aspect markers, and (b) in the future infinitive, this morphology is (exceptionally) temporal, as shown in (21):

- (21) eí tiná pou Danaōn éti élpetai
 if INDE.M.ACC.SG by chance Dan.M.GEN.PL yet hope.PRS.IND.3SG
 exapatésein
 deceive.INF.FUT.ACT
 'if by chance he hopes to utterly deceive yet another of the Danaans' (Homer, *Iliad* 9.371)

Now, as far as the future infinitive is concerned, it has been frequently taken for granted that what is at issue is aspect rather than tense. But, as argued by Miller (2002: 34–35), future infinitives (cf. example 21) definitely coded tense contrast, if by tense contrast we mean the ability of a complement structure to realize a range of possible tenses independent of that of the matrix verb. Additionally, as

Markopoulos (2009: 30) pointed out, some approaches did not consider the fact that the Ancient Greek future in general could convey both perfective and imperfective readings, being aspectually neutral. As Miller (2002: 34) succinctly summarizes, “Infinitives can have tense morphology. When they do they can be used exactly as fully tensed complement clauses”.

As mentioned above (see Section 3), the existence of infinitival clauses that serve to convey ID and exhibit root-like properties is unexpected from a theoretical point of view, since this state of affairs is in apparent violation of the Tensed S Constraint. The solution supported by the three arguments adduced in this section is to assume that at STAGE A and at STAGE B infinitival complements were tensed, thus conforming to the Tensed S Condition and qualifying as parts of Discourse Shells that allow for root-like operations.

There are also good reasons for assuming that this non-canonical state of affairs which represents a rather marked option of the Discourse Projection Parameter was in force only for a relatively limited time span in the history of Greek. There is strong evidence that infinitives began to lose this exceptional status in later stages of Greek. A reliable diagnostic criterion seems to be the behavior and the distribution of the future infinitive. As soon as Greek infinitival constructions are no longer capable of coding tense contrast or, alternatively formulated, when they lose their tense dimension, the future infinitive is expected to decline. An indication of the decline of the future infinitive is the fact that the form of the future infinitive is no longer used regularly and productively from the 4th century BCE onwards. The distribution of the future infinitive in the formation of the periphrastic future by means of *méllo* suggests that the future infinitive begins to decline as a category (cf. Markopoulos 2009: 28, Joseph 1983: 6–36 on finiteness).

5.3 STAGE B: Mood shift (T_3)

In this period, the optative of indirect discourse (henceforth OID) was generalized in its use, as example (22) illustrates.

- (22) élexan hótí pémpseie sphās
 say.IND.AOR.ACT.3PL that send.OPT.AOR.ACT.3SG REFL.M.ACC.PL
 ho Indōn basileús
 ART.M.NOM.SG Indian.M.GEN.PL king.M.NOM.SG
 ‘they said that the king of the Indians had sent them’ (Xenophon, *Cyropaedia* 2.4.7)

It has to be pointed out that the change to the optative represented a choice in expression and was not syntactically obligatory. By retaining the subjunctive or indicative of DD, the speaker was able to present the indirect clause as direct and as

being produced at the moment of speaking. The use of OID presented the indirect clause with less immediacy, as if it were something for which the speaker was unwilling to bear absolute responsibility (cf. also Philippaki-Warbuton 2007: 594).

5.4 STAGE B and clause connecting words (T_1) again: Additional specialized properties of complementizers

The principles governing the distribution of *hóti* or *hōs* after verbs of speaking can be outlined as follows: Some verbs take either a *hóti*- or *hōs*-clause or an infinitival clause. Affirmative clauses are usually followed by the infinitive or *hóti*; on the other hand, *hōs* is preferred to *hóti* if a writer wishes to mark a statement as an opinion, a pretext, as untrue, when the MC is negative, or when the SC is negative (cf. Smyth 1918: 582).³ The same holds for verbs of thinking, cf. Thucydides, 3.88. Example (23) illustrates this point.

- (23) légei d' hōs hubristḗs eimi
 say.IND.PRES.ACT.3SG PTC that.COMP insolent.M.NOM.SG be.IND.PRS.1SG
 'he says that I am an insolent person' (Lysias, *On the Refusal of a Pension*
 I,15)

5.5 Classical Greek: Some interpretations and conclusions

Our investigation has shown that, as far as the domain of subordinating conjunctions (T_1) is concerned, the usage of *hóti* 'that' gets established even in contexts in which at STAGE A the matrix verb used to require exclusive use of infinitival complements. At the same time, its distribution is redefined, since at this stage we witness the establishment of the complementizer *hōs* in a specialized context. There is a systematic meaning difference between the two complementizers, which persists throughout the whole Classical period.

With regard to root vs. non-root characteristics, there is no essential asymmetry between MCs and finite SCs, since both types of clauses behave in an identical manner as far as both word order and transformations like AP preposing over 'be' and negative constituent fronting are concerned. We also claim that, for this period of Greek, an exceptional status must be allotted to infinitival structures, since they display maximal isomorphism as compared to finite complement structures, as examples (18) and (20) convincingly suggest.

The distribution of competing constructions that serve as complements of the same classes of verbs can be summed up as follows: In Classical Greek, the

3. Cf., e.g., Xenophon, *Anabasis* 1.3.5.

complement of verbs of perceiving, believing, showing, and speaking is primarily rendered by an infinitival clause, the participle being an alternative construction; in addition to these constructions, the complement of verbs of perceiving, showing, speaking (not of verbs of believing) is often formed by means of an indirect question, and the development of this use is the construction with *hóti* (strictly *hó,ti* an indirect interrogative particle), which is allowable with these same verbs. Lastly, *hōs* with a finite verb is also in use with verbs of speaking, hearing etc. (cf. Blass & Debrunner 1961[1905]:203).

6. STAGE C: Non-literary *koine* texts

According to Horrocks (2007:623–625), the whole system of formal devices that serve for coding indirect speech underwent a series of serious changes in the Hellenistic period. The use of infinitival and participial constructions was drastically reduced, and the optative of indirect discourse in past-time contexts was abandoned. The accusative + infinitive complement clauses after verbs of thought and belief were widely replaced by *hóti* finite clauses with an indicative verb, whereas accusative + infinitive complement constructions after impersonal modals and control verbs of ‘wanting’ and expecting were replaced by finite clauses containing a subjunctive verb introduced by *hína* or *hópōs*.

6.1 STAGE C and mood shift (T₃)

The STAGE B rule respecting mood shift, i.e. the use of OID after past tenses, lost its force during STAGE C, and no longer holds for NT Greek with the exception of Luke, who occasionally uses the OID, as in Luke 8:9 after past tenses in the matrix clause, but this applies only to certain indirect questions. In all other cases of dependence, the OID has retreated before either the subjunctive or the indicative (cf. Jannaris 1897:452).

- (24) *élegon* *gàr hóti exéstē*
 say.IND.IMPF.ACT.3PL for that lose.consciousness.IND.AOR.MID.3SG
 ‘because they kept saying, that he’s out of his mind’ (Mark 3:21)

It must be pointed out that OID was no longer available as a device for marking ID at STAGE C. It is also noteworthy that the reason why optative fell out of use cannot be morphological, since the primary optative in MCs was still a living category of NT Greek.

6.2 STAGE C: Subordinating conjunctions (T_1), person shift (T_2) and the T_1 - T_2 Cluster

One of the typical alleged DD properties of a great number of *hóti*-complement constructions after verbs of **speaking** in NT Greek is the high frequency with which the pleonastic or quotative *hóti* occurs. Example (25) illustrates this use.

- (25) *kai tóte homologéō autoīs hóti oudépote*
 and then confess.IND.FUT.ACT.1SG them.DAT.3PL that never
égnōn humās
 know.AOR.IND.ACT.1SG you.ACC.PL
 'Then I will tell them plainly, "I never knew you!"' (Mat. 7: 23)

As correctly observed by Jannaris (1897:472), in this popular mode of DD it is very common to indicate the dependence of the verbatim clause by placing before it the conjunction *hóti*, which then seems redundant (in New Testament about 120 times) and corresponds to our quotation marks.

Within the framework adopted in the current study, these frequent instances of NT use can be best characterized in terms of interaction between T s or T -clusters. The unmarked option of the combination T_1 (subordinating conjunctions) and T_2 (person shift) that was characteristic for STAGE B can be captured as follows: Whenever T_1 (subordinating conjunctions) occurs, an automatic activation of T_2 (person shift) is expected. But in NT Greek (STAGE C) this maxim is not necessarily obeyed, or in other words the traits T_1 and T_2 can be dissociated from each other, although the default instantiation of this cluster remains the STAGE B value. This dissociation creates a parataxis effect, as the output has been mostly interpreted by traditional Koine-Greek accounts.

6.3 STAGE C: T_1 in *hōs*-complement clauses?

According to Jannaris (1897:412), at STAGE B only *hóti* and the less assertive *hōs* were in general usage. At the end of STAGE B, *hōs* retreated from the language and the T_1 domain had to be defined anew. A fine distinction concerning ID features cannot be maintained any longer in non-literary Post-Classical Greek.

7. Post-Koine Greek (STAGE D)

In this section, we do not present a detailed analysis of the traits of STAGE D. We confine ourselves to briefly commenting on the development of three characteristic phenomena from STAGE C in later Greek. In Post-Koine Greek, we observe

a drastic reduction in the frequency of pleonastic *hóti* in general, as well as the avoidance of T_1 – T_2 clustering violations (where T_1 = subordinating conjunctions, and T_2 = person shift) in complement clauses (see Section 6.2). The use of tense shift (T_4) instead of mood shift (T_3) in the context of reported speech after past tenses in the matrix clause becomes the absolute rule in Post-Koine Greek. The translation into Modern Greek of the example (17), which displays mood shift, is given here as (26), without mood shift but with tense shift, and illustrates this development.

- (26) i stratiótes dhiepístosan óti o
 ART.M.NOM.PL soldiers.M.NOM.PL realize.IND.AOR.ACT.3PL that ART.M.NOM.SG
 fóvos tus ítan avásimos
 fear.M.NOM.SG POSS.3PL be.IND.IMPF.3SG groundless.M.NOM.SG
 ‘the soldiers recognized that their fear was groundless’

Reflecting on a more complex type of cycle of linguistic change, we cannot fail to note that, on the surface, STAGE D resembles the hypothetical precursor of STAGE A as far as the following points are concerned: absence of verbal infinitive (T_0), no mood shift (T_3) but rather tense shift (T_4) as the typical means to signal ID in *that*-complements.

8. Summary and conclusions

In this study, we have shown that Hermann’s traits can be optimally exploited for discovering diachronic developments and tendencies in the history of Greek, and for describing them, provided that they are conceived of as a system, whose members enter a number of well defined relations to each other. We have extended the set of traits by adding T_0 in order to attain a more adequate account of the Greek data. We have shown that the interaction of Ts that we explored resembles the pattern displayed by the T_1 – T_9 interaction, (subordinating conjunction and verb placement) that has been extensively studied in the historical syntax of Germanic. At the same time we have paid close attention to the development of several types of MCs and SCs in association with the capacity to exhibit root-like properties.

One of the major aims of our study has been to demonstrate that the diachronic development of Ts (and T-clusters) can help us locate several kinds of cycles of linguistic change. A simple kind of cycle was documented in the cases of the emergence (by means of reinterpretation and/or grammaticalization) of the complementizers *hóti* and *hōs* (relative pronoun introducing adjoined relative clauses → complementizer introducing embedded complement clauses → quotative particle

introducing main clauses). We have also been able to follow the development and decline of the OID as an individual trait.

Apart from that, we have been in a position to trace the development of more complex kinds of cycle that involve a T-cluster (e.g., a pattern involving a subordinating conjunction, mood shift, person shift). More specifically and on the basis of the foregoing discussion, we can draw a number of generalizations pertaining to the use of individual traits as well as the relationship among Ts, especially in connection with cluster formation.

- Throughout all four stages of the history of Greek, in contrast to the Germanic languages with the exception of English, the traits T_1 (subordinating conjunctions) and T_9 (verb placement) do not constitute a T-Cluster.
- The traits T_1 (subordinating conjunctions) and T_3 (mood shift) may constitute a cluster, if T_1 is realized as *hóti* or *hōs*, only during STAGE B, cf. Section 5.3 example (22).
- The trait T_3 (mood shift) is used at STAGE A and STAGE B and, in some isolated cases, at STAGE C in the domain of indirect questions (a construction in which the complementizer position is not occupied by *hóti* or *hōs*); cf. Section 4.2 fn. 2 and Section 6.1 Luke 8:9.
- At all other stages of the history of Greek (STAGE A, C and D), the cluster T_1 (subordinating conjunctions) – T_4 (tense shift) regularly occurs; cf. Section 4.2 example (13), Section 6.1 example (24), Section 7 example (26).
- One might assume that the exploitation of T_3 (mood shift) in a subordinate clause that is introduced by a complementizer blocks the utilization of T_4 (tense shift) and vice versa.
- The cluster T_1 (subordinating conjunctions) – T_2 (person shift) applies almost obligatorily at STAGE B or, in other words, it attains the status of a quasi-grammatical principle.
- At STAGE C, the cluster T_1 (subordinating conjunctions) – T_2 (person shift) allows numerous exceptions, cf. Section 6.2 example (25).
- At STAGE D, the cluster T_1 (subordinating conjunctions) – T_2 (person shift) re-attains the status that it used to have at STAGE B.

By considering more cross-linguistic data from other old Indo-European languages like Latin, Indo-Iranian, Hittite and Armenian we can refine our analysis and ultimately contribute to a better understanding of the mechanisms that were at work in PIE syntax and their development. In addition, there are indications of a correlation between the stages of dramatic change in the implementation of traits of indirect discourse and the periods of time when we have evidence that intensive language contact was at play, e.g. both at STAGE C (*Koine*-stage) and later at the period of the genesis of various “Balkan Sprachbund” phenomena, cf. also

Hock (1982) on similar phenomena concerning the Sanskrit quotative and Joseph (1983: 179–241). In future work, we plan to explore the role that language contact and convergence area considerations have played in inducing changes pertaining to the ID signaling devices of the languages involved.

References

- Blass, Friedrich & Albert Debrunner. 1961. *A Greek Grammar of the New Testament and Other Early Christian Literature: A Translation and Revision of the Ninth–Tenth German Edition Incorporating Supplementary Notes of A. Debrunner* (translated by Robert W. Funk). Cambridge: Cambridge University Press.
- Chomsky, Noam. 1973. Conditions on Transformations. *A Festschrift for Morris Halle* ed. by Stephen R. Anderson & Paul Kiparsky, 232–286. New York: Holt, Rinehart and Winston.
- Christidis, A.F., ed. 2007. *A History of Ancient Greek: From the Beginnings to Late Antiquity*. Cambridge: Cambridge University Press.
- Emonds, Joseph. 1970. *Root and Structure-Preserving Transformations*. Cambridge, MA: MIT PhD dissertation.
- Emonds, Joseph. 1976. *A Transformational Approach to English Syntax*. New York: Academic Press.
- Emonds, Joseph. 2004. Unspecified Categories as the Key to Root Constructions. *Peripheries: Syntactic Edges and their Effects* ed. by David Adger, Cécile de Cat & George Tsoulas, 75–120. Dordrecht: Kluwer.
- Emonds, Joseph. 2012. Augmented Structure Preservation and the Tensed S Constraint. *Main Clause Phenomena: New Horizons* ed. by Loebke Aelbrecht, Liliane Haegeman & Rachel Nye, 23–46. Amsterdam: John Benjamins. DOI: 10.1075/la.190
- van Gelderen, Elly. 2011. Language Change as Cyclical: A Window on the Language Faculty. *Studies in Modern English* 27.1–23.
- Gipfert, Jost. 2004. Indo-European Word Order in Main and Subordinate Clauses in a Diachronic Perspective. *Analecta homini universali dedicata: Arbeiten zur Indogermanistik, Linguistik, Philologie, Politik, Musik und Dichtung. Festschrift für Oswald Panagl zum 65. Geburtstag*, Vol. I ed. by Thomas Krisch, Thomas Lindner & Ulrich Müller, 48–68. Stuttgart: Heinz.
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer. 1991. *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press.
- Hermann, Eduard. 1894. *Gab es im Indogermanischen Nebensätze?* Jena: University of Jena PhD dissertation.
- Hermann, Eduard. 1895. *Gab es im Indogermanischen Nebensätze?* *Zeitschrift für Vergleichende Sprachforschung* 33.481–535.
- Hock, Hans Henrich. 1982. The Sanskrit Quotative: A Historical and Comparative Study. *Studies in the Linguistic Sciences* 12.39–85.
- Hooper, Joan & Sandra Thompson. 1973. On the Applicability of Root Transformations. *Linguistic Inquiry* 4.465–491.
- Horrocks, Geoffrey. 2007. Syntax: From Classical Greek to the Koine. In A.F. Christidis, ed., 618–631.

- Jannaris, Anthony N. 1897. *An Historical Greek Grammar Chiefly of the Attic Dialect*. Hildesheim: Georg Olms. (Reprinted 1968.)
- Joseph, Brian. 1983. *The Synchrony and Diachrony of the Balkan Infinitive: A Study in Areal, General and Historical Linguistics*. Cambridge: Cambridge University Press.
- Kiparsky, Paul. 1995. The Indo-European Origins of Germanic Syntax. *Clause Structure and Language Change* ed. by Adrian Battye & Ian Roberts, 140–170. Oxford: Oxford University Press.
- Lehmann, Winfred P. 1980. The Reconstruction of Non-Simple Sentences in Proto-Indo-European. *Linguistic Reconstruction and Indo-European Syntax: Proceedings of the Colloquium of the 'Indogermanische Gesellschaft', University of Pavia, 6–7 September 1979* ed. by Paolo Ramat, 113–144. Amsterdam: John Benjamins. DOI: 10.1075/cilt.19.10leh
- Lühr, Rosemarie. 2008. Competitive Indo-European Syntax. *Principles of Syntactic Reconstruction* ed. by Gisella Ferraresi & Maria Goldbach, 121–159. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302.07luh
- Markopoulos, Theodore. 2009. *The Future in Greek: From Ancient to Medieval*. Oxford: Oxford University Press.
- Miller, Gary. 2002. *Nonfinite Structures in Theory and Change*. Oxford: Oxford University Press.
- Monro, David B. 1891. *A Grammar of Homeric Greek*. 2nd ed. Oxford: Clarendon Press.
- Philippaki-Warbuton, Irene. 2007. The Syntax of Classical Greek. In A.F. Christidis, ed., 590–598.
- Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. *Elements of Grammar: A Handbook of Generative Syntax* ed. by Liliane Haegeman, 281–337. Dordrecht: Kluwer.
- Roberts, Ian. 2007. *Diachronic Syntax*. Oxford: Oxford University Press.
- Roberts, Ian & Anna Roussou. 2003. *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486326
- Schwyzler, Eduard. 1988. *Syntax und syntaktische Stilistik*. München: Beck'sche Verlagsbuchhandlung.
- Sevdali, Christina. 2007. *Infinitival Clauses in Ancient Greek: Overt and Full Subjects, the Role of Case and Focus*. Cambridge: Cambridge University PhD dissertation.
- Smyth, Herbert Weir. 1918. *Greek Grammar*. Cambridge, MA: Harvard University Press.
- Vikner, Sten. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford: Oxford University Press.

Proto-Indo-European verb-finality

Reconstruction, typology, validation*

Hans Henrich Hock

University of Illinois at Urbana-Champaign

Although the reconstruction of Proto-Indo-European as verb-final is widely accepted, there continue to be dissenting opinions (e.g. Friedrich 1975). See e.g. Pires & Thomason (2008), who question the fruitfulness of Indo-European syntactic reconstruction. In this article I address two issues: First, the reconstructable subordination strategies, including relative-correlative structures, are perfectly in conformity with verb-final typology — *pace* Lehmann (1974) and Friedrich (1975) who considered relative clauses with finite verbs and relative pronouns incompatible with SOV. Second, verb-final reconstruction makes it possible to account for prosodic and segmental changes that single out finite verbs, such as the non-accentuation of Vedic finite verbs and *i*-apocope preferentially targeting finite verbs in Italic, Celtic, and Baltic-Slavic. Both developments find a natural, prosodically motivated explanation if we accept PIE as SOV, but not if we do not accept that reconstruction. These facts show that, *pace* Pires & Thomason (2008), the reconstruction of PIE as verb-final is a fruitful hypothesis.

1. The issues

There is a broad consensus that the unmarked constituent order of PIE was SOV or verb-final; e.g. Lehmann (1974), Watkins (1976), Mallory & Adams (1997). However, there is also a long tradition that questions the assumption. Wackernagel took the non-accentuation of simple finite verbs in Vedic main clauses as evidence for cliticization and hence assumed, with some reservations, that PIE main-clause order was verb-second (1892).

* A previous version of this article was read at the Workshop on Proto-Indo-European Syntax and Its Development, Thessaloniki, 1–2 April 2011. I am grateful to Nikolaos Lavidas and Leonid Kulikov for organizing the Workshop, and I have benefited from comments by Joseph Emonds and Thórhallur Eythórsson, as well as several referees. As usual, they may not agree with me on all details.

Wackernagel's account was accepted by Behaghel (1929) and used as an explanation of Germanic word order differences between main and dependent clauses. The account was also used by Friedrich (1975) as one of his arguments against Lehmann's (1974) typological approach to PIE syntax. However, the typological issue of PIE relative clauses played a more important role in the dispute between Friedrich and Lehmann.

Skepticism continues into the present. For instance, Pires & Thomason (2008) question the fruitfulness of Indo-European syntactic reconstruction, and Fortson remarks "As a matter of fact, the claim that many or most of the older Indo-European languages were verb-final has never been fully verified" (2010: 157).¹ See also the dialogue between Campbell & Harris (2002) on one hand and Lightfoot (2002a, 2002b) on the other concerning the GENERAL feasibility of syntactic reconstruction. For a recent evaluation of this issue see Hock (2010a).

In this article I briefly present empirical evidence favoring the reconstruction of PIE as verb-final, and then address two major issues that provide further support for PIE as verb-final. First, *pace* Lehmann and Friedrich, the relativization strategies reconstructable for the proto-language are not in conflict with verb-final typology, but are widespread in SOV languages. Second, verb-final reconstruction makes it possible to account for a number of prosodic and segmental changes that single out finite verbs in various Indo-European languages, a fact that shows that the verb-final hypothesis is a fruitful one.

The remainder of this article is organized as follows. Section 2 briefly summarizes the comparative evidence for PIE verb finality. Section 3 addresses contested issues regarding the reconstruction of relative clauses with finite verbs and the extent to which such clauses are compatible with SOV typology. Section 4 adduces prosodic evidence that supports the reconstruction of PIE as verb final and, at the same time, shows that the verb-final reconstruction is a fruitful hypothesis. Section 5 presents the conclusions.

2. Comparative evidence for verb-finality

The many skeptical voices notwithstanding, the evidence for PIE verb-finality (or SOV) is overwhelming. All the branches of ancient Indo-Iranian, Hittite and

1. Some scholars, such as Thomason (2006) and Harris (2008), accept the possibility of reconstruction for limited time depths such as Proto-Romance, but question the applicability of the method for Proto-Indo-European. Others, including Keydana (in press) have come to accept the method after earlier skepticism, "as long as we respect the limits (of external reconstruction) and restrain ourselves from speculating without sound empirical evidence".

the other early Anatolian languages, and Latin have clear evidence for this order, and given that they are not direct neighbors, that evidence should in principle be sufficient. For Germanic, too, recent work has established verb-finality as the original order, with V2 coming in as an innovation; for different perspectives see Hock (1982a), Pintzuk (1991), Eythórsson (1995, 1998), Dewey (2006), Suzuki (2008), and Section 4.1 below. Berneker (1900) finds both verb-initial and verb-final patterns in early Slavic (non-translation) texts; the common occurrence of verb-initial order in epics can be attributed to genre, while the SOV of didactic prose cannot be explained in this way and must therefore be considered unmarked (Hock 2000). While Modern Lithuanian is basically SVO, the syntactic examples cited by Schleicher (1856) are predominantly SOV (with or without V2); see also Hock (1982a). Even Greek, for which there has been a great amount of uncertainty and difference of opinion (see, e.g., Cervin 1990, Keydana in press), has been argued to have OV syntax in Homer, changing to VO syntax in later Greek (Taylor 1990, 1994). As for Celtic, Bergin (1938) has demonstrated the marginal persistence of verb finality in Middle Irish, beside dominant verb-initial order; see also Watkins (1962, 1963). What is especially significant is that in the European members of the Indo-European family the strongest evidence for SOV is found in their earliest stages, and that the V2, SVO, or VSO prevalent in later stages is an innovation. Under the circumstances, a reconstruction postulating anything other than SOV for Proto-Indo-European would be highly dubious.

Note however that this is a statement about the word order “pattern” or “construction” that can be reconstructed for PIE. What kind of synchronic grammar would generate this pattern is heavily dependent on the grammatical theory one adopts. For instance, in a grammatical framework such as Kayne’s (1994), which does not recognize a difference between base-generated right branching and left branching, the verb-final order would have to be accounted for through movement.

3. Relative clauses, SOV typology, and the reconstruction of PIE

This section addresses contested issues regarding the reconstruction of relative clauses with finite verbs and the extent to which such clauses are compatible with SOV typology. Section 3.1 deals with general typological issues. Section 3.2 argues that relative-correlative structures, without center-embedding of the relative clause, are compatible with SOV typology. Section 3.3 addresses the fact that the relative-correlative structures reconstructable for PIE differ in some interesting ways from relative-correlatives in other SOV languages. Section 3.4 discusses some apparent problems for the reconstruction. A summary and conclusion for the entire section is presented in Section 3.5.

3.1 General typological issues

While the evidence in Section 2 supports reconstructing PIE as SOV, questions have been raised as to whether this reconstruction is typologically compatible with other aspects of syntax that can be reconstructed, especially the structure of relative clauses. The foundation for this line of thinking is Greenberg's (1966) pioneering work on correlations between different aspects of word and clause order, and later modifications and additions to that work.

3.1.1 *Lehmann vs. Friedrich*

In the hands of Lehmann (1974) and Friedrich (1975), this line of thinking led to what Watkins (1976) has called the "straightjacket" of typology. Both scholars agreed that SOV correlates with prenominal relative clauses lacking relative pronouns, and that postnominal relative clauses with relative pronouns are a feature of VO languages. Against this background, Lehmann claimed that because PIE must be reconstructed as SOV, it cannot have had relative clauses containing relative pronouns. Conversely, Friedrich argued that the existence of such relative clauses in early Indo-European precludes reconstruction of PIE as SOV.

Underlying this argumentation is the notion that reconstructed languages must be "typologically consistent". However, natural languages frequently are typologically "inconsistent". For instance, English has SVO and prepositions, but the order adjective + noun, as well as genitive + noun as one option (the *s*-genitive); German has SOV with V2, but prepositions, and a prevalent noun + genitive order. Why should reconstructed languages be any more "consistent" than ordinary natural languages? (See Hock 1986/1991: 618–621, which is part of a more general discussion of attempts to invoke typology as an evaluative criterion in comparative reconstruction, as well as the recent remarks by Wichmann 2008: 6 regarding reconstruction in terms of ancillary properties.)

3.1.2 *Subordination and the finiteness constraint*

There is a more productive and meaningful approach that does not look for "typological consistency" in reconstructed languages, but rather addresses issues that have been subsumed under the notion "Strict SOV".

In its most extreme form, "Strict SOV" is defined as a typology that permits only one finite verb per sentence.² In complex sentences, therefore, subordinate

2. This definition causes some difficulties for Japanese, whose prenominal relative clauses appear to have finite verbs. However, the verb of these structures cannot (normally) be followed by sentence-final elements, such as *masita* 'pst'. Moreover, in earlier Japanese the relative structure is a type of nominalization, with a special form of the verb ("*rentai*") and with genitive marking of the subject; see, e.g., Shibatani (1987).

structures must be non-finite. An early characterization of this view is that of Caldwell (1913) in reference to Dravidian:

In every sentence there is but one finite verb, which is the last word in the sentence, and the seat of government; and all the verbs which express subordinate actions or circumstances, whether antecedent or contemporaneous, assume an indeterminate, continuative character, as verbal participles or gerundials, without the need of conjunctions or copulatives of any kind; so that the sense (and more or less the time also) waits in suspense for the authoritative decision of the final governing verb. (Caldwell 1913:488)

See also such publications as Kornfilt (1987) for Turkish and Turkic, Shibatani (1987) for Japanese, and Steever (1987) for Dravidian.

The precise nature of non-finite subordination may vary across different languages traditionally considered “Strict”. Common devices are infinitives, participles, and converbs, as well as nominalizations. The same devices are, of course, found in many VO languages (such as English), but the common perception is that in these languages they coexist with finite subordination and that the latter is more basic; see, e.g., Lehmann (1974, 1978), Friedrich (1975), Emeneau (1980), and Thomason & Kaufman (1988).

3.2 Finite relativization and SOV

In the following I focus on just one aspect of subordination, namely relativization, not only because this was a major issue of disagreement between Lehmann and Friedrich, but also because it is an issue where the traditional view, reflected in Caldwell’s statement, has been shown to be problematic.

First, starting with Downing (1973, 1978), Hale (1975), Andrews (1975/1985), and Comrie (1981), it has become clear that as an alternative to non-finite, participial relativization, many SOV languages employ relative-correlative constructions of the type in (1). The structures consist of a relative clause, containing a relative pronoun and a finite verb, and a main — or correlative — clause, containing a finite verb and commonly (but not necessarily) a correlative pronoun. Moreover, the relative clause is juxtaposed to the correlative clause, not inserted into it.

- (1) [RP ... V]_{RC} [CP ... V]_{CC}
 (+ finite) (+ finite)

While Keenan (1985) claimed that structures of this sort are found in “loose” SOV languages, it has now been shown that they are also found in languages usually considered “Strict”, including Dravidian (2), Turkish (3), and Burushaski (4); see Hock (1989) as well as Lipták (2009), both with references. More than that, they

seem to be a characteristic feature of (a subset of) SOV languages; VO languages do not seem to have such structures as their unmarked relativization strategy (regarding the post-RC particles of modern southern Dravidian, Burushaski, and other languages, see Hock 2008). An apparent exception, cited widely in the literature (e.g., Lipták 2009), is the existence of relative-correlatives in South Slavic languages. Closer examination, however, suggests that these are marked structures, used in indefinite/generic function, very much like English *who steals my purse steals trash*. South Slavic unmarked relative constructions are center-embedded, postnominal.

- (2) [evan nanrāka uṛaikkiṛāṇ]_{RC} =ō
 RP.NOM.SG.M hard work.PRS.3SG.M PCL
 [avan vaṛkaiyil munṇēruvāṇ]_{CC}
 CP.NOM.SG.M life.LOC.SG. succeed.FUT.3SG.M
 ‘Who works hard will progress in life.’ (Tamil)
- (3) [ne yıkılmış-sa]_{RC} [softalar yıkılmıştır]_{CC}
 RP demolish.PASS.COND.3SG bigot.PL demolish.PST.3PL
 ‘What(ever) has been demolished, the bigots have demolished.’ (Turkish)
- (4) [ámin daúčkuin imánimi]_{RC} ke
 RP hunter become.PRT.3SG PCL
 [íne buáa échap oóscói]_{CC}
 CP cow.GEN meat NEG-eat.PRS.3SG
 ‘Whoever has become a hunter, does not eat beef.’ (Burushaski)

The precise syntactic analysis of these structures is still a matter of debate, but starting with Andrews (1975/1985) and Hale (1975) there has been an increasing tendency to consider the relative clause to be adjoined to the correlative clause. An exception is Bhatt (2003), who argues for embedded, adnominal base generation of structures with single relative pronouns, but adjoined base generation if the relative clause contains more than one relative pronoun. See Davison (2009) for a critique of Bhatt’s account.

Let us at this point set aside the issue of syntactic analysis and turn our attention to early Indo-European. As noted by Watkins (1976), there is ample evidence for reconstructing relative-correlatives for the proto-language. See also the detailed studies by Haudry (1973) for Old Latin, Garrett (1994) for Hittite, Hettrich (1988) and Hock (1989) for Sanskrit, and the examples in (5–7). To these must be added the evidence of Homeric Greek (Chantraine 1963:236) and even early Germanic (Hock 1991), which however has the order CC–RC; see (8–9). Note further that in Greek and Latin, these structures are clearly archaic, giving way to post-NP constructions as in (10–11).

- (5) [de bacanalibus quei foideratei esent]_{RC}
 with bacchanal.ABL.PL.N RP.NOM.PL.M associated.NOM.PL.M be.SBJV.3PL
 [ita exdeicendum censuere]_{CC}
 thus declare.GDV.ACC.SG.N decide.PRF.3PL
 ‘(About those) who might be associated with the bacchanals, they decided to declare as follows.’ (SCBacch. 2–3)
- (6) [nu ku-it [(LU)]GAL-uš te-ez-zi]_{RC}
 CONN RP.ACC.SG.N king.NOM.SG.M say.PRS.3SG
 [nu a-pa-a-at i-ya-mi]_{CC}
 CONN CP.ACC.SG.N do.PRS.1SG
 ‘Now, what the king says, (now) that I will carry out.’ (KBo 17.1–7:Vs.2.17–18)
- (7) [yām u ha + eva tat paśavo manuṣyeṣu
 RP.ACC.SG.M PCLS then cattle.NOM.PL.M man.LOC.PL.M
 kāmam aroham]_{RC}
 desire.ACC.SG.M obtain.IPF.3PL
 [tam u ha + eva paśuṣu kāmam rohati]_{CC}
 CP.ACC.SG.M PCLS cattle.LOC.PL.M desire.ACC.SG.M obtain.PRS.3SG
 ‘What very desire the cattle then obtained among men, that very desire he (now) obtains among cattle.’ = ‘The very desire that the cattle then obtained among men he now obtains among cattle.’ (ŚB 2.1.2.7)
- (8) [hoús mèn speúdontas ídoi ...]_{RC}
 RP.ACC.PL.M PCL eager.ACC.PL.M see.OPT.3SG
 [toús mála tharsúneske]_{CC}
 CP.ACC.PL.M PCL encourage.ITER.IPF.3SG
 ‘which eager (Danaans) he might see, to them he would encourage ...’ (Il. 4.232–233)
- (9) [Ða wæs æt ðam geongan grim andswaru eðbegete
 then be.PST.3SG for that young.DAT.SG.M grim answer.NOM.SG.F easy-to-get
 þam]_{CC} [ðe Ø ær his elne forleas]_{RC}
 CP.DAT.SG.M RM (RP.NOM.) before his courage.DAT.SG.N lose.PST.3SG
 ‘Then a grim answer was easy to get for the young one who had earlier lost his courage.’ (Beow. 2860–61)
- (10) Ariaîos [hón hēmeîs êthélomen
 A.NOM.SG.M RP.ACC.SG.M we.NOM.PL want.IPF.1PL
 basiléa kathistánai]_{RC} ... hēmâs kakôs poieîn peirâtai
 king.ACC.SG.M establish.INF we.ACC.PL bad.ADV do.INF try.PRS.3SG
 ‘Ariaeus, whom we wanted to establish as king, ... is trying to treat us badly.’ (Xen. An. 3.2.5)

- (11) quomodo mortuos [qui istic sepultus
 as dead.NOM.SG.M RP.NOM.SG.M. here buried.NOM.SG.M
 est]_{RC} nec loqui nec sermonare potest
 be.PRS.3SG neither speak.INF nor converse.INF be able.PRS.3SG
 ‘Just as the dead (person) who is buried here can neither speak nor converse’
 (CIL 818, late 1st c. BC)

Significantly, as in the “Strict” non-Indo-European SOV languages with relative–correlatives, the relative clauses are not incorporated into the correlative clause but juxtaposed to it. In this sense, then, the relative–correlatives that we must reconstruct for PIE are perfectly compatible with the typology even of “Strict” SOV languages.³ (Recall that VO languages do not seem to have such structures as their unmarked relativization strategy.)

There are, however, a number of features in which Indo-European differs from (most) other SOV languages with relative–correlatives, and some of these features are typologically unusual enough to require reconstructing them for PIE. In addition, Probert (2006) has proposed an analysis of Hittite relative clauses which poses potential problems for this reconstruction (see also Keydana in press), and therefore needs to be addressed.

3.3 Some special features of early relative–correlatives

One typologically unusual feature is that (almost) all the early Indo-European languages permit both the order RC–CC and CC–RC; see, e.g., (12–13) from Vedic Sanskrit and Greek, beside (7–8) above. (As noted above, early Germanic has the order CC–RC.)

- (12) [tvám tám deva jihváyā bādhvasva]_{CC}
 you.NOM.SG CP.ACC.SG.M god.VOC.SG.M tongue.INS.SG.F drive-off.IMP.2SG
 [márto yó no jighāmsati]_{RC}
 mortal.NOM.SG.M RP.NOM.SG.M we.OBL slay.DESID.PRS.3SG
 ‘You, O God, drive off with your tongue that evil-doer who, a mortal, tries to slay us.’ (RV 6.16.32)
- (13) [allà khèrè tònn mèn katatháptein]_{CC}
 but need CP.ACC.SG.M PCL bury.INF

3. In a series of publications I have argued that it is this adjunction, rather than matrix-clause incorporation, which exempts these relative clauses from the otherwise active constraint against multiple finite verbs; see e.g. Hock (1989, 2008).

[hós ke thánēisi]_{RC}
 RP.NOM.SG.M PCL die.SBJV.3SG

‘But it is necessary to bury the one who is dead.’ (Il. 19.228)

A second, even more unusual feature is the evidence of Sanskrit, Greek, and Latin suggesting that relative and correlative clauses are not adjoined, but (asyndetically) conjoined. See Hock (1989, 1992), and note the examples in (14–16). In these, a single relative clause, containing two relative pronouns, is syntactically linked to two separate correlative clauses, each with its own correlative pronoun. Significantly, the verb accentuation in (14) shows that *tat tad bhavati* is a separate correlative (or “main” clause), a fact which precludes the analysis of (14) as in (14’). Structures of this sort can only be explained in terms of conjunction — a single relative clause cannot be simultaneously adjoined to two different “main” clauses (or underlyingly attached to two different NPs in two different clauses, for that matter).

- (14) [sā_i vai daiivī vāg]_{CC}
 CP.NOM.SG.F PCL divine speech.NOM.SG.F
 [yayā_i yad yad_j eva vadati]_{RC}
 RP.INST.SG.F RP.ACC.SG.N PCL speak.PRS.3SG
 [tat tad_j bhavati]_{CC}
 CP.NOM.SG.N come about.PRS.3SG

‘That_i is divine Speech by which_i whatever_j one speaks, that_j comes about.’ ≈
 ‘Whatever one speaks by means of divine Speech comes about.’ (BAU 1.3.27)

- (14’) * [sā_i vai daiivī vāg]_{CC}
 [yayā_i [yad yad_j eva vadati]_{RC} tat tad_j bhavati]_{CC}_{RC}

- (15) [tois d’ anéstē | Kálkhas_i ...]]_{CC}
 that.DAT.PL.M PCL stand-up.AOR.3SG Calchas.NOM.SG.M
 [hós_i éidē tá t’ eónta
 RP.NOM.SG.M know.PLPF.3SG that.ACC.PL.N & being.ACC.PL.N
 tá t’ essómena pró t’ eónta |
 that.ACC.PL.N & be.FUT.PTCP.ACC.PL.N before & being.ACC.PL.N
 kai néess’ hērēsát’ ... |
 & ship.DAT.PL.F lead.AOR.3SG
 hèn_j dià mantosúnēn]_{RC}
 RP.ACC.SG.F by soothsaying.ACC.SG.F
 [tén_j hoi póre Phoibos Apóllōn]_{CC}
 CP.ACC.SG.F he.DAT.SG give.AOR.3SG Ph-A.NOM.SG.M

‘But among them arose Calchas_i ... who_i knew what is, what will be, and what has been and (who) led the ships ... by the soothsaying that Phoebus Apollo had given him.’ (Lit.: ‘and (who) led the ships ... by which_j soothsaying, that_j soothsaying Ph. A. had given him.’) (Il. 1.68–72)

- NEG polished.NOM.SG.M CP.ABL.PL.F art.ABL.PL.F

[quas_i qui_j tenent]_{RC}
 RP.ACC.PL.F RP.NOM.PL.M hold.PRS.3SG

[Ø_j eruditi appellantur]_{CC}
(CP.NOM.PL.M) erudite.NOM.PL.M call.PASS.PRS.3PL

'... not polished in those_i arts which_i who_j have (they_j) are called erudite.' ≈
'... not polished in those arts on account of having which people are called
erudite.' (Cicero, Fin. 1.7.26)

Further evidence that the relationship between relative and correlative clauses is a fairly loose one, as would be expected in conjoined structures, comes from constructions in which the relative pronoun (and its NP) lacks a coreferent in the main clause or is picked up by a non-coreferring element, as in (17–19).

- (17) [yá indra yátayas tvā |
RP.NOM.PL.M Indra.VOC.SG.M Yati.NOM.PL.M you.SG.ACC

bhṛgavo yé ca tuṣṭuváh] _{RC}
 Bhṛgu.NOM.PL.M RP.NOM.PL.M & praise.PRF.3PL

[máma + íd ugra śrudhī hávam]_{CC}
I.GEN.SG PCLE strong.VOC.SG.M hear.IMP.2SG call.ACC.SG.M

‘Which Yatis, O Indra (praised) you, and which Bhṛguś have praised (you), O powerful one, hear my call.’ ≈ ‘Even if the Yatis and Bhṛguś have praised you, powerful Indra, nevertheless hear MY call.’ (RV 8.6.18)

- (18) [bélteron]_{CC} [hòs pheúgōn prophúgēi]_{RC}
better.NOM.SG.N RP.NOM.SG.M fleeing.NOM.SG.M escape.AOR.SBJV.3SG

'It is better if one escapes by fleeing.' (Lit. '[It is] better who escapes fleeing') (Il. 14.8).

- (19) [calcem partiaro coquendam qui
limestone.ACC.SG.F share.DAT.SG.N cook.GDV.ACC.SG.F RP.NOM.PL.M
dant]_{RC} [ita datur]_{CC}

give.PRS.3PL thus give.PASS.PRS.3SG

‘(People) who give limestone for burning on shares, it is given as follows.’
= ‘If people give’ (Cato, Agr.16)

In Hock (1989), I presented further evidence that Sanskrit relative-correlatives involve conjunction, rather than adjunction or derivation from an underlying post-nominal, center-embedded construction. As (20) shows, unlike “normal” relative clauses (whether adjoined or embedded), Sanskrit relatives may contain interrogatives. Since then I have found evidence that Sanskrit also permits imperative

modality in relative clauses, as in (21) — again, something not found in “normal” relative clauses.⁴ The latter phenomenon is also found in Latin, as in (22).⁵

- (20) śaryāto ha vā ikṣāṁcakre
 Śaryāta.NOM.SG.M PCLS reflect.PERPF.MID.3SG
 [yat **kīm** akaram] _{RC}
 RADV what.ACC.SG.N do.IPF.SG.1
 [tasmād idam āpadi] _{CC} + iti
 CADV this.ACC.SG.N get into.IPF.MID.1SG QUOT.
 ‘Śaryāta thought, “Because I have done what?, therefore I have gotten into this.” ≈ ‘... “What have I done to get into this?”’ (ŚB 4.1.5.4)
 (Thus also ŚB 1.7.3.19; a similar structure with *kva* ‘where’ at ŚB 5.1.3.13)
- (21) tyaje prāṇān na+ eva dadyān kapotaṁ (|)
 abandon.PRES.1SG breath.ACC.PL.M NEG PCL give.OPT.1SG dove.ACC.SG.M
 saumyo hy ayaṁ kiṁ na jānāsi ... |
 gentle.NOM.SG.M for this.NOM.SG.M Q NEG know.PRS.2SG
 [yathā kleśaṁ mā **kuruṣva**+ iha ... (|)] _{RC}
 RADV. trouble.ACC.SG.M NEG do.IMP.2SG here
 [na+ ahaṁ kapotam arpayiṣye kathamcit] _{CC}
 NEG I.NOM.SG. dove.ACC.SG.M hand-over.FUT.1SG anyhow
 ‘I abandon my life, but I may not at all give the dove; for he is gentle, don’t you know ...? So that “don’t you make” trouble here ..., I will not hand over the dove in any way.’ = ‘... so that you don’t make trouble here ...’
 (Mahābhārata 3, App. 21/5.82)
- (22) [Hic ille iam vertetur orbis] _{CC}
 here this.NOM.SG.M now turn.FUT.PASS.3SG cycle.NOM.SG.M
 [cuius naturalem motum atque circuitum
 RP.GEN.SG. natural movement.ACC.SG.M & going around.ACC.SG.M
 a primo **discite** agnoscere] _{RC}
 from beginning.ABL.SG.N learn.IMP.2PL discern.INF

4. This example was not included in Hock (1989) or (1992). Other examples (with second-plural imperatives) are found in Oberlies (2003: 173, n. 3). Note that only second-person imperatives are relevant, since the more hortative-like first and third persons can even be used in questions (examples in Speijer 1886: 264 and Meenakshi 1983: 174). The interpretation of RV 1.172.2 as containing an imperative in a relative clause is problematic; see Hettrich (1988: 140).

5. Greek, too, offers imperatives in relative clauses, as in *planēn phrásō hēn eggráphou* ... ‘I will describe (your) wandering, which inscribe-you [IMPERATIVE] ...’ = ‘which you should inscribe’ (Aeschylus, *Prometheus* 788). However, these only appear in later Attic and Ionic (Schwyzer 1949: 344).

‘Here now this cycle (of constitutions) will be turned whose natural movement and revolution “learn you” [IMPERATIVE] to discern from the beginning.’ = ‘... you should learn ...’ (Cicero, *De republica* 2.45; see Leumann et al. 1977b: 570–571 for further examples and references)

Given the typologically unusual nature of the early Indo-European RC–CC constructions, it is highly unlikely that they originate from independent innovations in the daughter languages. Relative–correlatives, with conjunction of the relative and correlative clauses, must therefore be reconstructed for Proto-Indo-European.

3.4 Apparent difficulties

There are, however, some remaining problems which might argue against the conclusion that PIE only had conjoined relative-correlatives. One of these is the existence of “replative” relative clauses, which is addressed in Section 3.4.1; the other is Probert’s claim (2006) that Old Hittite has embedded relative clauses, which is discussed in Section 3.4.2.

3.4.1 “Replative” relative clauses

One obvious potential difficulty for the view that PIE relative clauses are conjoined to their correlative counterparts is the existence of “replative” relative clauses, as in (23–25), where a relative clause seems to be embedded into a main clause as one of its constituents.

- (23) *tām etad devās ca paryāsatē*
 that.ACC.SG.F here god.NOM.PL.M & sit-around.PRS.3PL
 [*ye ca + ime brāhmaṇāḥ ...*]_{RC}
 RP.NOM.SG.M & this.NOM.PL.M brahmin.NOM.PL.M
 ‘And around it here sit both the Gods and (those) who (are) these brahmins ...’ (ŚB 1.3.3.8)
- (24) *tá t’ éldetai [hós k’ epideuēs]*_{RC}
 that.ACC.PL.N PCLE desire.PRS.3SG RP.NOM.SG.M PCL lacking.NOM.SG.M
 ‘which (he) who is lacking (them) may desire.’ (Il. 5.481)
- (25) *minime=que male cogitantes sunt*
 minimally=& bad.ADV. thinking.NOM.PL.M be.PRS.3PL
 [*qui in eo studio occupati sunt*]_{RC}
 RP.NOM.PL.M in that undertaking.ABL.SG.N occupied.NOM.PL.M be.PRS.3PL
 ‘And (those) who are occupied with this undertaking are minimally engaged in bad thoughts.’ (Cato, Agr. Intro)

Hettrich, in his account of Vedic (1988: 601, 608, 788, etc.) points out that replacive RCs usually appear at the margin of the clause to which they appear to belong. This suggests that they are not fully embedded into the “matrix” clause, and Hettrich proposes that the structures have undergone ellipsis, the “missing” head of the matrix clause being easily supplied from the context (1988: 694). See the alternative versions of (23–25) in (23'–25').

- (23') *tām etad devās ca paryāsatē*
 that.ACC.SG.F here god.NOM.PL.M & sit-around.PRS.3PL
 [te ca]_{CC} [ye + ime brāhmaṇāḥ ...]_{RC}
 CP.NOM.PL.M & RP.NOM.SG.M this.NOM.PL.M brahmin.NOM.PL.M
 ‘And around it here sit both the Gods and those who (are) these brahmins ...’
- (24') [tā t' éldetai hoûtos]_{CC}
 that.ACC.PL.N PCL desire.PRS.3SG CP.NOM.SG.M
 [hós k' epideuês]_{RC}
 RP.NOM.SG.M. PCL lacking.NOM.SG.M.
 ‘which (possessions he) who is lacking (them) may desire.’
- (25') [minime=que male cogitantes sunt ii]_{CC}
 minimally=& bad.ADV. thinking.NOM.PL.M be.PRS.3PL CP.NOM.PL.M
 [qui in eo studio occupati sunt]_{RC}
 RP.NOM.PL.M in that undertaking.ABL.SG.N occupied.NOM.PL.M be.PRS.3PL
 ‘And those who are occupied with this undertaking are minimally engaged in bad thoughts.’

The appropriateness of this account is supported by parallel structures such as the Sanskrit ones in (26). (Additional research on this issue would be desirable for other early Indo-European languages.) Corresponding to the “replacive” structure in (26a), the parallel version of the *Mādhyandina* recension shows the same structure followed by a correlative (26b), presumably with right extraction of the correlative pronoun, or perhaps truncated from something like (26c).

- (26) a. *etēna + u eva vyajayata*
 this.INSTR.SG.N PCLS win.IPF.3SG
 [yā + iyam asya vijitis]_{RC}
 RP.NOM.SG.F this.NOM.SG.F he.GEN.SG victory.NOM.SG.F
 ‘Through this he won what victory (there is) of him now.’ = ‘Through this he won the victory that now is his.’ (ŚB (K) 1.3.4.9)
- b. *etēna + u eva vyajayata*
 this.INSTR.SG.N PCLS win.IPF.3SG
 [yā + asya iyam vijitis]_{RC}
 RP.NOM.SG.F he.GEN.SG this.NOM.SG.F victory.NOM.SG.F

[tām]_{CC}

CP.ACC.SG.F

‘Through this he won what victory (there is) of him now, that one.’ =

‘Through this he won the victory that now is his.’ (ŠB (M) 2.4.4.15)

- c. etēna + u eva vyajayata

this.INSTR.SG.N PCLS win.IPF.3SG

[yā + asya iyam vijitis]_{RC}

RP.NOM.SG.F he.GEN.SG this.NOM.SG.F victory.NOM.SG.F

[tām eva vyajayata]_{CC}

CP.ACC.SG.F PCL win.IPF.3SG

‘Through this he won, whatever victory is his now, that he won.’

“Replacives”, thus, do not constitute an obstacle to reconstructing conjoined RC–CCs for Proto-Indo-European (further apparent problems are discussed in detail in Hock 1989, with ample references to Hettrich 1988).

3.4.2 Old Hittite “embedded” relatives

A recent paper by Probert (2006) has raised what appear to be more serious questions as to whether relative–correlatives (however analyzed) are to be reconstructed, and her skepticism has been taken up by Keydana (in press).

According to Probert, Old Hittite offers two different types of relative construction. One has both a “conjunction” or “boundary marker” and a resumptive pronoun in the main clause, see (27 = 6), the other has neither (28–29). Probert finds only one example that seems to unambiguously violate this distinction, by containing a resumptive pronoun (written heterographically) but no conjunction.

- (27) [nu ku-it ((LU))GAL-uš te-ez-zi]_{RC}
CONN RP.ACC.SG.N king.NOM.SG.M say.PRS.3SG

[nu a-pa-a-at i-ya-mi]_{CC}
CONN CP.ACC.SG.N do.PRS.1SG

‘Now, what the king says, (now) that I will carry out.’

(KBo 17.1–7:Vs.2.17–18)

- (28) ku-e-la=aš ar-ḫi a-ki
RP.GEN=he.NOM.SG land.LOC.SG die.PRS.3SG
1 ME ^{GIŠ}gi-pé-eš-šar A.ŠĀ kar!-aš-ši-i-ez-zi
one hundred gipeššar.ACC land cut off.PRS.3SG

‘... whoever’s land (he_i) dies on, (he_i) shall cut off 100 *gipeššar* of the land (of that person_i).’ (KBo 6.2 i 7–8. Laws series one, §6)

- (29) pa-ap-re-ez-zi ku-iš
be impure.PRS.3SG RP.NOM.SG.M

3 *GÍN* *KÛ.BABBAR* pa-a-i
 3 shekel.ACC silver give.PRS.3SG
 ‘(the one) who is impure (he) gives three shekels of silver’
 (KBo 6.2 i 57. Laws series one § 25)

According to Probert, the relative clauses in most sentences of the type in (28–29) function as subjects of the corresponding matrix clause (2006: 54–55) and should therefore be analyzed as fronted from an underlyingly subject position embedded in the matrix clause (2006: 73–74); see Figure 1.

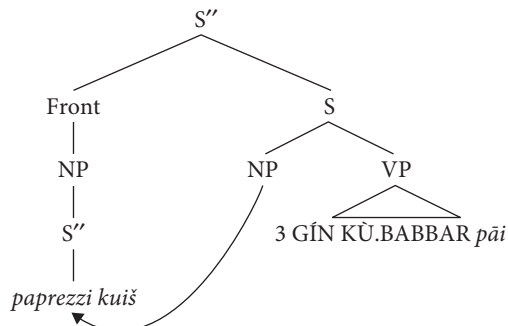


Figure 1. Probert's (2006: 74) account of example (29)

Probert further offers example (30) as evidence for embedding, with part of the main clause preceding the relative clause and part of it following.

- (30) nu a-pé-「e」-[el É-SÛ] 「ku-e-la」^{GIŠ}e-ia-an
 CONN CP.GEN house.NOM-his RP.GEN eyan tree.NOM
 a-aš-ki-iš=ši ša-「ku-wa」-a-an 「a」-[pé-ni-iš-ša-an]
 gate.LOC=he.LOC visible.NOM likewise
 ‘of hi(m his house) of **whoever at his gate an eyan-tree is visible** is
 (likewise).’ I.e. ‘The house of him at whose gate an *eyan*-tree is visible is
 likewise exempt.’ (KBo 6.2 ii 61–62, Laws series one, § 50)

Perhaps significantly, the structure differs from Probert's normal “embedded” structures by having a “boundary” marker as well as a correlative pronoun in the main clause part preceding the relative clause. More important, the passage is immediately preceded by the conditional clause in (31) and forms part of a longer conditional structure initiated by that clause (see Hoffner 1997: 61–62).

- (31) 「ma-a-an」^{URU}A-ri-in-na 11 ITU-aš ti-ez-zi
 if Arinna.LOC 11 month.NOM.SG. enter.PRS.3SG
 ‘If the 11th month arrives in Arinna’ (KBo 6.3 II 1)

The complete translation therefore should run something like this: ‘If the eleventh month arrives in Arinna, the house of him at whose gate there is an *eyan* tree, likewise, with implicit carry-over of ‘exempt’ from the passage that precedes our example. The passage, thus, is multiply complex, and the part that follows the relative clause is not only structurally reduced but may also be dependent on both the preceding RC and the initial conditional clause. Moreover, the expression ‘likewise’, whether in Hittite form (as in our example) or written heterographically (as *QAT-TAM-MA(=pát)*), frequently occurs at the end of extended conditional structures, a common feature of Hittite law texts. It is therefore possible that the final ‘likewise’ of our example has been right-extracted, influenced by the model of the other conditional structures.

At any rate, the interpretation of the passage is too uncertain for (30) to be considered reliable evidence for Old Hittite relativization.

Regarding the other structures, Probert notes that her “embedded” type is limited to Old Hittite and thus must be considered an archaism within the history of the language. Moreover, she entertains the idea that the relative-correlative type in (27), which persists into later Hittite, results from reanalysis of the embedded type (2006: 76). She concludes that this may suggest a need to reconsider the usual PIE reconstruction, but leaves open the alternative that her “embedded” type was a relatively recent innovation that did not persist (2006: 79).

The (near-)complementarity of the relative-correlatives and “embedded” relatives unearthed by Probert is certainly interesting and constitutes an important new insight into the syntax of Old Hittite. At the same time, problems remain. One of these concerns her formal claim that the embedded relative has been fronted from subject position. More important for present purposes is the fact that there is room for adopting her alternative conclusion, namely that the “embedded” type may have been a transitory phenomenon.

Let me start with the issue of the formal analysis. Although in five of the eight Old Hittite “embedded” constructions that Probert has identified, the relative clause can be considered to function as subject of its main clause (to use her terminology), this does not hold true for the remaining three. Two of these are of the type in (28), involving a genitival relation (and the head of the genitive in 28 is a locative); and one has a dative relation. These constructions, nearly half of the total, would cause difficulties for her “fronting” account from the subject position of the main clause.

This problem can be mitigated by assuming that the structures are replacives, with deleted CP “head”. But in that case her “embedded” structures are simply analyzable as truncated relative-correlatives, similar to what we find in the other early Indo-European languages. As noted earlier (Section 3.4.1), replacives typically occur at the margin of the clause with which they are associated, a fact which would

provide a good motivation for the (left-)peripherality of Probert's "embedded" relatives, no matter what their syntactic status may be within the "main" clause.

To lay the foundation for an alternative account, under which the "embedded" relatives are a transitory phenomenon, it is useful to disaggregate Probert's two defining features for these structures and to examine their historical motivations, in terms both of diachronic developments specific to Hittite (or Anatolian) and of the larger Indo-European picture.

As is well known, the Hittite "conjunctions" or "boundary markers", such as the repeated *nu* in (28), are an innovation.⁶ They are not normally employed in the other early Indo-European languages; they are frequently absent in Old Hittite (not only in Probert's "embedded" RCs, but also in conditionals and even more generally; see Hoffner 1997: 12, van den Hout 2003: 181), and they become (near-)obligatory in later Hittite.

The absence of conjunctions or boundary markers in structures like (28–29) can therefore be considered an archaism. Note also Probert's comment that "In the language as a whole ... the marking of clause boundaries is increasingly ubiquitous after the Old Hittite period, and so it is not surprising that the real variation that arises lies in the possibility of leaving the resumptive element implicit" (2006: 78).

As for the "resumptive pronoun", by contrast, there is no evidence that its presence — or absence — is an innovation. Rather, the evidence of both Hittite and, as we have seen, the other early Indo-European languages shows that correlative pronouns could be deleted. In this regard, then, Probert's "embedded" relatives are no more archaic (or innovative) than her relative-correlatives. These facts permit the following alternative account.

Probert's "embedded" constructions reflect Indo-European replacive relatives with deleted correlative pronouns. Her relative-correlatives are reflexes of the same construction, but with overt correlative pronouns, plus the additional, innovative feature of "boundary marking". The (near-)restriction that structures without overt resumptive elements also must lack "boundary markers" would then be a temporary, transitional phenomenon in Old Hittite, a possibility envisioned by Probert.⁷

6. A parallel to this typologically rare phenomenon is found in the Middle East Iranian language Sogdian, in which (nearly) all clauses have a second-position enclitic "complementizer" $=(\text{a})t < \text{uta}$ 'and, also'; see Yoshida (2009: 314–315 and *passim*).

7. One of the referees raises the question whether the difference between this traditional stage and the later period can be explained in terms of a change (reanalysis) in the syntactic behavior of the boundary marker. At this point I am not prepared to answer this question with any degree of certainty. It is *a priori* possible to interpret the boundary marker as establishing greater syntactic independence of the constituent elements of complex sentences and thus requiring overt specification of constituents in each clause; as a consequence correlative-pronoun deletion would no longer be possible. However, without a fuller investigation of all complex clauses,

A shortcoming of the present account would seem to be that it fails to provide a motivation for this transitional stage. On the other hand, by relating the difference between Probert's two types of structures to the independently attested innovation of increasingly (near-)obligatory use of "boundary markers" and to the independently attested possibility of resumptive-pronoun deletion in the other early Indo-European languages, it manages to provide a sound historical account for the Old Hittite situation.

3.5 Conclusions regarding relativization

It therefore seems reasonable to stick to the reconstruction of conjoined relative-correlatives for Proto-Indo-European and the conclusion that this particular type of relativization is eminently compatible with SOV typology, *pace* Lehmann (1974) and Friedrich (1975).

This, however, only covers part of the story. There are also considerable differences between the early Indo-European languages. Hittite vastly prefers left-conjoined relative clauses, while early Germanic has the opposite order. Greek and Latin tend to shift toward post-nominal relative clauses and so, eventually, does Germanic. Sanskrit does not participate in this development, and only some of the modern Indo-Aryan languages are beginning to use post-nominal relative clauses.

There are differences, too, as regards the use of finite subordination (relative-correlatives) and non-finite subordination (participles, converbs, infinitives, and nominalizations). The Romans called the Greeks *Philométokhoi*, which can be loosely translated as 'participle-happy'. Non-finite relativization (by means of participles) is also common in Latin and Sanskrit. On the other hand, Hittite and early Germanic use non-finite relativization much more sparingly.

Finally, with its conjoined relative-correlatives and the availability of both RC-CC and CC-RC, Indo-European differs markedly from other SOV languages. Some SOV languages, in fact, do not have relative-correlatives at all (including Japanese, Korean, and Classical Tibetan). Languages such as Dravidian, Turkic, and Burushaski permit only the order RC-CC, and there is no evidence for conjoined, rather than adjoined status of the relative clauses. Within these languages, some require post-RC particles (e.g. modern southern Dravidian and Burushaski, as in examples 2 and 4), others do not (e.g. Turkic, as in 3).

There is, thus, ample room for more research, both on early Indo-European and in terms of crosslinguistic typological work.

both before and after the (near-)obligatoriness of boundary markers, such an account is entirely speculative.

4. Prosodically motivated changes confirming the verb-final reconstruction

I conclude this paper by demonstrating that the Indo-European verb-final reconstruction is a fruitful hypothesis, helping to explain segmental and suprasegmental changes in four different branches of Indo-European — Vedic Sanskrit, Italic, Insular Celtic, and Baltic-Slavic — which would be difficult to motivate without the hypothesis.

4.1 Suprasegmental effects of verb finality

As is well known, (simple) Vedic finite verbs are unaccented in main clauses (unless initial in the clause or poetic line), but accented in dependent structures; see (32).

- (32) a. *agne* [*yám* *yajñám* ...
 Agni.VOC.SG.M RP.ACC.SG.M sacrifice.ACC.SG.M
paribhúr *ási*]_{DC}
 surrounding.NOM.SG.M be.PRS.2SG
 [*sá* *íd* *devéṣu* *gacchati*]_{MC}
 CP.NOM.SG.M PCL god.LOC.PL.M go.PRS.3SG
 ‘O Agni, the sacrifice ... that you surround goes to the Gods.’ (RV 1.1.4)
- b. [*tásmai* *víśaḥ* *svayám evá namante*]_{MC}
 CP.DAT.SG.M people.NOM.PL.F self PCL bow.PRS.3PL
 [*yásmin* *brahmá* *púrva* *éti*]_{DC}
 RP.LOC.SG.M brahmin.NOM.SG.M first.NOM.SG.M go.PRS.3SG
 ‘Even the common people bow to him for whom the brahmin goes first.’
 (RV 4.50.8)

A comprehensive, prosodically motivated account for this difference was first provided by Klein (1992) with reference to Hock (1986/1991).⁸ According to Klein, non-accentuation of main-clause verbs arose in unmarked, or “canonical” constructions with the main-clause verb in sentence-final position, and with resolution of the conflict between the high pitch of the verb accent and the low pitch of sentence-final falling intonation through loss of verb accent.

Klein’s account has been developed further by Hock (2014), who argues that the main-clause accent loss results from an earlier stage of accent retraction in canonical utterance-final position and that this retraction also accounts for the well-known accent retraction of Greek finite verbs. The major ingredients of the

8. A full discussion of the prosodic impact of utterance-finality and its special effects in verb-final languages is presented in Hock (1999), with references to earlier literature and ample cross-linguistic evidence, including a near-parallel to the Vedic accent loss (or retraction) found in Modern Persian. For the accentual effects of verb-finality see also Ladd (1996).

Under the circumstances, an approach that accounts for Germanic V2 in terms of internal developments would be preferable. For the time being see Hock (1982a, 1986/1991: 330–336).

4.2 Segmental effects of verb finality

In two recent articles (Hock 2006, 2007) I have argued that the preferential application of *i*-apocope in finite verbs in Italic, Insular Celtic, and Baltic-Slavic can likewise be accounted for as the result of canonical utterance-finality of the verb in an SOV language. (See Hock 2010b for a comprehensive discussion.)

For instance, in Latin there is a difference in the scope of *i*-apocope between nouns and verbs. Nominal apocope is limited, such that disyllabics resist apocope (34a), but there is no such restriction for finite verbs (34b).⁹ Old Latin *tremonti* (*Carmen Saliare*) may possibly be a finite verb form with preserved final *-i*, but the text is notoriously difficult to interpret. In the absence of similar Old Latin structures, the significance of this form is difficult to assess.

- | | | | | |
|------|-----|----------|---|-------------------|
| (34) | a. | *animali | > | animal ‘animated’ |
| | vs. | *mari | > | mare ‘sea’ |
| | | *pedi | > | pede ‘by foot’ |
| | b. | *esti | > | est ‘is’ |
| | | *eyti | > | it ‘goes’ |
| | | *weniti | > | uenit ‘comes’ |

Traditional accounts note the difference but generally fail to provide an explanation; see, e.g., Leumann et al. (1977a: 92–93). Some scholars, such as Sihler (1995: 69) and Meiser (1998: 73–74) propose “quasi-morphological” accounts (Sihler’s term), with Meiser specifying the condition as distinguishing between an *i* that was always unaccented (in verbal endings) and an *i* that may show accent variation (in nominal endings). The only exception is Weiss (2009: 146–147) who, apparently unaware of Hock (2006), states that “The verbal endings, which were typically utterance-final ..., were subject to an utterance-final reduction, which did not normally affect the nominal forms”.

The notion of morphologically conditioned sound change, however, is problematic, especially if phonological explanations are readily available (Hock 1976). In our case, such an explanation is in fact available, in terms of the canonical utterance-final position of finite verbs in SOV languages and the fact that this is the

9. Leumann is probably correct in attributing the apparent apocope of *-i* in adverbs and particles, such as *eti* > *et*, to syncope in close collocations. Consider the parallel *nequ(e)-dum* > *necdum* (Leumann et al. 1977a: 93).

position of greatest prosodic weakness; see again Hock (1999) for details. This specific explanation is in large measure indebted to Harms's account (1964, 1990) for the loss of final *-e* in Finnish finite verbs.

4.3 Verb-finality effects as evidence that the SOV reconstruction is a fruitful hypothesis

Now, the prosodic explanations for Vedic (and Greek) finite-verb accentuation and preferential *i*-apocope in finite verbs clearly rests on the reconstruction of PIE as having unmarked SOV order, for only under this assumption is it possible to provide a phonologically grounded explanation for the various phenomena in terms of responses to utterance finality. The explanation, thus, provides additional support for the reconstruction of Proto-Indo-European as verb-final.

Moreover, by providing a prosodically motivated explanation for sound changes that would otherwise have to be considered morphologically (or syntactically) motivated, the present account shows that the reconstruction of PIE as verb-final is a fruitful hypothesis — *pace* Pires & Thomason's (2008: 50) claim that “no generally-accepted research results have shown any hypothesis about PIE word order to be scientifically fruitful, enabling historical linguists to explain things that are otherwise mysterious.”

5. Summary and conclusions

There is, thus, strong evidence for the hypothesis that the unmarked order of PIE was verb final. As shown in Section 2, the earliest attestations of Indo-European languages overwhelmingly have verb-final order, even Germanic which early on begins to change toward V2. The SOV reconstruction receives further support from phonological and prosodic changes which target finite verbs and which can be phonetically motivated only by assuming that they were conditioned by utterance-finality (Section 4). A major concern in early typologically informed publications (Lehmann 1974, Friedrich 1975) was the comparative evidence for relative clause structures containing relative pronouns and finite verbs, for such structures were considered incompatible with SOV typology. More recent research has shown that relative clauses of this sort are perfectly compatible with SOV typology, as long as they occur in relative–correlative constructions, i.e. structures not center-embedded into matrix clauses. The extensive discussion in Section 3 confirms that it is this kind of “SOV-compatible” relative structures that needs to be reconstructed for PIE and that apparent evidence for embedded structures turns out to be inconclusive.

Something that has not been addressed in this article is the question of how “SOV” or “verb finality” and relative–correlative constructions should be accounted for in a formal grammar. Answers to this question are heavily dependent on the grammatical theory one adopts. Thus, as noted in Section 2, a grammatical framework such as Kayne’s (1994), which does not recognize a difference between base-generated right-branching and left-branching, would have to account for verb-final order through various movements whose major motivation would be to align constituents in the right surface order. The case may be better for relative–correlatives, for as seen in Section 3.3, the evidence of early Indo-European suggests that they were conjoined, rather than adjoined. For Sanskrit, Davison (2009) suggest to account for this as CP (rather than IP) adjunction.

Finally, the findings of this article should not be taken to mean that PIE was a “Strict SOV” language in the way that, say, Dravidian languages can be said to be. As noted in Section 3.3, the fact that we must reconstruct not only the order RC–CC, but also CC–RC is typologically unusual; other SOV languages only have the order RC–CC. Moreover, as is well known, all of the early Indo-European languages have an alternative, verb-initial word order. Now, it is true that verb-initial order is marked (see, e.g., Dressler 1969, as well as Bauer 2011 for Hittite); and as noted in Section 4, the order DC–MC (i.e. RC–CC) was unmarked, rather than CC–RC (witness the fact that Finality-induced accent loss applies only in the MC/CC, not in the DC/RC). Still, the fact remains that among the SOV languages, early and Proto-Indo-European holds a typologically unusual position, as far as both relative-clause syntax and word order are concerned. It is tempting to speculate whether there is a correlation between the relative-clause exceptionality and the word order one; but I am not aware of any evidence or arguments that would raise this idea beyond the level of speculation.

Abbreviations

In addition to the “Standard Abbreviations” of the Leipzig glosses, the following abbreviations are used.

+	an indication that Sanskrit sandhi has been resolved for greater clarity	ITER	iterative
&	‘and’ (conjunction)	MID	middle voice
AOR	aorist	OPT	optative
CADV	correlative adverb	PCL	particle
CC	correlative clause	PERPF	periphrastic perfect
CONN	connective	PLPF	pluperfect
CP	correlative pronoun	PRT	preterit
		RC	relative clause

DESID	desiderative	RADV	relative adverb
GDV	gerundive	RM	relative marker
IPF	imperfect	RP	relative pronoun

References

- Andrews, Avery D., III. 1975 [1985]. *Studies in the Syntax of Relative and Comparative Clauses*. Cambridge: Massachusetts Institute of Technology PhD dissertation. (Reprinted in “lightly retouched version”, 1985, New York/London: Garland.)
- Bauer, Anna. 2011. Verberststellung im Hethitischen. *Indogermanistik und Linguistik im Dialog: Akten der XIII. Fachtagung der Indogermanischen Gesellschaft Salzburg* ed. by Thomas Krisch, Thomas Lindner, Michael Crombach & Stefan Niederreiter, 39–48. Wiesbaden: Reichert.
- Behaghel, Otto. 1929. Zur Stellung des Verbs im Germanischen und Indogermanischen. *Zeitschrift für vergleichende Sprachforschung* 56.276–281.
- Bergin, Osborn J. 1938. On the Syntax of the Verb in Old Irish. *Ériu* 12.197–214.
- Berneker, Erich Karl. 1900. *Die Wortfolge in den slavischen Sprachen*. Berlin: Behr.
- Bhatt, Rajesh. 2003. Locality in Correlatives. *Natural Language & Linguistic Theory* 21.485–541. DOI: 10.1023/A:1024192606485
- Caldwell, Robert. 1913. *A Comparative Grammar of the Dravidian or South-Indian Family of Languages*. 3rd edition, revised and edited by J.L. Wyatt & T. Ramakrishna Pillai. Reprinted 1974, New Delhi: Oriental Books Reprint Corporation.
- Campbell, Lyle & Alice C. Harris. 2002. Syntactic Reconstruction and Demythologizing “Myths and the Prehistory of Grammars”. *Journal of Linguistics* 38.599–618. DOI: 10.1017/S002226702001706
- Cervin, Richard Stuart. 1990. *Word Order in Ancient Greek: VSO, SVO, SOV, or All of the Above?* Urbana-Champaign: University of Illinois PhD dissertation.
- Chantraine, Pierre. 1963. *Grammaire homérique. 2: Syntaxe*. Paris: Klincksieck.
- Comrie, Bernard, ed. 1987. *The World’s Major Languages*. Oxford: Oxford University Press.
- Comrie, Bernard. 1981. *Language Universals and Linguistic Typology: Syntax and Morphology*. Chicago: University of Chicago Press.
- Davison, Alice. 2009. Adjunction, Features, and Locality in Sanskrit and Hindi-Urdu Correlatives. In Anikó Lipták, ed., 223–262. DOI: 10.1075/lfab.1.10dav
- Dewey, Tonya Kim. 2006. *The Origins and Development of Germanic V2: Evidence from Alliterative Verse*. Berkeley: University of California PhD dissertation.
- Downing, Bruce T. 1973. Correlative Relative Clauses in Universal Grammar. *Minnesota Working Papers in Linguistics and Philosophy of Language* 2.1–17.
- Downing, Bruce T. 1978. *Some Universals of Relative Clause Structure. Universals of Human Language. Vol. 4: Syntax* ed. by Joseph H. Greenberg, 375–418. Palo Alto, CA: Stanford University Press.
- Dressler, Wolfgang U. 1969. Eine textsyntaktische Regel der indogermanischen Wortstellung. *Zeitschrift für vergleichende Sprachforschung* 83.1–25.
- Emeneau, Murray B. 1980. *Language and Linguistic Area: Essays selected by A. S. Dil*. Stanford, CA: Stanford University Press.

- Eythórsson, Thórhallur. 1995. *Verbal Syntax in the Early Germanic Languages*. Ithaca, NY: Cornell University PhD dissertation.
- Eythórsson, Thórhallur. 1998. Zur historisch-vergleichenden Syntax des Verbums in Germanischen. *Sprache und Kultur der Indogermanen; Akten der X. Fachtagung der Indogermanischen Gesellschaft, 22–28. September 1996* ed. by Wolfgang Meid, 395–410. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Ferraresi, Gisella & Maria Goldbach, eds. 2008. *Principles of Syntactic Reconstruction*. Amsterdam: John Benjamins. DOI: 10.1075/cilt.302
- Fortson, Benjamin W., IV. 2008. *Language and Rhythm in Plautus: Synchronic and Diachronic Studies*. Berlin: de Gruyter. DOI: 10.1515/9783110210026
- Fortson, Benjamin W., IV. 2010. *Indo-European Language and Culture: An Introduction*. 2nd ed. Oxford: Blackwell.
- Friedrich, Paul. 1975. *Proto-Indo-European Syntax*. Butte, MT: Journal of Indo-European Studies.
- Garrett, Andrew. 1994. Relative Clause Syntax in Lycian and Hittite. *Die Sprache* 36.29–69.
- Greenberg, Joseph H. 1966. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. *Universals of Language: Report of a Conference Held at Dobbs Ferry, New York, April 13–15, 1961* ed. by Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press.
- Hale, Kenneth. 1975. Gaps in Grammar and Culture. *Linguistics and Anthropology: In Honor of C. F. Voegelin* ed. by Marvin Dale Kinkade, Kenneth Locke Hale & Oswald Werner, 295–315. Amsterdam: John Benjamins.
- Harms, Robert T. 1964. Review of The Structure and Development of the Finnish Language by Lauri Hakulinen (Bloomington, IN: Curzon, 1960). *Word* 20.105–114.
- Harms, Robert T. 1990. Synchronic Rules and Diachronic “Laws”: The Saussurean Dichotomy Reaffirmed. *Research Guide on Language Change* ed. by Edgar C. Polomé, 313–323. Berlin: Mouton de Gruyter.
- Harris, Alice. 2008. Reconstruction in Syntax: Reconstruction of Patterns. In Gisella Ferraresi & Maria Goldbach, eds., 73–96. DOI: 10.1075/cilt.302.05har
- Harris, Alice C. & Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620553
- Haudry, Jean. 1973. Parataxe, hypotaxe et corrélation dans la phrase latine. *Bulletin de la Société de Linguistique de Paris* 68.147–186.
- Hettrich, Heinrich. 1988. *Untersuchungen zur Hypotaxe im Vedischen*. Berlin: de Gruyter. DOI: 10.1515/9783110892123
- Hock, Hans Henrich. 1976. Review of An Introduction to Historical and Comparative Linguistics by Raimo Anttila (New York: Macmillan, 1972). *Language* 52.202–220. DOI: 10.2307/413217
- Hock, Hans Henrich. 1982a. AUX-cliticization as a Motivation for Word Order Change. *Studies in the Linguistic Sciences* 12:1.91–101.
- Hock, Hans Henrich. 1982b. Clitic Verbs in PIE or Discourse-based Verb Fronting? Sanskrit *sá hovāca gárgyaḥ* and Congeners in Avestan and Homeric Greek. *Studies in the Linguistic Sciences* 12:2.1–38.
- Hock, Hans Henrich. 1986/1991. *Principles of Historical Linguistics*. 1st and 2nd ed. Berlin: Mouton de Gruyter.
- Hock, Hans Henrich. 1989. Conjoined We Stand: Theoretical Implications of Sanskrit Relative Clauses. *Studies in the Linguistic Sciences* 19:1.93–126.

- Hock, Hans Henrich. 1991. On the Origin and Development of Relative Clauses in Early Germanic, with Special Emphasis on Beowulf. *Stæfcræft: Studies in Germanic linguistics* ed. by Elmer A. Antonsen & Hans Henrich Hock, 55–89. Amsterdam: John Benjamins. DOI: 10.1075/cilt.79.05hoc
- Hock, Hans Henrich. 1992. Reconstruction and Syntactic Typology: A Plea for a Different Approach. *Explanation in Historical Linguistics* ed. by Garry W. Davis & Gregory K. Iverson, 105–121. Amsterdam: John Benjamins. DOI: 10.1075/cilt.84.08hoc
- Hock, Hans Henrich. 1999. Finality, Prosody, and Change. *Proceedings of LP '98* ed. by Osamu Fujimura, Brian D. Joseph & Bohumil Palek, 15–30. Prague: Karolinum Press.
- Hock, Hans Henrich. 2000. Genre, Discourse, and Syntax in Sanskrit. *Textual Parameters in Older Languages* ed. by Susan C. Herring, Pieter van Reenen & Lene Schøsler, 163–195. Amsterdam: John Benjamins. DOI: 10.1075/cilt.195
- Hock, Hans Henrich. 2006. The Insular Celtic Absolute: Conjunct Distinction Once Again: A Prosodic Proposal. *Proceedings of the Sixteenth UCLA Indo-European Conference. Los Angeles, Nov. 5–6, 2005* ed. by Karlene Jones-Bley, Martin E. Huld, Angela Della Volpe & Miriam Robbins Dexter, 153–172. (= The Journal of Indo-European Studies Monograph, 50.) Washington: Institute for the Study of Man.
- Hock, Hans Henrich. 2007. Morphology and i-apocope in Slavic and Baltic. *Proceedings of the Eighteenth UCLA Indo-European Conference. Los Angeles, Nov. 3–4, 2006* ed. by Karlene Jones-Bley, Martin E. Huld, Angela Della Volpe & Miriam Robbins Dexter, 65–76. (= The Journal of Indo-European Studies Monograph, 53.) Washington: Institute for the Study of Man.
- Hock, Hans Henrich. 2008. Dravidian Syntactic Typology: A Reply to Steever. *Annual Review of South Asian Languages and Linguistics* ed. by Rajendra Singh, 164–198. Berlin: Mouton de Gruyter.
- Hock, Hans Henrich. 2010a. Review of Ferraresi & Goldbach, eds. (2008). *Kratylos* 55.1–12.
- Hock, Hans Henrich. 2010b. Phrasal Prosody and the Indo-European Verb. Paper at the Arbeitstagung “The Indo-European Verb” of the Indogermanische Gesellschaft, UCLA, September 2010.
- Hock, Hans Henrich. 2014. Vedic Verb Accent Revisited. *Vedic and Sanskrit Historical Linguistics. Papers from the 13th World Sanskrit Conference* ed. by Jared Klein & Elizabeth Tucker, 153–178. Delhi: Motilal Banarsidass.
- Hoffner, Harry A. 1997. *The Laws of the Hittites: A Critical Edition*. Leiden: Brill.
- van den Hout, Theo. 2003. Studies in the Hittite Phraseological Construction I: Its Syntactic and Semantic Properties. *Hittite Studies in Honor of Harry A. Hoffner, Jr. on the Occasion of his 65th Birthday* ed. by Gary M. Beckman, Richard Henry Beal & Gregory McMahon, 177–204. Winona Lake, IN: Eisenbrauns.
- Kayne, Richard. 1994. *The Antisymmetry of Syntax*. (= Linguistic Inquiry Monographs, 25.) Cambridge, MA: MIT Press.
- Keenan, Edward. 1985. Relative Clauses. *Language Typology and Syntactic Description, Volume II: Complex Constructions* ed. by Timothy Shopen, 141–170. Cambridge: Cambridge University Press.
- Keydana, Götz. In Press. Indo-European Syntax. *Comparative Indo-European Linguistics: An International Handbook of Language Comparison and the Reconstruction of Indo-European* ed. by Matthias A. Fritz & Jared S. Klein. Berlin: de Gruyter.
- Klein, Jared S. 1992. *On Verbal Accentuation in the Rgveda*. (= American Oriental Society Essay, 11.) New Haven, CT: American Oriental Society.

- Kornfilt, Jaklin. 1987. Turkish and the Turkic Languages. In Bernard Comrie, ed., 619–644.
- Kuhn, Hans. 1933. Zur Wortstellung und -betonung im Altgermanischen. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 57.1–109. DOI: 10.1515/bgsl.1933.1933.57.1
- Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
- Lehmann, Winfred P. 1974. *Proto-Indo-European Syntax*. Austin: University of Texas Press.
- Lehmann, Winfred P. 1978. Conclusion: Toward an Understanding of the Profound Unity Underlying Languages. *Syntactic Typology: Studies in the Phenomenology of Language* ed. by Winfred P. Lehmann, 395–422. Austin: University of Texas Press.
- Leumann, Manu, Johann B. Hofmann & Anton Szantyr. 1977a. *Lateinische Laut- und Formenlehre*. München: Beck.
- Leumann, Manu, Johann B. Hofmann & Anton Szantyr. 1977b. *Lateinische Syntax und Stilistik*. München: Beck
- Lightfoot, David W. 2002a. Myths and the Prehistory of Grammars. *Journal of Linguistics* 38.113–136.
- Lightfoot, David W. 2002b. More Myths. *Journal of Linguistics* 38.619–626.
- Lipták, Anikó. 2009. The Landscape of Correlatives: An Empirical and Analytical Survey. In Anikó Lipták, ed., 1–46. DOI: 10.1075/lfab.1.02lip
- Lipták, Anikó, ed. 2009. *Correlatives Crosslinguistically*. Amsterdam: John Benjamins. DOI: 10.1075/lfab.1
- Mallory, James P. & D.Q. Adams. 1997. *Encyclopedia of Indo-European Culture*. London: Fitzroy-Dearborn.
- Meenakshi, K. 1983. *Epic Syntax*. New Delhi: Meherchand Lachhmandas.
- Meiser, Gerhard. 1998. *Historische Laut- und Formenlehre der lateinischen Sprache*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Oberlies, Thomas. 2003. *A Grammar of Epic Sanskrit*. (= Indian Philology and South Asian Studies, 5.) Berlin: de Gruyter. DOI: 10.1515/9783110899344
- Pintzuk, Susan. 1991. *Phrase Structures in Competition: Variation and Change in Old English Word Order*. Philadelphia: University of Pennsylvania PhD dissertation.
- Pires, Acrisio & Sarah G. Thomason. 2008. How Much Syntactic Reconstruction is Possible? In Gisella Ferraresi & Maria Goldbach, eds., 27–72. DOI: 10.1075/cilt.302.04pir
- Probert, Philomen. 2006. Clause Boundaries in Old Hittite Relative Sentences. *Transactions of the Philological Society* 104.1.17–83. DOI: 10.1111/j.1467-968X.2006.00165.x
- Schleicher, August. 1856. *Handbuch der litauischen Sprache, 1: Grammatik*. Prag: Calve.
- Schwyzler, Eduard. 1949. *Griechische Grammatik 2: Syntax und syntaktische Stilistik*, completed and edited by Albert Debrunner. München: Beck.
- Shibatani, Masayoshi. 1987. Japanese. In Bernard Comrie, ed., 855–880.
- Sihler, Andrew L. 1995. *New Comparative Grammar of Greek and Latin*. Oxford: Oxford University Press.
- Speijer, Jakob S. 1886. *Sanskrit Syntax*. Leiden: Brill.
- Steever, Sanford B. 1987. *The Serial Verb Formation in the Dravidian Languages*. Delhi: Motilal Banarsidass.
- Suzuki, Yasuko. 2008. *Finite Verb Stress and Clitic Verbs in Old English Beowulf*. Urbana-Champaign: University of Illinois PhD dissertation.
- Taylor, Ann. 1990. *Clitics and Configurationality in Ancient Greek*. Philadelphia: University of Pennsylvania PhD dissertation.
- Taylor, Ann. 1994. The Change from SOV to SVO in Ancient Greek. *Language Variation and Change* 6.1–37. DOI: 10.1017/S0954394500001563

- Thomason, Sarah G. 2006. Syntactic Reconstruction. *Encyclopedia of Language and Linguistics*, 2nd ed., Vol. 10 ed. by Keith Brown, 397–401. Oxford: Elsevier.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Wackernagel, Jacob. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen* 1.333–345.
- Watkins, Calvert. 1962. *Indo-European Origins of the Celtic Verb*. Dublin: Institute for Advanced Studies.
- Watkins, Calvert. 1963. Preliminaries to a Historical and Comparative Analysis of the Old Irish Verb. *Celtica* 6.1–49.
- Watkins, Calvert. 1964. Preliminaries to the Reconstruction of Indo-European Sentence Structure. *Proceedings of the 9th International Congress of Linguists* ed. by Horace G. Lunt, 1035–1045. The Hague: Mouton.
- Watkins, Calvert. 1976. Towards Proto-Indo-European Syntax: Problems and Pseudo-Problems. *Papers from the Parasession on Diachronic Syntax* ed. by Sanford Steever, Carol A. Walker & Salikoko S. Mufwene, 305–326. Chicago: Chicago Linguistic Society.
- Weiss, Michael. 2009. *Outline of the Historical and Comparative Grammar of Latin*. Ann Arbor: Beech Stave Press.
- Wichmann, Søren. 2008. The Study of Semantic Alignment: Retrospect and the State of the Art. *The Typology of Semantic Alignment* ed. by Mark Donohue & Søren Wichmann, 3–23. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199238385.003.0001
- Yoshida, Yutaka. 2009. Sogdian. *The Iranian Languages* ed. by Gernot Windfuhr, 279–335. London: Routledge.

Hittite *pai-* ‘go’ and *uwa-* ‘come’ as Restructuring Verbs^{*}

Bernhard Koller

University of California, Los Angeles

The present article provides a new syntactic analysis of the Hittite phraseological construction involving the verbs *pai-* ‘go’ or *uwa-* ‘come’ and a second finite verb. Most approaches have treated the construction as monoclausal in terms of verb serialization (Garrett 1990). This study will take a different approach, arguing that *pai-* and *uwa-* select a phrasal complement. The features that apparently set the construction apart from other cases of embedding in Hittite will be explained as effects of Restructuring (Rizzi 1982), a phenomenon Hittite also exhibits outside of the construction in question. It will further be argued that *uwa-* functions as a raising verb while *pai-* functions as a control verb, accounting for the differences in syntactic behavior between the two.

1. Introduction

The object of the present study is the syntactic construction in Hittite labeled “phraseological construction” (van den Hout 2003), which appears to have a somewhat special status within Hittite grammar. Even though the construction involves two finite verbs, the first being one of the two motion verbs *pai-* ‘go’ and *uwa-* ‘come’, it exhibits certain properties that give it the appearance of a monoclausal structure. The most prominent of these properties is that, while in general second-position clitics in Hittite are typically associated with the clause they appear in, the two verbs in a phraseological construction share a single clitic cluster. This feature (among others to be introduced below) led various scholars (such as Garrett 1990) to treat this construction as a form of verb serialization, i.e. as mono-clausal.

* I would like to thank the participants of the 20th International Symposium on Theoretical and Applied Linguistics, as well as two anonymous reviewers, for their helpful remarks and suggestions. I would also like to thank Dominique Sportiche for his advice and comments on an early version of this paper.

The goal of this article is to offer an alternative analysis of the phraseological construction that renders it unnecessary to attribute to it the unique status within Hittite grammar it would receive under a serialization analysis. Instead, I demonstrate that the construction can receive a fairly straightforward biclausal analysis if one treats Hittite *pai-* and *uwa-* as restructuring verbs. The second verb would then be part of an embedded clause, which shows properties very similar to infinitival clauses. Finally, I account for the different behaviors of *pai-* and *uwa-* in this construction by treating *pai-* as a control verb and *uwa-* as a raising verb. Although Section 4.2 includes a brief discussion of the possible origin of the construction in question (specifically, how the development proposed in van den Hout 2010 can be related to the current proposal), the primary purpose of this article is to provide a **synchronic** account of the internal structure of phraseological sentences in Middle- and Neo-Hittite, where the construction was already fully developed.

Section 2 provides an overview of the main properties of the construction in question, showing what led a number of scholars to analyze it as a form of verb serialization. In Section 3, I give a brief introduction to the phenomenon of Restructuring and demonstrate that certain Hittite verbs exhibit Restructuring properties with infinitival clauses. Section 4 shows that the same phenomenon can be taken to be involved in the phraseological construction, despite the embedded verb being finite in these cases. I begin by showing that the Hittite construction parallels the behavior of Restructuring verbs in Salentino of Brindisi (as analyzed by Terzi 1996). In Section 4.2 I discuss the peculiar position of the subject with phraseological ‘come’ and ‘go’ compared to its unmarked position with verbs selecting infinitival complements. Finally, Section 4.3 discusses a possible structural difference between the two verbs occurring in the construction in question.

2. The Hittite phraseological construction

Hittitologists use the term “phraseological” or “serial verb” construction to refer to sentences containing a finite form of the verb *pai-* ‘go’ or *uwa-* ‘come’ and a second verb in sentence-final position, showing the same inflection in tense, mood, person and number (the agreeing verb forms are underlined in the examples). The two verbs share a single subject, which, when expressed by an accented noun phrase, usually follows the first verb.² This is illustrated in example (1), in which the subject ‘Ishtar, queen of Niniveh’ follows the imperative form *paiḍdu* ‘may she go’.

2. The term ‘accented’ is to be understood as contrasting with subjects that are left unpronounced (‘null-subjects’), as well as subjects expressed by enclitics (such as *=as*), which are always realized as part of the clitic cluster following the first accented word of a sentence (see

- (1) *paiddu=wa=smas=at* ^dIŠTAR-is ^{URU}Ninuwas SAL.LUGAL-as
 go.3SG.IMP=QUOT=for.you=it Ishtar.NOM.SG Nineveh.GEN.SG queen.NOM.SG
hannāu
 decide.3SG.IMP
 ‘may Ishtar, queen of Nineveh, go (and) decide it for you.’ (ABoT 48 7–9,
 example from Disterheft (1984: 222))

Since Hittite is a SOV and (partial) pro-drop language, the word order of (1) gives the impression that we are dealing with two consecutive sentences (‘May she go. May the queen of ...’). However, this analysis cannot be correct because of the distribution of the enclitics. Hittite has a set of second-position clitics which accumulate after the first accented word of a clause. In the phraseological construction, however, there is only a single clitic cluster, and it contains clitics that can only be selected by the second verb. In (1), for example, the two enclitics *=smas* ‘for you’ and *=at* ‘it’ are both arguments of the second verb *hannāu* ‘decide’, not the initial verb, to which they attach. Clitics unambiguously belonging to the second verb are boldfaced in the examples. The remaining clitics (such as the quotation particle *=wa*) take scope over the whole sentence.

Example (1) further illustrates another property that distinguishes phraseological *pai-* and *uwa-* from their regular motion-verb counterparts concerning the distribution of subject clitics: Hittite is a partial pro-drop language in that enclitic subject pronouns are exclusively selected by unaccusative verbs, as was demonstrated by Garrett (1996). The motion verbs *pai-* and *uwa-* regularly select subject clitics when occurring in isolation, while their phraseological counterparts do not.³

- (2) *taknas=at* ^dUTU-as KASKAL-an *paiddu*
 earth.GEN.SG=it sungoddess.GEN.SG way.ACC.SG go.3SG.IPV
 ‘Let it go the way of the Sungoddess of the earth.’ (KUB XVII 10 IV 12–13,
 example from Tjerkstra 1999: 50)
- (3) *pāndu=wa=z* *asandu*
 go.3PL.IMP=QUOT=REFL settle.3PL.IMP
 ‘Now let them proceed to settle down.’ (KBo 3.1+ ii 13–15, example from
 Hoffner & Melchert 2008: 329)

Hoffner & Melchert 2008: 410 for the exact distribution of subject clitics within this chain of clitics).

3. Garrett backs up his generalization by showing that the set of verbs that select subject clitics lines up with the verbs that select the auxiliary *es-* ‘to be’ over *har(k)-* ‘to have’ in the periphrastic perfect (see Hoffner & Melchert 2008: 311 for examples involving *pai-*).

Example (2) shows the subject clitic =*at* ‘it’ being selected by the motion verb *pai-*, while in (3), where *pai-* occurs as part of a phraseological construction, no subject clitic is present. It thus follows that the initial clitic cluster in phraseological constructions will only include a subject clitic if it is selected by the second verb.

- (4) n=**as=mu** uet GİR.MEŠ-as kattan haliyattat
 COMP=SBJ=me come.3SG.PST feet.DAT/LOC.PL down kneel.3SG.PST
 ‘... and she proceeded to prostrate herself at my feet.’ (KUB 14.15 iv 28–29,
 example from Hoffner & Melchert 2008: 327)

That (4) is indeed phraseological follows from the fact that the enclitic =*mu* can only be associated with the second verb, not *uet*. However, since phraseological *uwa-* does not select subject clitics as discussed above, the clitic =*as* must be selected by the second verb *haliyattat*. This is confirmed by Garrett’s analysis (1996: 97) that *haliya-* is among the verbs that can occur with subject clitics. Furthermore, even though subordinate clauses in Hittite are usually introduced by a complementizer, such an item never intervenes between the two verbs of a phraseological construction. Finally, as first noted by van den Hout (2003), *pai-* never occurs with change-of-state verbs, while no such restrictions hold for *uwa-*.

The exact meaning of the construction is still under dispute and some of the proposals that have been made differ radically from each other. To give a selective overview: Disterheft (1984) takes the construction to mark a temporally consecutive event, similar to an adverb; Dunkel (1985) claims that it indicates perfective aspect; Neu (1995) argues for a future tense meaning; van den Hout (2003) concludes that it marks “transition between two actions that are closely and often causally related” (van den Hout 2003: 200). Rieken (2010), in contrast, takes the different possible interpretations of the construction as representing different degrees of grammaticalization of *pai-* and *uwa-*. While some instances of the construction still show the two verbs in their original meaning as motion verbs (possibly with the added semantic element of purpose), others set an event in relation to a reference point in time (immediate in the case of *pai-*, distant in the case of *uwa-*). My discussion will mostly disregard the question of the semantics of the construction and instead focus on its syntactic properties.

The properties described above (and especially the distribution of the clitics) have led various scholars (such as Disterheft 1984, Garrett 1990, Luraghi 1993, van den Hout 2003, Hoffner & Melchert 2008, Rieken 2010) to analyze the phraseological construction as monoclausal in terms of a serial verb construction. Though there is a certain degree of variation in the use of the term, properties that are typically attributed to verb serialization, e.g. by Aikhenvald 2006, are that it is monoclausal, denotes a single event, and involves a single set of morpho-syntactic

features (such as person, tense, etc.). One example of a language using motion verbs in serial verb constructions is Cantonese.

- (5) lei⁵ lo² di¹ saam¹ lai⁴
 you take PL clothing come
 'Bring some clothes' (example from Aikhenvald 2006: 21)

The example in (5) is a typical case of a serial verb construction in which a verb from an open class combines with a motion verb of a closed class to form a complex predicate, in this case with the meaning 'bring'.

While the available evidence is in principle consistent with such an analysis, it would attribute to *pai-* and *uwa-* a rather marked status within Hittite syntax in that they would be the only two verbs permitting serialization.^{4,5} Even though the motion verbs 'come' and 'go' are among the verbs featuring most commonly in serialization, they are typically not the only ones in a serializing language, which is one of the reasons Aikhenvald (2006: 46) rejects such an analysis for double-verb constructions such as English *go eat*.

3. Restructuring and Clitic Climbing

'Restructuring' (Rizzi 1982) has become the term for the well-known phenomenon that certain verbs render their (usually infinitival) embedded clauses transparent for various extraction processes. One of these processes is Clitic Climbing, which moves a clitic associated with an embedded clause into the main clause:

- (6) Piero li venne/andò/tornò a chiamare alla stazione.
 Piero them came/went/came back to call at station
 'Piero came/went/came back to call them at the station.'

4. The more restrictive definition of serial verb constructions proposed in Baker (1989), which requires direct objects selected by the first verb to be shared by the second, would rule out a serial analysis for the Hittite construction. Baker predicts on theoretical grounds that in true serial verb constructions an initial unaccusative verb cannot be followed by an unergative verb, a configuration that is clearly attested in Hittite with *pai-* and *uwa-*. Jaeggli & Hyams (1993) also use the lack of object sharing to reject a serial verb analysis for aspectual *come* and *go* in English. However, Baker's treatment of serial verb constructions is rather controversial. A detailed criticism of his account, based both on theoretical and empirical considerations, is given in Durie (1997).

5. This is true at least from a synchronic viewpoint. Luraghi (1993) takes the quotation particle =*wa* to have originated from the first member of a serial verb construction involving a verb of saying. While this etymology of =*wa* certainly seems plausible, it is not entirely clear from her discussion why the development from a verb selecting a clausal complement to a particle necessarily requires us to posit the intermediary step of a serial verb construction.

In the Italian example above (from Rizzi 1982:5) the clitic *li* ‘them’ is the direct object of the embedded verb *chiamare* ‘call’, even though it precedes the verb in the main clause.

Another type of extraction is so-called Long Scrambling, which moves an argument of the embedded verb into the main clause.

- (7) weil der Hans **den Zaun** begonnen hat zu streichen
 because the Hans the fence begun has to paint
 ‘because Hans has begun to paint the fence’

Similar to the Italian case, the German example (from Wurmbrand 2001: 337) features a direct object, *den Zaun* ‘the fence’, that has been extracted from its clause into a position preceding the verb of the main clause *begonnen* ‘begun’.

Both of the phenomena illustrated above can be observed with infinitival clauses in Hittite:

- (8) m^{ān}=**mu** idalauwanni=ya kuis waggariya[w]anzi
 if=me wickedness.DAT/LOC.SG=and INDEF.NOM.SG rebel.against.INF
 sanhazi
 seek.3SG.PRS
 ‘if someone in wickedness seeks to rebel against me’ (KUB 21.47, KUB 23.82 rev. 23 i 31–32, from Hoffner & Melchert 2008: 334)
- (9) anzel=za=ka[n] ÉRIN.MEŠ-an ÉRIN.MEŠ <LÚ>KÚR walhuwanzi
 our=REFL=LOC troops.ACC.SG troops.enemy.NOM.SG strike.INF
 zikkezzi
 begin.3SG.PRS
 ‘The enemy’s troops will begin to attack our troops.’ (KBo 10.7, iii 15–16, example from Hoffner & Melchert (2008: 332)

The sentence in (8) is an example of Clitic Climbing, similar to the Italian example in (6). The enclitic =*mu* ‘me’ is realized in the main clause despite its origin as the direct object of the embedded verb *waggariyawanzi* ‘to rebel against’. That the enclitic is located in the main clause in (8) is evidenced by the fact that it precedes the nominative subject *kuis* ‘someone’, which unambiguously belongs to the main clause. The sentence in (9) is an example of Long Scrambling, similar to (7). The phrase *anzel* ÉRIN.MEŠ-*an* ‘our troops’ is the direct object of the embedded verb *walhuwanzi* ‘strike’, yet it occurs in the main clause, preceding the main clause subject ÉRIN.MEŠ <LÚ>KÚR ‘enemy troops’.

In what follows I adopt the analysis of Terzi (1996) for Restructuring. Terzi takes the transparency observed with clausal complements of restructuring verbs to be caused by movement and adjunction of the embedded tense head (T) to the tense head of the matrix clause. Furthermore, for T-raising to be possible the

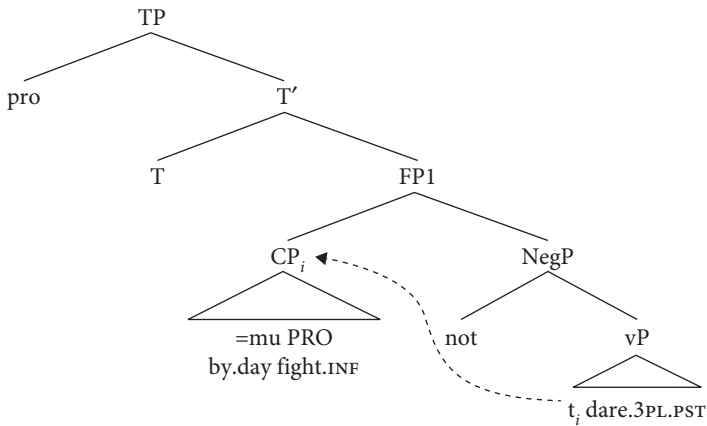
subject positions of the two clauses have to be coindexed, i.e. they have to share a single subject. This restricts restructuring to raising and control contexts.

To illustrate this analysis for Hittite, let us look at another example of clitic climbing out of an infinitival clause.

- (10) nu=**mu**=za namma UD.KAM.HI.A zahhiyauwanzi ÛL []
 COMP=me=REFL any.longer days.LOC.PL fight.INF NEG
 handalliyer
 dared.3PL.PST
 ‘They no longer dared to fight me by day’ (KBo 4.4 iii 62–64, translation from Craig Melchert, p.c.)

The enclitic =*mu* ‘me’ is the direct object of the embedded infinitive *zahhiyauwanzi* ‘fight’. That it must have moved out of the embedded clause is evidenced by the fact that it precedes the adverb *namma*, which clearly belongs to the main clause. What follows is a simplified derivation for (10).

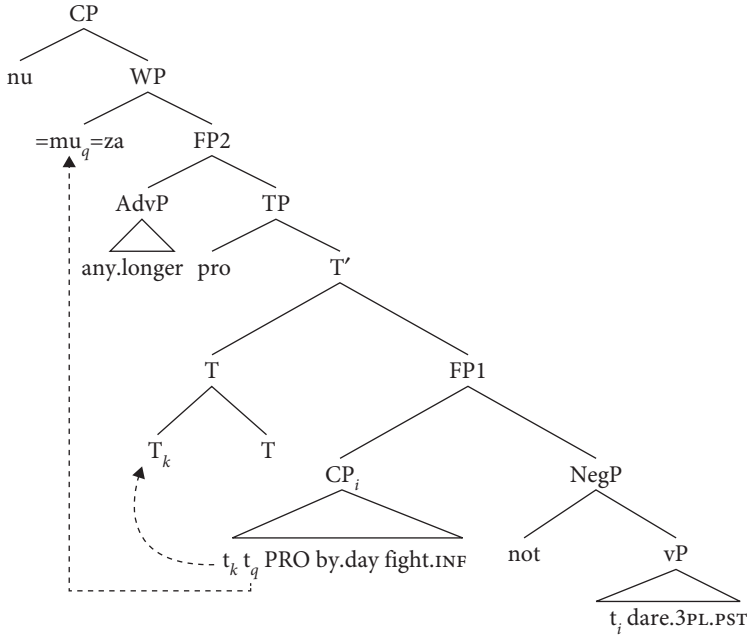
- (11) Movement of embedded clause to FP1



The tree in (11) shows that the embedded clause moves out of the complement position of the verb into a functional projection to the left of the negation (FP1). I take FP1 to be the obligatory landing site for all arguments of the verb in Hittite,

which accounts for the fact that their unmarked position is not adjacent to the verb that selects them, but precedes the negation.^{6,7}

(12) T-raising + clitic climbing



The tree representation in (12) shows how the embedded tense head moves and adjoins to the matrix T, passing through the empty complementizer position (C) of the embedded clause. This process renders the embedded clause transparent and allows the direct object clitic to move to the Wackernagel position of the main clause.

4. The phraseological construction as an instance of Restructuring

We have seen that Hittite exhibits restructuring with infinitival clauses, a cross-linguistically common phenomenon. In what follows I will argue that the

6. A more comprehensive study of Hittite clause structure would be required to identify FP1 with one of the various functional projections above VP proposed in the syntactic literature (such as AgrObj, etc.). Since such an enterprise lies outside the scope of this article, I leave the label for this node intentionally vague.

7. Indefinites are exceptional in this respect, in that they typically occur directly preceding or following the verb.

4.1 Restructuring and finite clauses

(13) Karlu_i lu voli pro_{i/i'}_j kkatta.
Karlu it want.3SG buy.3SG
'Karlu wants to buy it.' (Terzi 1996: 288)

(14) *Lu voggyu ku kkattu.
it want.1SG M buy.1SG
'I want to buy it.' (Terzi 1996:285)

(15) a. Maria e vulutu ku mandza.
 Maria want._{3SG.PST} M eat._{3SG.PRS} (Terzi 1996:287)

b. *Maria e vulutu mandza.
 Maria want._{3SG.PST} eat._{3SG.PRS} (Terzi 1996:287)

In (15) we see that the omission of the modal particle *ku*, which is a prerequisite for clitic climbing, requires two finite verbs to agree with each other in person, number and tense.

Let us now return to the construction with *pai*- and *uwa*- in Hittite:

- (16) *n=at=za=kan paizzi ēpzi*
 COMP=it=REFL=LOC go.3SG.PRS take.3SG.PRS
 ‘... and (she) goes-takes it.’ (KUB 7.53 + 12.58 iv 16–17 (NH), example from van den Hout 2003: 192)

If we adopt Terzi’s analysis of Salentino, apply it to Hittite examples such as (16) and take the clitics to have originated in an embedded clause through clitic climbing, three of the major properties of the phraseological construction, which have been described in Section 2, are immediately accounted for: The restriction to a single subject follows from the requirement that the subject positions of the two clauses have to be coindexed (as in 13). Raising and adjunction of the embedded tense head to the matrix tense head accounts for the agreement between the two verbs (as in 15a), as well as the lack of intervening functional heads such as *nu* (as in 14).⁸

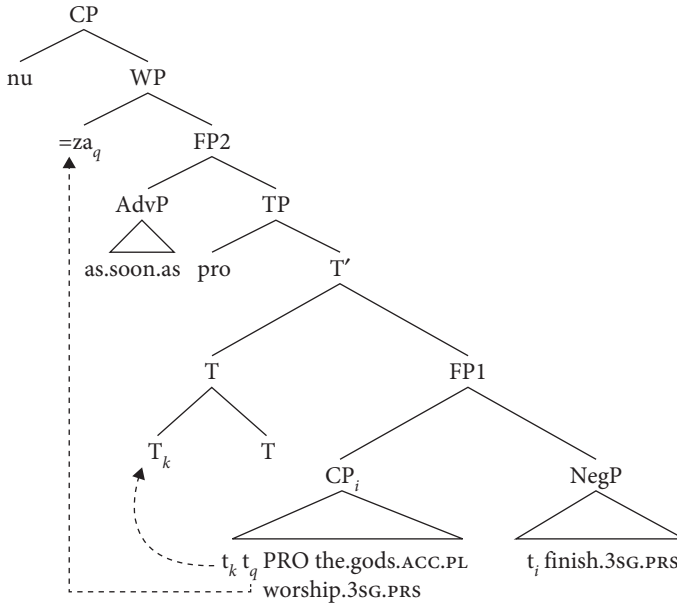
Further evidence that this analysis is on the right track comes from examples such as the following.

- (17) *nu=za mahhan DINGIR.MEŠ iyazzi zinnāi*
 COMP=REFL as.soon.as gods.ACC.PL worship.3SG.PRS finish.3SG.PRS
 ‘As soon as he finishes worshipping the gods ...’ (KBo 18.15, 11–13, example from Hoffner & Melchert 2008: 325)

Here we see that restructuring in Hittite is not restricted to the finite clauses embedded under *pai*- and *uwa*-. Example (17) features a finite embedded clause in the same position as infinitival clauses, accompanied by clitic climbing. The enclitic *=za* unambiguously belongs to the embedded verb since *iya*- has to combine with *=za* to mean ‘worship’ instead of ‘make, do’ (as noted by Boley 1993). The following tree gives a partial derivation for (17).

8. It should be noted here that the Salentino construction differs slightly from Hittite in that the embedded verb is restricted to the present tense in Salentino (Terzi 1996: 287 fn. 15).

(18) Partial derivation for (17)



The structure in (18) illustrates that, if we allow for restructuring with finite clauses, sentences such as (17) can receive exactly the same structure as the infinitival clause in (10). In each case the embedded clause is selected by a control predicate in the main clause. While subject control usually involves non-finite embedded clauses (such as in *I want PRO to leave*), a number of languages do show control with finite embedded clauses. This can be illustrated by the following example from Hebrew.

- (19) Gil₁ hivtiax [še-ec₁ yitna’heg yafe] ...
 Gil promised that-ec (=PRO) behave.3SG.FUT.M well
 ‘Gil promised to behave ...’ (example from Landau 2004: 824)

Other well-known examples of this sort come from various languages of the Balkan Sprachbund, such as Albanian, Bulgarian, Greek and Romanian. Landau posits as a condition for constructions such as (19) in Hebrew that the embedded clause bear dependent tense, i.e. “a situation where the tense of the embedded clause is constrained by (though, crucially, not necessarily identical to) the matrix tense” (Landau 2004: 822). This requirement is borne out in Hittite in a different way than it is in Hebrew and the Balkan languages. While in both of the latter the embedded verb does not co-vary in tense with the matrix verb (e.g. the embedded verb is restricted to the future tense in Hebrew), Hittite does indeed require the tense of the embedded verb to be identical to the tense of the matrix verb. Despite

this difference, Hittite does pattern with the Balkan languages and Hebrew in that the tense of the embedded verb is determined by the matrix verb.

The analysis in (18) further allows us to account for cases such as the following.

- (20) nu uwandu apel URU-an DINGIR^{MES} URU^{URU} HATTI QATAMMA
 COMP come.3PL.IMP his.city.ACC.SG gods.of.Hatti.NOM.PL likewise
 GIR-it isparrandu
 foot.INS.PL stomp.3PL.IMP
 ‘The gods of Hatti shall likewise stomp on his city with their feet (of whoever breaks these oaths).’ (CTH 427, KBo 6.34 III 26–28, trans. based on the German trans. in Neu 1995)

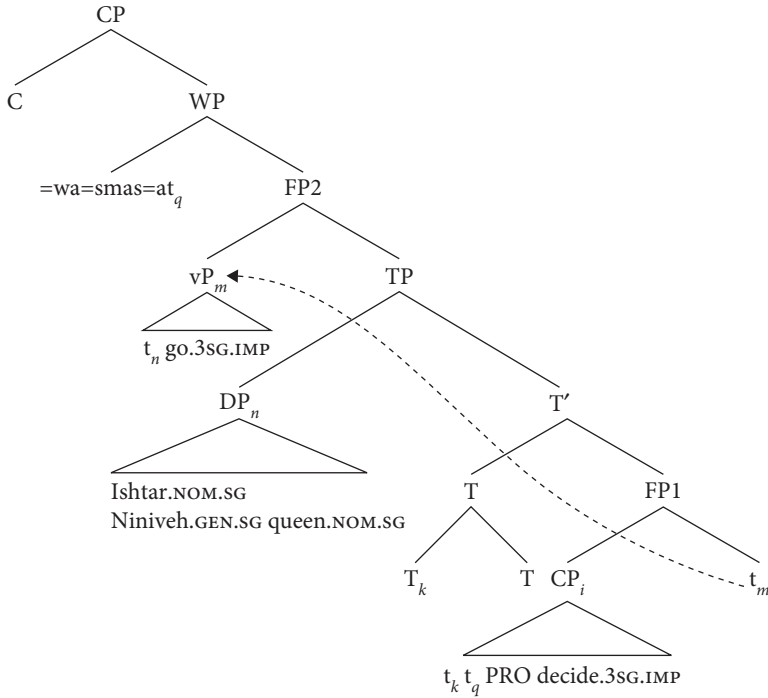
Here the direct object of the second verb appears before the subject. These cases can now be treated as instances of Long Scrambling of the same type as we have already seen with infinitival clauses (see 9 above). The only difference between examples such as (17) and (20) that has yet to be accounted for are the different positions of the matrix verbs.

4.2 The position of the matrix verb

The only property that still sets the phraseological construction apart from infinitival embedded clauses is the fact that the matrix verb consistently precedes the subject.

- (21) paiddu=wa=sma=at ^dIŠTAR-is ^{URU}Ninuwas
 go.3SG.IMP=QUOT=for.you=it Ishtar.NOM.SG Nineveh.GEN.SG
 SAL.LUGAL-as hannāu
 queen.NOM.SG decide.3SG.IMP
 ‘may Ishtar, queen of Nineveh, go (and) decide it for you.’ (ABoT 48 7–9, example from Disterheft 1984:222)

I take the position of the main verb in cases such as (21, repeated from 1) to be due to remnant *v*P-movement.

(22) Remnant *v*P-movement

After all the arguments have been removed, the verb phrase, now only containing the verb itself, moves to a functional projection above the subject (FP2). A functional projection between the complementizer position (C) and the subject position is supported by the process of Fronting, which can target verbs as well as arguments. The fact that whole phrases (as opposed to single words) can occur in this position shows that we are dealing with a phrasal position. The syntactic operation responsible for placing the verb before the subject in the phraseological construction is the same as the one responsible for verb fronting, even though the purpose of the two seems to be rather different. Verb fronting usually has functions pertaining to information structure and typically involves the “topicalizing/contrastive conjunction *-a/-ma*” (Hoffner & Melchert 2008:407). Remnant *v*P movement in the phraseological construction, on the other hand does not seem to involve modification of information structure. However, verb movement with no discernible semantic function is not restricted to phraseological *pai-* and *uwa-*, but can also affect other verbs taking clausal complements:

- (23) nu mahhan MU.KAM-za mehur [tiyazi] selias
 COMP when year.NOM.SG time.NOM.SG start.3SG.PRS seli.DAT/LOC.PL
 sunnumanzi
 fill.INF
 ‘and when the year, (that is,) the season/time for filling of the seli-s arrives.’
 (KUB 21.17 iii 9–10 (NH), from Hoffner & Melchert 2008: 243)
- (24) nu [ŪL tarhuzzi] hāsi
 COMP NEG can.3SG.PRS open.3SG.PRS
 ‘He cannot open (it).’ (KUB 17.10 i 33, from Hoffner & Melchert 2008: 325)

The examples above show further cases in which a sentential complement has been stranded to the right of the verb. In (23) the verb has moved over an infinitival clause while in (24) the whole phrase including the verb and the negation has been moved to the left of the embedded clause *hāsi*. This also shows that this kind of movement has nothing to do with phonological weight, since *hāsi* is significantly lighter than *ŪL (=natta) tarhuzzi*.

One caveat that should be noted about the examples above is that while (24) is ambiguous with respect to the landing site of the dislocated phrase, (23) clearly features movement to a projection below the subject position. On the other hand, phraseological *pai*- and *uwa*- usually seem to target the higher projection FP2 (at least wherever the presence of an overt subject allows us to make a judgment).

A possible diachronic explanation for the position of the subject is provided by van den Hout (2010), who traces the phraseological construction back to asyndetically linked clauses:

- (25) [nu]=za kēdani EZEN-NI [e]kuten ispiyatten
 COMP=REFL this.DAT.LOC feast.DAT.LOC drink.2PL.IMP satiate.2PL.IMP
 [n]inkatten
 satisfy.2PL.IMP
 ‘At this feast eat, drink, satiate (and) satisfy yourselves!’ (KUB 36.97 iv? 4–6 (NS), example from van den Hout 2010)
- (26) t]=as paizzi LŪ.MEŠMEŠEDI-an pēran tiezzi
 COMP=SBJ go.3SG.PRS guards.GEN.PL in.front step.3SG.PRS
 ‘He goes (and) steps in front of the bodyguards.’ (KUB 20.12+ 2–3 (OS), example from van den Hout 2010)

After demonstrating that asyndetically conjoined verbs could share a single Wackernagel clitic (the reflexive particle =*za* in 25), van den Hout argues that the phraseological construction started out in Old Hittite sentences such as (26), where *pai*-/ *uwa*- share a subject clitic with the verb of a second clause. Once the subject clitic had been reanalyzed as belonging exclusively to the second verb, any

clitic selected by the latter could be incorporated into the initial clitic cluster of the whole sentence (such as in 1).

Luraghi (1993) posits a different account that treats the phraseological construction as inherited due to similar constructions in Latin: “Motion verbs were originally (i.e. in Proto-Indo-European) initial in imperative sentences; apparently, however, they could be preceded by pronouns governed by a verb which occurred later in the sentence ...” (Luraghi 1993: 277). Hittite would then have extended this pattern to indicative forms of motion verbs and the development of a rigid SOV pattern (as opposed to a more flexible word-order in PIE) would have led to a reanalysis to a serial verb construction. One problem with this approach is that the verb-initial pattern Luraghi posits for PIE motion verbs (i.e. imperatives preceding pronouns selected by a later verb) would already have to involve some form of verb serialization or restructuring. If this were the case, however, we would predict that clitic climbing with imperatives of *pai-* and *uwa-* should already have been possible in Old Hittite. The fact that no such examples are attested in that stage of the language would have to be treated as a complete accident.

I follow van den Hout’s explanation with the slight modification of taking sentences like (25–26) (before reanalysis) to represent coordinated verb phrases inside a single clause. The possibility of sharing the same clitics results from the two *v*Ps sharing a single set of functional projections:

- (27) $[_{CP} t=as_i [_{TP} [_{vP} t_i paizzi] [_{vP} t_i^{L\acute{U}.ME\check{S}} ME\check{S}EDI-an p\acute{e}ran tiezzi]]]$

The (simplified) structure in (27) shows how the single Wackernagel clitic represents the subject of both coordinated verb phrases. Subsequently *v*P₂ would have been reanalyzed as part of its own clause embedded under the matrix verb *paizzi*, whose position must now be taken as due to movement in accordance with Hittite’s basic SOV pattern. At the same time the initial subject clitic has been reanalyzed as the result of clitic climbing and is no longer selected by the matrix verb.

- (28) $[_{CP} t=as_i [_{TP} [_{vPx} paizzi] [_{CP} [_{TP} [_{vP} t_i^{L\acute{U}.ME\check{S}} ME\check{S}EDI-an p\acute{e}ran tiezzi]]] t_x]]$
(simplified representation not including all traces/functional projections)

The outlined reanalysis could have been further facilitated if restructuring with finite clauses, such as in (17), was already available at the time, although I am not aware of any attested examples in Old Hittite.

Thus, van den Hout (2010) provides a plausible diachronic scenario of how the motion verbs regularly came to precede their clausal complement. However, this by itself cannot explain what would have motivated the *v*P-movement outlined in (22) in the synchronic grammar of a Hittite speaker, who did not have access to any form of diachronic information.

It should also be noted that the reanalysis outlined in (27–28) does not account for the fact that the phraseological verbs almost always precede their subjects, i.e. that remnant ν P fronting consistently targets FP2 instead of FP1. Van den Hout (2010) points out that a number of attested Old Hittite sentences (including examples with *pai-* and *uwa-*) show VSO order. However, it is not entirely clear from van den Hout's discussion why $V_{motion}S$ sequences should have been taken as the sole model for the creation of phraseological construction with overt subjects. Furthermore, as pointed out to me by Craig Melchert (p.c.), the VSO examples of *pai-* and *uwa-* van den Hout bases his argument on are all taken from texts translated from Hattic (as shown by Rizza 2007: 73–83). Thus their status as evidence for Hittite syntax is questionable.

As we have seen in this section in examples (23–24), ν P-fronting over an embedded clause does occur outside the phraseological construction and the existence of a phrasal position preceding the subject (here FP2) can be independently supported by numerous examples of constituent fronting (Hoffner & Melchert 2008). Why the synchronic grammar of Hittite seems to consistently force phraseological *pai-* and *uwa-* to target this projection (as opposed to remaining in situ or targeting FP1), on the other hand, must for the moment remain an open question.

4.3 *uwa-* as a raising verb

Finally, I address the issue of the syntactic difference between *pai-* and *uwa-* in the phraseological construction. So far we have only seen examples with *pai-*, which I have treated as a control construction, i.e. both the embedded verb and the matrix verb assign a theta role to their subject positions. In what follows I argue that the different syntactic behavior of *pai-* and *uwa-* can be accounted for by treating the latter as a raising verb.

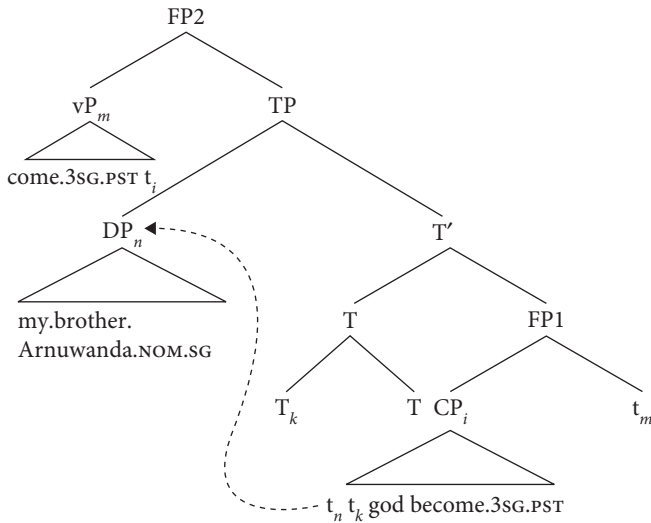
The following examples illustrate several ways in which *uwa-* is more flexible in its usage than *pai-*:

- (29) nu măn uezzi Ê dUTU URUPÚ-na
 COMP if come.3SG.PRS temple.of.the.Sungoddess.of.Arinna.NOM.SG
 parā [ha]ppineszi
 become.richer.3SG.PRS
 'And if the temple of the Sungoddess of Arinna comes-becomes richer,'
 (KUB 26.43(+) obv. 56 (NH), from van den Hout 2003: 185)

- (30) EGIR-pa=ma uet ^mAmmunas DUMU.LUGAL BA.UŠ
 later=CONJ come.3SG.PST prince.Ammunas.NOM.SG die.3SG.PST
 ‘and later it happened that Ammuna, the prince, died.’ (KBo 3.1+ ii 31–32
 (OH/NS), example from van den Hout 2003: 200)
- (31) GI[M-a]n=ma=za uet ŠEŠ-IA [^mArnuwandas
 when=CONJ=REFL come.3SG.PST my.brother.Arnuwanda.NOM.SG
 DINGIR-LIM kisat]
 god.NOM.SG become.3SG.PST
 ‘When it happened that my brother Arnuwanda became a god, (I, My
 Majesty, sat down on my father’s throne).’ (KUB 19.49+ i 19 (NH), example
 from van den Hout 2003: 200)

Example (29) shows that, unlike *pai-*, *uwa-* can occur with inanimate subjects. The three examples further illustrate the observation of van den Hout (2003) that only *uwa-* occurs with change-of-state verbs and in constructions that do not seem to involve any form of voluntary movement on the part of the subject (as reflected in the translation “it happened that ...”). This flexibility in the usage of *uwa-* indicates that it does not assign an external theta role to its subject position, as is typical for raising verbs. What follows is a partial derivation for the sentence in (31). The only difference to the examples with *pai-* is that the subject of the matrix clause originates in the embedded clause.

- (32) Partial derivation of (31)



Treating *uwa-* as a raising verb further allows us to account for the data below, which are apparent counterexamples to the agreement requirement.

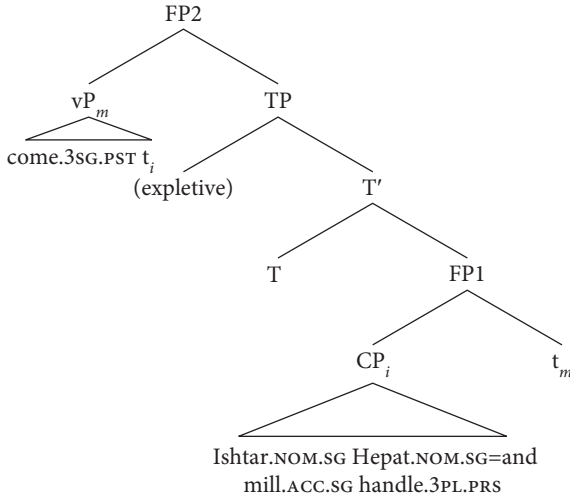
- (33) nu uezzi ANA DINGIR.MEŠ NINDA.GU[(R₄.RA.HI.A
 COMP come.3SG.PRS gods.DAT/LOC breads.ACC.PL
 DUG_iispanduzzi)] UDU_Uauliuss=*a* kars[(antari)]
 wine.rations.ACC.PL sheep.offerings.ACC.PL=and neglect.3PL.PRS
 ‘Thick breads, wine rations, and sheep offerings to the gods they will come-
 neglect.’ (KUB 24.3+ ii 14’–15’, example from van den Hout 2003: 190)
- (34) nu uezzi=ma ^dIŠTAR-is ^dHepatuss=*a* [N^{A4} A]RA₅
 COMP come.3SG.PRS=CONJ Ishtar.NOM.SG Hepat.NOM.SG=and mill.ACC.SG
 [apsi]la mallanzi
 self grind.3.PL.PRS
 ‘... while Istar and Hepat themselves will come-handle the mill.’ (KUB
 33.103 ii 6–8, example from van den Hout 2003: 190)

In (34) we find a coordinated subject (i.e. ^dIŠTAR-is and ^dHepatus combined by the conjunction *a* ‘and’) agreeing with the 3pl verb *mallanzi*, while the matrix verb *uezzi* is in the 3sg. Hittite generally allows both singular and plural agreement with coordinated subjects, which is why van den Hout (2003: 190) interprets this as an example where each of these two options is borne out by one of the two verbs. Under the present restructuring analysis, however, this is impossible, since the two verbs do not separately agree with the subject, but agree with each other via T-raising. Hence, if T-raising takes place the two verbs must surface with the same tense and ϕ -features. I therefore follow Dunkel (1985) in taking (34) as an impersonal construction.

The lack of agreement between the two verbs can be accounted for by the fact that raising verbs typically allow two options for filling their subject position: a) the subject in the main clause can originate in the embedded clause (as in *John seems to t_i be angry*) and b) the subject position of the main clause can be filled by an expletive while the full subject remains in the embedded clause (as in *It seems that John is angry*).

While we saw the first option in (32), the following tree for sentence (34) illustrates the second.

(35) Partial derivation of (34)



The subject [Ishtar and Hepat] remains in the embedded clause, while the subject position is filled by a silent expletive triggering default agreement on the matrix verb.

It should be noted that under the present analysis, examples such (31) are actually ambiguous, and the tree in (32) only represents one of the two possible structures. Since subject raising in (32) would be string-vacuous, it is impossible to tell whether the 3sg form of *uwa-* is due to raising and agreement with the embedded verb, or whether it just exhibits default inflection. On the other hand, cases such as (20), where the two verbs show any inflection other than the 3sg, can only be due to raising.⁹

9. Rieken (2010) cites examples involving *paiddu* (3.SG.IPV of *pai-*) that show lack of agreement between the two verbs.

(i) *paiddu=wa innarauwantes innarauwantes pahhasnuwantes asandu*
 go.3SG.IMP=QUOT strong.ones.NOM.PL strong.NOM.PL protected.NOM.PL be.3PL.IMP
 'Henceforth, let the strong ones be strong and protected.'
 (KBo 17.88+ III 23'f., example from Rieken 2010:233)

If Rieken is correct in analyzing those cases of *paiddu* as uninflected particles instead of verbs (comparing them to particles like *ehu*), sentences such as (i) would just constitute single clauses.

5. Conclusion

It has been shown that a restructuring approach to the phraseological construction can elegantly account for most of its major features, i.e. the position of the enclitics, the restriction to a single subject, and the lack of an overt complementizer introducing the embedded clause. We have also seen that, treating the difference between the behavior of *pai-* and *uwa-* as one between control and raising provides a straightforward account for the aberrant-looking cases that lack agreement between *uwa-* and the second verb in the construction.

By treating the phraseological construction as biclausal, we can rid it of its unique status in the grammar of Hittite and instead analyze it on par with common cases of finite and infinitival embedded clauses. What has to remain an open issue at this point is the fact that *pai-* and *uwa-* consistently precede the subject, which does not seem to be the case with other instances of embedding. Another question I am leaving open for future research is what determines whether a given control verb will select a finite or infinitival clause as its complement, considering that Hittite provides both options.

References

- Aikhenvald, Alexandra Y. 2006. Serial Verb Constructions in Typological Perspective. *Serial Verb Constructions: A Cross-linguistic Typology* ed. by Alexandra Y. Aikhenvald & R.M.W. Dixon, 1–68. Oxford: Oxford University Press.
- Baker, Mark. 1989. Object Sharing in Serial Verb Constructions. *Linguistic Inquiry* 20.513–553.
- Boley, Jacqueline. 1993. *The Hittite Particle -z/-za*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Disterheft, Dorothy. 1984. Non-Final Verbs in Hittite. *Historische Sprachforschung* 97.221–227.
- Dunkel, George E. 1985. IE Hortatory *éy, *éyte: Ved. éta ... stāvāma, Hitt. *ehu=wa it*, Hom. *ei δ' ἄγε*. *Münchener Studien zur Sprachwissenschaft* 46.47–79.
- Durie, Mark. 1997. Grammatical Structures in Verb Serialization. *Complex Predicates* ed. by Alex Alsina, Joan Bresnan & Peter Sells, 289–354. Stanford: CSLI Publications.
- Garrett, Andrew James. 1990. Hittite Enclitic Subjects and Transitive Verbs. *Journal of Cuneiform Studies* 42.227–242. DOI: 10.2307/3515906
- Garrett, Andrew James. 1996. Wackernagel's Law and Unaccusativity in Hittite. *Approaching Second: Second Position Clitics and Related Phenomena* ed. by Aaron L. Halpern & Arnold M. Zwicky, 85–133. Stanford: CSLI.
- Hoffner, Harry A. & H. Craig Melchert. 2008. *A Grammar of the Hittite Language*. Winona Lake, IN: Eisenbrauns.
- Jaeggli, Osvaldo A. & Nina M. Hyams. 1993. On the Independence and Interdependence of Syntactic and Morphological Properties: English Aspectual Come and Go. *Natural Language & Linguistic Theory* 11.313–346. DOI: 10.1007/BF00992916

- Landau, Idan. 2004. The Scale of Finiteness and the Calculus of Control. *Natural Language & Linguistic Theory* 22.811–877. DOI: 10.1007/s11049-004-4265-5
- Luraghi, Silvia. 1993. Verb Serialization and Word Order: Evidence from Hittite. *Historical Linguistics 1989: Papers from the 9th International Conference on Historical Linguistics, New Brunswick, 14–18 August 1989* ed. by Henk Aertsen & Robert J. Jeffers, 267–281. Amsterdam: John Benjamins. DOI: 10.1075/cilt.106.20lur
- Neu, Erich. 1995. Futur im Hethitischen? *Verba et Structurae: Festschrift für Klaus Strunk zum 65. Geburtstag* ed. by Heinrich Hettrich, Wolfgang Hock, Peter-Arnold Mumm & Norbert Oettinger, 195–202. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Rieken, Elisabeth. 2010. Die periphrastischen Konstruktionen mit *pai-* "gehen" und *uwa-* "kommen" im Hethitischen. *Investigationes Anatolicae: Gedenkschrift für Erich Neu* (=Studien zu den Bog̃ azköy-Texten, 52) ed. by Jörg Klinger, Elisabeth Rieken & Christel Rüster, 217–239. Wiesbaden: Harrassowitz.
- Rizza, Alfredo. 2007. *I Pronomi enclitici nei testi etei di traduzione dal hattico*. (= *Studia Mediterranea*, 20.) Pavia: Iuculano.
- Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris. DOI: 10.1515/9783110883718
- Terzi, Arhonto. 1996. Clitic Climbing from Finite Clauses and Tense Raising. *Probus* 8.273–295. DOI: 10.1515/prbs.1996.8.3.273
- Tjerkstra, Françoise A. 1999. *Principles of the Relation between Local Adverb, Verb and Sentence Particle in Hittite*. Groningen: STYX Publications.
- van den Hout, Theo. 2003. Studies in the Hittite Phraseological Construction I: Its Syntactic and Semantic Properties. *Hittite Studies in Honor of Harry A. Hoffner, Jr. on the Occasion of his 65th Birthday* ed. by Gary Beckman, Richard Beal & Gregory McMahon, 177–204. Winona Lake: Eisenbrauns.
- van den Hout, Theo. 2010. Studies in the Hittite Phraseological Construction II: Its Origin. *Studia Anatolica in memoriam Erich Neu dicata* (=Hethitica, XVI) ed. by René Lebrun & Julien De Vos, 191–204. Louvain-La-Nouve: Peeters.
- Wurmbrand, Susanne. 2001. *Infinitives: Restructuring and Clause Structure*. Berlin: Mouton De Gruyter.

Reconstructing passive and voice in Proto-Indo-European*

Leonid Kulikov and Nikolaos Lavidas
Ghent University & Aristotle University of Thessaloniki

This article examines various aspects of the reconstruction of the passive in Proto-Indo-European (PIE), foremost on the basis of evidence from the Indo-Aryan (Early Vedic) and Greek branches. In Proto-Indo-European the fundamental distinction within the verbal system is between the active and middle, while specialized markers of the passive are lacking and the passive syntactic pattern is encoded with middle inflection. Apart from the suffix **-i̯(e/o)-* (for which we cannot reconstruct a passive function in the proto-language) and several nominal derivatives, we do not find sufficient evidence for specialized passive morphology. The role of the middle (and stative) in the expression of the passive in ancient IE languages raises important theoretical questions and is a testing ground for the methods of syntactic reconstruction. We will examine the contrast between non-specialized and specialized markers of the passive in Early Vedic and Greek. Most Indo-European languages have abandoned the use of middle forms in passive patterns, while Greek is quite conservative and regularly uses middle forms as passives. In contrast, Indo-Aryan has chosen a different, anti-syncretic, strategy of encoding detransitivizing derivational morphology, though with the middle inflection consistently preserved in passive *ya*-presents. These two branches, Indo-Aryan and Greek, arguably instantiate two basic types of development: a syncretic type found in many Western branches, including Greek, and an anti-syncretic type attested in some Eastern branches, in particular in Indo-Aryan.

* We would like to thank Jóhanna Barðdal and Silvia Luraghi as well as two anonymous reviewers of the *Journal of Historical Linguistics* for valuable comments and criticisms on earlier drafts of this article. We are also thankful to the audience of the Workshop “Proto-Indo-European Syntax and its Development” (Aristotle University of Thessaloniki, April 1–2, 2011) for important remarks and suggestions — in particular, to Dorothy Disterheft, Thórhallur Eythórsson, Chiara Gianollo, Hans Henrich Hock, Thomas Krisch, and Ianthi Maria Tsimpli.

1. Introduction

The aim of this article is to identify the problems of reconstructing the category of the passive in Proto-Indo-European (PIE), with a focus on evidence from Early Vedic and Greek. We will examine the non-specialized markers of the passive in Early Vedic and Greek in order to show their role in the development of the encoding of the passive pattern. More specifically, we will show that the change from non-specialized to specialized passive markers indicates a morphological parallelism with regard to the evolution of markers of valency-changing categories such as causative.

For the purposes of the present survey, we will primarily follow the approach developed by the Leningrad/St. Petersburg Typology Group, which offers a useful calculus of relationships between the two main levels of representation of the structure of a clause.¹ These levels include: (i) the level of semantic ‘macroroles’ (for example, Actor or Undergoer; see, among many others, Dowty 1990) and (ii) the level of syntactic functions, such as Subject [S], Direct Object [DO], Indirect Object [IO], Oblique Object [Obl], which are often directly corresponding to case-marking (nominative, accusative, dative).²

Section 2.1 presents the main characteristics of the passive (syntactic) pattern and passivization following the Leningrad/St. Petersburg Typology Group and focusing on Vedic and Greek data. Section 2.2 discusses the marking of the passive in different IE languages and shows the role of the mediopassive conjugation in ancient IE languages as well as the lack of evidence for specialized markers for the passive pattern. Section 3 supports the claim of the absence of specialized markers for the passive pattern using evidence from Greek. In Section 4, we analyze the Vedic passive, which again is expressed by both non-specialized and specialized markers. We argue that middle morphology in Vedic is not regularly used for the passive syntactic pattern, as well as that middle morphology is not the only morphology found for the passive syntactic pattern. Section 5 concludes the main remarks of the article regarding the problems in reconstructing the passive and presents a sketch of a hypothesis about the PIE sources of the passive and the syncretic vs. anti-syncretic types in the development of the PIE verb.

1. For an outline of this methodology, see, for example, Geniušienė (1987), Mel’čuk (1993), and Kulikov (2011a). For a typological overview of passive, see the seminal work by Siewierska (1984).

2. This correspondence is not always of straightforward character, however. Thus, in the case of ‘non-canonical’ case-marking of the core relations, subject and object, additional and more sophisticated syntactic criteria are necessary for elucidating the grammatical relation of the noun phrases. We would like to thank Jóhanna Barðdal (p.c.) for drawing our attention to this issue.

2. Background

2.1 Passive and Passivization

The most important theoretical concept that can be determined regarding the two levels of representation (semantic and syntactic) and that enables us to capture the rich variety of voices is that of SYNTACTIC PATTERN, which is a pattern of mapping semantic roles onto syntactic functions, cf. the concept of ‘diathesis’ in the terminology of the Leningrad/St. Petersburg Typology Group (see e.g. Geniušienė 1987) and the notion of ‘valence pattern’ (Haspelmath & Müller-Bardey 2004). The pattern in which the Actor is mapped onto the Subject and the Undergoer is mapped onto the Direct Object is the most common, unmarked way of representing an event and therefore can be regarded as the basic, or the neutral, syntactic pattern of a simple transitive verb (for the notion of basic pattern, see, in particular, Shibatani 2006: 257ff.). This pattern is illustrated by the following Ancient Greek and Sanskrit (Vedic) sentences:

- (1) Homeric Greek (Hom. Od. 3.379)
 hế toi kai patér’ esthlòn en
 REL.NOM.SG.F you:DAT.SG and father:ACC.SG noble:ACC.SG.M among
 Argeioisin etíma
 Argive:DAT.PL honor:IMPF.ACT.3SG
 ‘who honored your noble father too among the Argives’
- (2) Vedic (RV 9.109.11)
 tám te sotáro rásam mádāya
 that:ACC.SG.M your presser:NOM.PL sap:ACC.SG intoxication:DAT.SG
 punánti sómam mahé dyumnáya ...
 purify:PRS:3PL.ACT Soma:ACC.SG great:DAT.SG.M/N intoxication:DAT.SG
 ‘The pressers purify that sap of yours (= Soma) for great intoxication, they
 purify Soma for great splendor ...’

Syntactic patterns can be conveniently presented in tabular form. Thus, the syntactic pattern exemplified by examples (1–2) can be schematized as in (3):

- (3) Basic transitive pattern

Semantic argument level (role)	X (Actor)	Y (Undergoer)
Syntactic function level (case)	S (NOM) (hế; sotáro)	DO (ACC) (patér’(a); rásam)

Syntactic changes affecting transitivity can be described in terms of modifications of syntactic patterns (which may be accompanied by certain changes in verbal

forms, as in the case of the category of voice; see below). Thus, the modification of the basic (neutral) transitive pattern that results in the passive equivalent of a transitive clause typically suggests the following two syntactic phenomena: (i) the promotion of the initial Direct Object to the Subject (= the Subject of the passive pattern), and (ii) the demotion of the initial Subject (usually an Actor). The demotion of the Subject may amount either to its being realized as an Oblique Object (passive Actor) or to its removal from the clause. These phenomena (i–ii) are partly independent from each other, since the demotion of the initial Subject is not necessarily accompanied by the promotion of the initial Direct Object, as in the case of the impersonal (backgrounding) passive; see, for instance, Kulikov (2011a: 376 ff., and references therein). This change in pattern is exemplified by the passive examples in (4–5) and presented in tabular form in (6):

- (4) Homeric Greek (Hom. Od. 7.69)

hòs keínē perì kēri tetímētai ék te
so that:NOM.SG.F above heart:DAT.SG honor:PRF.MP.3SG by PTC
phílōn paídōn
beloved:GEN.PL.M child:GEN.PL.M
'So she has been honored exceedingly by her beloved children'

- (5) Vedic (RV 9.86.12)

s_uv-āyudhāḥ sotṛbhiḥ pū-ya-te vṛṣā
well-armed:NOM.SG.M presser:INS.PL purify-PRS.PASS-3SG.MID bull:NOM.SG
'[Soma], the well-armed bull, is being purified by pressers'

- (6) Passive pattern

X	Y	$\Rightarrow$	Y	X
S	DO		S	Obl/–

In languages in which the correspondence between syntactic functions and case-marking is relatively straightforward, pattern modification can also be formulated in terms of changes in case-marking. Thus, scheme (7) describes passivization in Ancient Greek and Sanskrit:

- (7) Passivization in Ancient Greek and Sanskrit

X	Y	$\Rightarrow$	Y	X
S (NOM)	DO (ACC)		S (NOM)	Ancient Greek: Obl (DAT, <i>hupó/ek</i> ... + GEN)/– Sanskrit: Obl (INS)/–

The technical concept of SYNTACTIC PATTERN was primarily introduced to reach a more accurate definition and a reasonable formalization of the traditional concept of VOICE, and, eventually, to determine its linguistic content. The category of

VOICE is determined based on the concept of syntactic pattern as follows: VOICE IS A REGULAR ENCODING OF SYNTACTIC PATTERNS THROUGH VERBAL MORPHOLOGY. Thus, many languages encode the passive pattern outlined above using a VERBAL MORPHEME, which, accordingly, is interpreted as the marker of the passive. In Sanskrit, these are the present passive suffix *-yá-* and the middle inflection (for details, see below); in Latin (in the present, imperfect and future tenses), the passive voice is expressed by special series of endings ('mediopassive' inflection); cf. 3SG.PASS *-tur* ~ 3SG.ACT *-t*. Furthermore, in Latin (in the tenses derived from the perfect stem — that is, in the perfect, pluperfect and future perfect) and in many modern Romance and Germanic languages, the passive is expressed using the auxiliary verb 'be' (or 'become') (Latin *esse*, Italian *essere*, French *être*, English *be*, German *werden* etc.) and the past (perfect) participle. Of course, defining the passive and other voices primarily in terms of syntactic patterns puts the focus on one particular aspect of voice, capturing such features as detransitivization, but leaving in shadow other aspects and functions of voice and passive, such as, in particular, change in topicality (see, above all, Givón 1981, Barðdal & Molnár 2003).

In many languages, the morphemes labeled 'passive' include the functions of non-canonical passive(s) or even non-passive patterns, such as the reflexive or the anticausative, which justifies a more morphologically oriented (i.e., form-oriented) study of voice.

2.2 The passive in ancient Indo-European languages

Scholars agree on the content of the category of the middle (mediopassive) conjugation in the ancient IE languages (and thus, presumably, in the IE proto-language), which is represented in Figure 1 (see, e.g., Neu 1968a, 1968b, Jankuhn 1969, Flobert 1975, Gonda 1979, Allan 2003, and, most recently and with particular focus on Greek and Indo-Iranian, Benedetti 2006).

Passive
Anticausative
Reflexive
Reciprocal
Auto-benefactive (subject version)

Figure 1. Mediopassive conjugation in ancient IE languages

The evidence for the passive use of middle (mediopassive) forms in the (ancient) IE languages is ample and does not require much discussion (see, for example,

Leumann & Hoffman 1928 for Latin, Friedrich 1967, Neu 1968a, 1968b, and Hoffner & Melchert 2008 for Hittite, and Krause 1968 for Gothic).³

More limited is the evidence for the existence of specialized passive morphemes. Thus, although there is no doubt that the present formative suffix **-j(e/o)-* can be safely reconstructed for the proto-language (see, e.g., Szemerényi 1990: 295ff.), evidence for its passive function is limited to three Eastern branches:⁴

- Indo-Aryan (Vedic): suffix *-yá-* + middle inflection
yuj ‘yoke, join’: *yunak-ti* ‘yokes, joins’ ~ *yujyáte* ‘is (being) yoked, joined’
bhr̥ ‘bear, carry’: *bhára-ti* ‘carries’ ~ *bhriyáte* ‘is carried’
- Iranian (Avestan): suffix *-iia-*
bar ‘carry’ ~ *bairiiate* ‘is carried’
kar ‘make’ ~ *kiriieiti* ‘is made’, *kiriieinti* ‘are made’ (note active inflection)
- Armenian: present formant *-i-*
berem ‘I carry’ ~ 1SG *berim* ‘I am carried’, 3SG *beri* ‘is carried’

The IE proto-language must also have had some nominal formations that were regularly used to mark passive patterns. These include, above all, the perfect (resultative) participles with **-to-*, **-no-*, reflected, for instance, in the Indo-Aryan (Vedic) *-ta-/na-* participles (cf. *yuj* ‘yoke, join’: *yuk-tá-* ‘yoked, joined’; *bhr̥* ‘bear, carry’: *bhr̥-tá-* ‘carried’), in the Greek *-to-s* participles (cf. *didáskō* ‘teach’: *didak-tós* ‘taught, learned’), in the Latin perfect passive participles in *-tus* and in the Slavic *-t-/n-* participles (cf. Old Church Slavonic *po-vi-ti* ‘envelop’: *po-vi-tō* ‘(is) enveloped’ or *nes-ti* ‘carry’: *nes-enō* ‘(is) carried’). Periphrastic passives with participles are also common in Hittite (see Hoffner & Melchert 2008: 304–305). However, this formation cannot be qualified as a specialized passive in the strict sense of the word, since the resultative participles with the suffixes **-to-* and **-no-* can also be derived from intransitive verbs (cf. Vedic subject resultatives *úpa-sanna-* ‘(having) set, sat near/upon’ or *gatá-* ‘gone’ derived from *úpa-sad* ‘sit near/upon’ and *gam* ‘go’).

3. Lack of specialized passive morphology in IE: Evidence from Greek

The supposed lack of a specialized passive marker in PIE finds its direct continuation, in particular, in Ancient Greek. We will show that there is no specialized

3. However, Hoffner & Melchert (2008) show that the middle was not the standard way of expressing passive in Hittite, and thus its usage in passive constructions cannot be compared to that attested in Latin or Greek.

4. On the syntactic reconstruction in the framework of the Construction Grammar, see Barðdal & Eythórsson (2012), Barðdal (2013).

passive marker in Ancient Greek, contrary to the traditional assumption (among others, Allan 2003, Gildersleeve 1900, Goodwin 1978 [1894], Humbert 1945, Kühner & Gerth 1963 [1898–1904], Meillet & Vendryes 1924, Schwyzler & Debrunner 1950, Smyth 1984 [1920], Stahl 1907) that Ancient Greek has **three voice paradigms**: ACTIVE, MIDDLE, and PASSIVE. The middle and passive paradigms are distinguished only in the future and the aorist; hence, the contrast is between active and mediopassive in all tenses except for future and aorist.

First, in Classical Greek, middle and (traditionally named) “passive” suffixes, in the future and the aorist where they are distinguished, are used interchangeably (“in free variation”) without any difference in meaning. In other words, “passive” morphology is not identified with the passive pattern; rather, it can have a reflexive or anticausative interpretation. Furthermore, the middle ending is also not identified with a specific syntactic pattern (reflexive or anticausative); instead, it can also mark the passive pattern.⁵ In example (8), a “passive” form is used in an intransitive non-passive construction (with a psych-verb), while in example (9), a middle form is used in a passive pattern.

- (8) Classical Greek (Isoc. Or. 12.20.1; 5–4 BC)
 hōs mèn oûn elupéthēn kaì sunetarákthēn
 how PTC PTC grieve:AOR.PASS.1SG and disturb:AOR.PASS.1SG
 akoúsas ...
 hear:PTCP.AOR.ACT.NOM.SG.M
 ‘How much I grieved and how much I was disturbed on hearing ...’
- (9) Classical Greek (Xen. Hell. 7.5.18; 5–4 BC)
 ekeínoi poliorkésointo hupò tôn antipálōn
 that:NOM.PL.M besiege:OPT.FUT.MID.3PL by ART.GEN.PL adversary:GEN.PL
 ‘They would be besieged by their adversaries.’

Second, regarding the morphological analysis of these types, in the future, both the “passive” morpheme *-(th)ē-* and the mediopassive *-omai* express the mediopassive voice (and mark passive or non-passive patterns as we have observed above); that is, we have a **double marking** of voice in these cases. In contrast, the aorist has only the “passive” morpheme *-(th)ē-*. The suffix *-ē-* in the aorist has been considered (by Humbert 1945, Jannaris 1968 [1897], Kühner & Gerth 1963 [1898–1904], Schwyzler & Debrunner 1950, Smyth 1984 [1920], among others) to be a suffix that marked the aorist with active verbs of the conjugation *-mi* and was later used in reflexive and anticausative (but not passive) patterns with verbs

5. We leave the issue of quantitative analysis of the data open for future research. What is important for us here is that middle forms in a passive pattern and passive forms (passive aorists and futures) in anticausative or reflexive patterns are possible.

with active morphology (see also Allan 2003, Luraghi 2010a). The suffix *-ē-* was primarily used in reflexive and anticausative (but not passive) patterns in Homer; Delbrück (1897) states that only two (out of the 22 examples) forms in *-ē-* have a purely passive interpretation (*eplégēn* ‘I was struck, smitten’, *etúpēn* ‘I was beaten’), and Grosse (1889) claims that, of the 129 examples with *-thē-* that he notes, 30 forms have a purely passive interpretation (for example, *ktathēnai* acquire:AOR. PASS.INF, *dotheíē* give:OPT.AOR.PASS.3SG). Similarly, Allan (2003) notes that of the 66 examples of passive aorists in *-thē-* in Homer, 19 examples have a passive interpretation, whereas of the 34 examples of passive aorists in *-ē-*, nine examples have a passive interpretation.

Three aspects of the distribution of active and mediopassive voice reveal the role of “passive” voice (thus called in accordance with tradition) and its non-specialized status in Ancient Greek. First, mediopassive morphology is frequently used with transitive verbs and denotes that the result of the verbal action concerns the subject (i.e., an auto-benefactive meaning; we refer here only to voice alternating verbs — the mediopassive is also used with deponents, but deponents cannot be in the active voice), see (10). Verbs with a reflexive (and active) reading and mediopassive morphology can also take a direct object in the accusative case in Ancient Greek, see (11), if the object is directly related to the subject (the construction is ungrammatical after the Hellenistic Greek period, when mediopassive verbs with a reflexive interpretation are obligatorily intransitive).

- (10) Classical Greek (Xen. Mem. 4.4.19; 5–4 BC)
 hóti hoi ánthrōpoi autoùs éthento?
 that ART.NOM.PL man:NOM.PL 3PL.ACC.M set:AOR.MID.3PL
 ‘[Could you say] that men made them [=those laws] for themselves?’
- (11) Classical Greek (Hdt. 4.75.7; 5 BC)
 ou gàr dè louóntai húdati tò parápan
 NEG PTC PTC wash:PRS.MP.3PL water:DAT.SG ART.ACC.SG absolutely
 tò sōma
 ART.ACC.SG body:ACC.SG
 ‘For they never ever wash their bodies with water.’

Second, in Homeric Greek, constructions with active verbs, an Actor argument in a PP or in the dative case, and an Undergoer argument as subject are possible. The role of the active voice in these examples was first discussed in detail by Jankuhn (1969), although Jankuhn did not try to account for them. Similar syntactic patterns (examples with intransitive verbs with active morphology and a PP or a dative denoting an Actor, with a passive interpretation, or a Cause, with a passive or anticausative interpretation) are also attested in Classical Greek (George

2005). Hence, in Classical Greek, sentences with verbs with active morphology, the Undergoer argument in the subject position, and an Actor in the dative case or in a PP are possible.

- (12) Classical Greek (Xen. Cyr. 7.1.48; 5–4 BC)

oud' autoí ge apéthnēiskon hupò hippéōn
 NEG 3PL.NOM.M PTC die:IMPF.ACT.3PL by cavalry:GEN.PL
 'They were not killed by any of the [enemy's] cavalry.'

It should be noted that Luraghi (2010a) has argued that the Ancient Greek data only show increasing grammaticalization of the passive (rather than the beginning of its introduction). According to Luraghi, the creation of the new value of passive for the voice category must be traced back to some late stage of PIE in which the middle voice sporadically began to be interpreted as a passive, depending on the context. Following Luraghi, we have to assume that the active/passive opposition developed in Ancient Greek when the mediopassive inflection became increasingly obligatory as a marker of transitivity alternations for every verb that can take an object.

Third, in Classical Greek, many examples of mediopassive verbs in passive patterns can also be found. The Actor in these structures is only present in exceptional cases (Ernout 1908–1909: 329 ff., Schwyzler 1943) despite its presence in intransitive patterns which are marked with active morphology, see (12) above. The necessity or not for an Agent (in a PP or NP in the dative) for the analysis of a construction as passive in Classical Greek has been the subject of discussion in the literature (George 2005, Jankuhn 1969, Kuryłowicz 1935, 1964, Meillet 1964 [1903], Schmidt 1963a, 1963b, Wackernagel 1904). In the examples where a PP appears, the choice of the PP depends on the following (George 2005): (a) the semantic class to which the verb belongs — verbs meaning 'give' and 'send' are accompanied by *ek, pará* + Genitive, while verbs meaning 'think' are accompanied by *prós, pará* + Dative; and (b) on the finiteness of the verb — participles frequently take different PPs from those attested with finite verbs.

Furthermore, in Classical Greek, the presence of the following arguments in the position of the subject of mediopassive verbs (and the assignment of nominative case to them) is possible: (a) an Actor + Benefactive; the person who acts and benefits from the action or the interested person (13a), (b) an Undergoer (Theme) or Recipient (13b), (c) a Recipient that is in the dative or accusative in the corresponding transitive pattern; in this case, the Undergoer (Theme) remains in the accusative because the mediopassive does not block the assignment of the accusative to the Undergoer (14).

(13) Classical Greek

a. (Soph. OT 1143; 5 BC)

hōs emautôi thrēmma threpsaímēn egó
 so myself:DAT.SG foster-child:ACC.SG rear:OPT.AOR.MID.1SG 1SG.NOM
 ‘So that I would rear (him) as my own foster child.’

b. (Xen. Hell. 2.3.15; 5–4 BC)

eí tis etimáto hupò toû
 because INDF.NOM.SG.M/F honor:IMPF.MP.3SG by ART.GEN.SG
 démou
 commons:GEN.SG
 ‘because somebody was honored by the commons.’

(14) Classical Greek (Thuc. 1.140; 5 BC)

(humeís) állo ti meízon euthùs
 2PL.NOM other:ACC.SG.N INDF.ACC.SG.N bigger:ACC.SG.N immediately
 epitakhthésesthe
 impose:FUT.PASS.2PL
 ‘You will be immediately imposed some greater demand.’

As can be seen from comparison of example (13b) with example (14), both canonical passives (where the accusative object of the active sentence is promoted to the passive subject) and passives with the subject of a Recipient (corresponding to the dative or the accusative of the active sentence) are possible. In the latter case (and if the verb is ditransitive), the Undergoer (Theme) remains in the accusative case in spite of the mediopassive form of the verb. This shows again that the mediopassive (and the “passive”) morphology does not encode a detransitivizing process in Ancient Greek (since a “direct object” in the accusative case can appear in spite of the passive pattern and the (medio)passive morphology).

4. Evidence from Vedic: Traces of IE passive?

In Old Indo-Aryan (Vedic), the relationship between the passive pattern and specialized forms is not straightforward either. Within the three main tense systems of the Vedic verb — present, aorist, and perfect — the passive is expressed by specialized forms (some of which take middle endings, however) and, only exceptionally, by non-specialized (bare) middle forms. Specialized passive formations include the following: (i) in the **present**, passives with the accented suffix *-yá-* and middle inflection (e.g., *yuj* ‘yoke, join’, 3sg. *yujyáte* ‘is (being) yoked, joined’, 3pl. *yujyánte* ‘are (being) yoked, joined’, participle *yujyámāna-*, etc.); (ii) in the **aorist**, mediopassive aorists with the endings *-i* and *-ran* (*-ram*) (3sg. with *-i* and 3pl. with

-*ran/-ram*, e.g., *yuj* ‘yoke, join’, 3sg. *áyoji*, 3pl. *ayujran*); and (iii) statives with the endings *-e* and *-re*, which form passives, lacking in the **perfect** (e.g., *hi* ‘impel’, 3sg. *hinvé* ‘(it) is/has been impelled’, 3pl. *hinviré* ‘(they) are/have been impelled’; for details, see Kümmel 1996 and Gotō 1997).⁶ Both mediopassive *i*-aorists and statives have defective paradigms.

The system of passive formation in early Vedic (that is, above all, in the language of the *R̥gveda*), includes, according to the *communis opinio*, in addition to specialized passive formations (*-yá*-presents, *i*-aorists, and statives), a number of non-specialized middle forms in all three tenses that purportedly encode the passive pattern. These non-specialized (‘bare’) middle forms (that is, present formations with middle inflection but without the suffix *-yá*-, middle aorists different from *i/-ran*-aorists, and middle perfects) are depicted by the narrow shadowed column in Table 1; the table shows that, together with specialized passive formations (*-yá*-presents, *i*-aorists, and statives), they are used in passive constructions.

Table 1. Passive in Vedic, traditional view

	Non-Passive		Passive
	Active	Middle	
Present			<i>-yá</i> -presents
Aorist			aorists in <i>-i/-ran</i>
Perfect			statives in <i>-e/-re</i>

As argued in Kulikov (2006a, 2006b, 2009), briefly recapitulated below, non-specialized middle forms are rarely used in passive constructions. The two large groups of non-specialized middle forms (which can be called ‘bare middles’, that is, middle forms without specialized passive morphology) employed in passive patterns, middle perfects and middle athematic participles with the suffix *-āna-*, have the special paradigmatic status of being grammatically ambiguous — that is, such forms are members of two paradigms (for instance, *hinvāná-* belongs to the paradigm of the middle present and to that of the stative, and *all* forms attested in passive constructions should be qualified as belonging to the latter; see below for details). Accordingly, such forms (for instance, *hinvāná-*) should be discarded as evidence for the passive usage of the non-specialized middle forms (in the case of *hinvāná-*, middle present formations without the passive suffix *-yá-*) in the passive pattern.

6. Synchronically, the middle perfect endings *-e* and *-re* (3sg. and 3pl.) are identical with the stative inflection; historically, they represent a secondary innovation, being borrowed from the stative paradigm, see Kümmel 2000: 52 (“‘Stativendungen’ des Präsens”). Thus, from a historical point of view, *-e* and *-re* are stative, not (middle) perfect, endings.

Athematic middle participles with the suffix *-āna-* exhibit unusual syntactic properties in early Vedic, particularly in the language of the Ṛgveda (RV). Whereas the corresponding finite forms are used only transitively, the *-āna-*participles are attested both in transitive and intransitive (passive) patterns (see already Delbrück 1888:264). For instance, the participle *hinvāná-* (root *hi* ‘impel’), taken by all grammars as the middle participle of the nasal present with the suffix *-nó/-nu-* (class V in the Indian tradition), occurs 18 times in intransitive (passive) patterns, as in 15a, and ten times in transitive patterns, as in 15b, in the Ṛgveda:

(15) Vedic

a. (RV 9.12.8)

sómo **hi-nv-ānó** arṣati
Soma:NOM.SG impel-PRS-PTCP.MID:NOM.SG.M flow:PRS:3SG.ACT
‘Soma, being impelled, flows.’

b. (RV 9.97.32)

índrāya pavase ... **hi-nv-ānó**
Indra:DAT.SG purify:PRS:2SG.MID impel-PRS-PTCP.MID:NOM.SG.M
vácam matibhiḥ kavínām
speech:ACC.SG thought:INS.PL poet:GEN.PL
‘You (sc. Soma) purify yourself for Indra ... impelling (your) speech
with the (religious) thoughts of the poets.’

In contrast, the finite middle forms produced from the same stem (3pl.mid. *hinváte* etc.), with which *hinvāná-* is supposed to belong, can only be used transitively, meaning ‘to impel sth.’, as in (16):

(16) Vedic (RV 9.65.11)

hi-nv-é vāṣeṣu vājínam
impel-PRS-1SG.MID prize:LOC.PL runner:ACC.SG
‘I spur on this runner [in the race] for prizes.’

Similarly, the participle *yujāná-* (root *yuj* ‘yoke’) occurs eight times in intransitive (passive) syntactic patterns (e.g., *rátho yuj-ānāḥ* ‘a chariot that has been yoked’) and 14 times in transitive syntactic patterns (as in *yuj-ānó haritā ráthe* ‘yoking two fallow [horses] to the chariot’). Vedic grammars treat *yujāná-* as a middle participle of the root aorist (see, for instance, Whitney 1885:132, Macdonell 1910:370). However, again, as in the case of *hinvāná-*, the corresponding finite forms (3sg. *áyukta* etc.) can only be used transitively. Kulikov (2006a, 2006b) has demonstrated that the grammatical characteristics of such passive *-āna-*participles should be reconsidered; here, we will only briefly summarize the conclusions of these articles. These participles are grammatically ambiguous, that is, they belong to the following two paradigms: (1) to the paradigm of (middle) root aorist and mediopassive

aorist and (2) to the paradigm of (middle) present and stative. Thus, the participle *hinvāná* in its transitive use, meaning ‘impelling’, belongs to the paradigm of the transitive nasal present (*hinváte*, etc.), but it is a member of the paradigm of the stative (3sg. *hinvé*, 3pl. *hinviré*), i.e., a stative participle, when used intransitively (in a passive syntactic pattern), meaning ‘impelled’. Similarly, *yujāná-* is a member of the paradigm of the (transitive) root aorist (*áyukta*, etc.) when used transitively (‘yoking’), but it is a member of the paradigm of the passive aorist (3sg. *áyoji*, 3pl. *ayujran*), that is, a passive aorist participle, when used in passive syntactic patterns (‘yoked’):

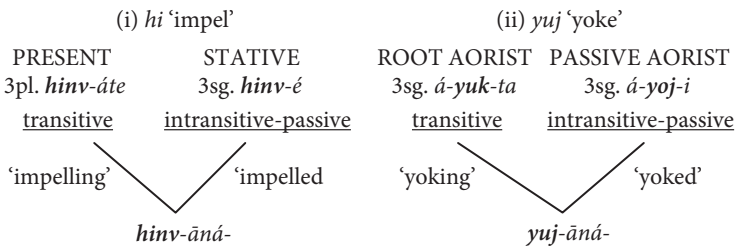


Figure 2. Syntax of the participles *hinv-āná-* (*hi* ‘impel’) and *yuj-āná-* (*yuj* ‘yoke’)

All in all, these data show that the *-āna*-participles attested in passive constructions cannot be considered non-specialized middle forms (e.g., middle present participles, in the case of *hinvāná-* ‘impelled’, or middle aorist participles, in the case of *yujāná-* ‘yoked’), and therefore should be discarded as evidence for the passive usage of non-specialized middle forms. As members of the paradigms of the specialized passive formations, that is, statives and passive aorists, forms such as *hinvāná-* ‘(that has) been impelled’ or *yujāná-* ‘(being) yoked’ should be treated separately from the homonymous middle present participles (*hinvāná-* ‘impelling’) or middle aorist participles (*yujāná-* ‘yoked’), which can never be used in passive patterns.

Another large group of middle forms used in passive syntactic patterns includes **middle perfects**. Most remarkably, only 3sg. and 3pl. middle perfects forms (with the endings *-e* and *-re*, respectively) are attested in passive constructions (for a detailed study of Vedic perfects, see Kümmel 2000). As argued in Kulikov (2006a), all such forms should be taken as statives built on perfect stems, rather than as middle perfects proper. For instance, the form *dadhé* (root *dhā* ‘put’) should be taken to be a 3sg. form of the middle perfect when meaning ‘has put’, as in (17a), and a 3sg. form of the stative with the passive interpretation ‘is put / has been put’, as in (17b):

(17) Vedic

a. (RV 9.18.4)

yó víśvāni vāryā vásūni
 who:NOM.SG.M all:ACC.PL.N desirable:ACC.PL.N goods:ACC.PL
 hástayor **dadh-é** ...
 hand:LOC.DU put:PRF-3SG.MID

‘The one who **holds / has put** all desirable goods in his hands ...’

b. (RV 1.168.3)

hásteṣu khādís ca kṛtís ca sám
 hand:LOC.PL brooch:NOM.SG and sword:NOM.SG and together
dadh-é
 put:STAT-3SG.MID

‘Brooch and sword **is put** in [your] hands.’

Furthermore, there are some reasons to assume that stative *-āna*-participles could have active counterparts, that is, participles derived from the stem of stative with the active participle suffix *-ant*. It has frequently been noted (e.g., Watkins 1969: 142ff., Schaefer 1994: 45f.) that the form *stavánt-*, which occurs three times in the family maṇḍalas of the Ṛgveda, at 2.19.5, 2.20.5, and 6.24.8 (only in the nom. sg. form *staván*), is attested in a passive construction, which is quite unusual for an active form, cf.:

(18) Vedic (RV 2.20.5c)

muṣṇānn uśāsaḥ sūryeṇa
 steal:PRS:PTCP.ACT:NOM.SG.M dawn:ACC.PL sun:INS.SG
staván
 praise:STAT:PTCP.ACT:NOM.SG.M

‘while (he), the **praised** one, abducted the dawns with the sun’

By virtue of its suffix accentuation and active morphology, this form cannot belong to the class I present *stávate*, which is only attested in the middle. However, its meaning and passive syntactic pattern plead for the connection of this form with the stative *stáve* as the active counterpart of the (middle) participle *stávāna-*.

The assumption of the existence of active stative participles may elucidate the paradigmatic status of some other forms with *-ant-* (most of which are traditionally regarded as adjectives). Watkins, who first drew attention to these formations (1969: 142ff., see also Schaefer 1994: 45f.), assumed that they represent active participles with the secondary accent shift marking their passive syntactic pattern. These participles may include the following, in addition to *stavánt-*: (i) *járant-* ‘old’ (i.e. ‘(having) grown old’); (ii) *pépiśat-* ‘adorned’ (RV 10.127.7, see Schaefer 1994: 45, 152f.; this form may point to the unattested stative **pépiše* ‘is adorned’ of the type *cékite*, on which see Schaefer 1994: 44); (iii) *mahánt-* ‘great’ (whose

parallelism with *stavánt-* was noted by Watkins 1969: 144) that may belong with the hapax stative *mahe* ‘is able’ (RV 7.97.2); and some others (see Kulikov 2006b: 59ff. for details). These forms represent a good structural parallel to Hittite participles with *-ant-* that form passives for transitive verbs (but hitherto have generally been considered isolated phenomena, without parallels in other IE branches), such as *kunant-* ‘killed’ (*kuen-* ‘to kill’) or *dant-* ‘taken’ (*dā-* ‘to take’); see, in particular, Hoffner & Melchert (2008), Šackov (2008), and Luraghi (2010b).

The hypothesized existence of active participles in the stative paradigm, which, in spite of their ‘active’ morphology, were employed in passive constructions, still further supports the connection between the stative formation and passive syntactic pattern. On the other hand, it serves as additional evidence against the traditional assumption on the straightforward connection between middle morphology and passive syntax in Vedic and thus, eventually, against taking such forms as *hinváná-* ‘impelled’ as non-characterized middles used in syntactic patterns.

The sub-paradigm of the present passive is, in fact, also defective. We primarily find 3sg. and 3pl. forms of the present tense and participles. Next to present tense forms, there are rare imperatives and only exceptional attestations of other moods, which makes the sub-paradigm of the present quite similar to those of the aorist and perfect — much more similar than it is described in traditional Vedic grammar, which usually pays little attention to the fact that some members of the present passive paradigm are virtually unattested in early Vedic.

The early Vedic passive paradigm is summarized in Table 2. Different types of shadowing show the status of the corresponding forms: dark grey means lacking and morphologically impossible, intermediate grey means morphologically

Table 2. The passive paradigm in early Vedic (from Kulikov 2006a)

		PRESENT		AORIST		PERFECT/STATIVE	
		INDICATIVE	IMPERATIVE	INDICATIVE	INJ.	INDICATIVE	IMPER.
SG	1	...	...				
	2	..., <i>yuyjáse</i>	[<i>dhīyasva</i>]			[<i>śṛṇviṣé</i>]	
	3	<i>sūyáte, yuyjáte</i>	[<i>dhīyātām</i>]	<i>ásāvi, áyoji</i>	..., <i>yoji</i>	<i>sunvé, yuyujé</i>	[<i>duháṁ</i>]
DU	1	...	...				
	2	...	...				
	3	[<i>ucyete</i>]	...				
PL	1	[<i>-panyámahe</i> (?)]	...				
	2	...	<i>yuyyadhvam</i>				
	3	..., <i>yuyjante</i>	[<i>badhyantām</i>]	..., <i>áyujran</i>	...	<i>sunvire, yuyujré</i>	
PTCP		<i>sūyámāna-, yuyjámāna-</i>		<i>s^hvāná-, yujāná-</i>		<i>sunvāná-, yuyujāná-</i>	

possible but unattested or only exceptionally attested (underdeveloped part of the paradigm), light grey means morphologically possible but rare.

The bare middle forms attested in the passive that remain after sifting through the Vedic evidence (i.e., after explaining the status of the passive *-āna*-participles and 3sg. and 3pl. middle perfects) form a small set. Most of these forms can be explained as secondary formations created on the basis of specialized passive formations such as *stāve* or *gr̥ne* (see Kulikov 2006a); cf. class I prs. *stāvate* ‘is praised’, class IX prs. *gr̥ñitē* ‘is praised’ (based on the statives *stāve* and *gr̥ne* ‘is praised’, as a type of back derivation [*Rückbildung*]); sigmatic aorists (mostly 3pl. forms) *ayukṣata* ‘(they) were yoked’, or *adr̥kṣata* ‘(they) were seen, visible, (they) appeared’.⁷ Thus, middle morphology cannot be said to serve as the regular **independent** marker of the passive voice — that is the morphological marker that can encode passive syntactic pattern on its own, without using **additional** (specialized) morphemes, such as the present passive suffix *-yá-*.

In this section, we have discussed the paradigm and functional status of both non-specialized markers (middle morphology) and specialized morphemes encoding the passive pattern in Early Vedic and Greek. Although passive morphology in the strict sense of the term (that is, morphological marking regularly encoding a passive syntactic pattern) cannot be reconstructed for the proto-language, passive **syntactic patterns** associated with **non-specialized** verbal morphemes can well be reconstructed for the Indo-European proto-language. Accordingly, the existence of the passive pattern can be posited for PIE verbal syntax in spite of the lack of the passive voice *sensu stricto*.

5. Concluding remarks: Proto-Indo-European sources of Indo-European passives

A detailed and well-substantiated reconstruction of the PIE voice and voice-related categories goes beyond the scope of the present article, which analyzed in detail the problems of such a reconstruction (for the evolution of the [early] PIE system of verbal endings, see especially Kortlandt 1979:66–68 *et passim*, 1981:128–129 *et passim*). Here, we will confine ourselves to several remarks on this issue. It is reasonable to assume that the IE categories ‘middle’ and ‘perfect’ (see above, Section 4) are historically related and most likely originate in one

7. Regarding the development of anticausative uses of passives made from verbs of perception, see Kulikov (2011b).

single proto-category.⁸ This hypothesis, going back as far as Kuryłowicz (1932) and Stang (1932), is primarily based on the fact that the sets of endings used by the middle voice and the active perfect share a number of features (but cf. also somewhat different views and the reconstruction in Sihler 1995:445, 564f.).⁹ Thus, in early PIE (= Stage I), the active/middle opposition would have been irrelevant for perfect forms. The vestiges of this system can still be found in early Vedic, where the active perfect of some verbs syntactically correspond to middle presents, i.e., they function as non-passive intransitives (for a typological discussion of this phenomenon in ancient Indo-European languages, see Kulikov 1999, where it is labeled ‘split causativity’); cf. the middle present *pádyate* ‘falls’ corresponding to the active perfect *papáda* ‘has fallen’ or the middle present *mriyáte* ‘dies’ corresponding to the active perfect *mamára* ‘has died’ (see, e.g., Hoffmann 1976, Jasanoff 1978:15, Kümmel 2000:296f., 370ff. *et passim*). Perfect forms of some verbs could be used both intransitively and transitively, thus being syntactically labile (see Kulikov 2014a: 1151–1152).

Similar remarks can be made for Greek: the early Greek perfect shows a clear preference for the active voice (some media tantum have an active perfect; see Szemerényi 1990 and, on the perfect, Di Giovine 1990–1996). Furthermore, the early Greek perfect shows preference for intransitives, whereas all perfect forms in Mycenaean Greek are either participles of intransitive verbs or participles of transitive verbs used in passive constructions (Hooker 1980:61–62). Perfect forms in Homeric Greek are mostly attested either for intransitive verbs or for transitive verbs that appear in intransitive syntactic patterns in the perfect: *pépoitha* (persuade:PRF.1SG) ‘I am persuaded’ (Luraghi et al. 2005).¹⁰

During the next stage (II = [standard] PIE), we may reconstruct a number of innovations resulting from a contamination and/or analogical rebuilding of inflection due to mutual influence of the endings belonging to different sets. In particular, some elements of the ‘perfect-stative’ inflection would have been introduced into the present paradigm (see Kortlandt 1979:67). These newly built forms must have retained the functional connection with statives, which were closely associated with intransitive syntax. This connection could be the origin of the middle

8. We are not going to discuss here the status of other verbal categories — in particular, the quite complicated issue of the origins of the aorist tense. There are some reasons to believe that this category is a recent addition to the verbal system, perhaps going back to some nominal formations (see especially Kortlandt 2009:57, 187, Kortlandt 2010).

9. Cf., for instance, Ved. 1sg.mid. (athematic secondary ending) *-i* (< *-H₂) ~ 1sg.prf.act. *-a* (< *-H₂e), 2sg.mid. *-thās* (secondary ending) ~ 2sg.prf.act. *-tha*, etc.

10. We would like to thank Silvia Luraghi (p.c.) for relevant remarks on this point.

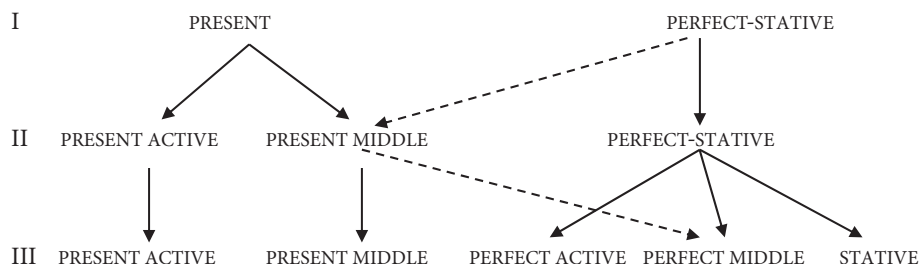


Figure 3. The emergence and development of the middle in IE (from Kulikov 2006a).

voice that was used to mark several intransitive patterns, such as passive, anticausative, reflexive and reciprocal.

Finally, during Stage III, in some IE dialects (in particular, in Proto-Indo-Iranian), the active/middle distinction was introduced into the perfect paradigm under the influence of the present system, which results in the universal character of the active/middle opposition applied across the paradigm (for details, see Renou 1925: 103ff., Jasanoff 1978: 16, 81f., Jasanoff 2003, Kümmel 2000: 94). This scenario is schematically represented in Figure 3.

The position of the passive within the PIE can accordingly be summarized as follows:

- *early PIE*
passive: expressed by some **statives**
- *(standard) PIE*
passive: (i) (some **statives**), (ii) **middle** forms
- *late PIE (differences between dialects)*
passive: (i) (rare with **statives**), (ii) **middle** forms, (iii) specialized morphemes (Ved. prs. -yá-, -i-aorist)

On the basis of the above discussion of the PIE sources of the passive morphology in individual Indo-European branches, the main tendencies in the development of the PIE verbal syntax for different parts of the IE area can be presented as follows (see Kulikov 2009 and cf. also Table 3). Several Western groups of IE either preserve the old syncretic marker of valency-reducing categories, the middle conjugation, or replace the old middle with a new middle morphology, which in most cases can be traced back to the PIE reflexive pronoun **s(u)e-* (see, for instance, Cennamo 1993). Such a scenario of the evolution of the middle is instantiated, for instance, by Latin as it evolves with its daughter languages (Romance) (see Kemmer 1993: 151–182, Cennamo 1993, 1998, Heidinger 2010). Greek, which forms one of the Western, syncretic, branches, exemplifies a slightly different scenario of development (see Lavidas 2009). The development of voice in Greek results in the new/innovative marking of the passive, the reflexive and the reciprocal

Table 3. Western vs. Eastern IE languages: Specialized passive morphemes, syncretic middle, and lability

<i>Western IE languages: Syncretic type</i> (Greek, Germanic, Romance, Slavic ...)	<i>Eastern IE languages: Anti-syncretic type</i> (Indo-Aryan, Iranian, Armenian ...)
Increase of lability	Decline of lability
Preservation of the old (PIE) middle morphology or emergence of the new middle	Degrammaticalization of the old Indo-European middle
Preservation of the syncretic encoding of passive syntactic pattern	Development of specialized passive morphology
Decline and disappearance of the old (PIE) causative morphology	Development and increase of productivity of the causative morphology

with a syncretic (mediopassive) morphology in contrast to the anticausative and all transitive patterns that become obligatorily marked with active morphemes after the loss of the auto-benefactive meaning (that was originally expressed by the mediopassive). Furthermore, a number of Romance and Germanic languages attest the rise and expansion of labile patterning (which becomes particularly common and productive in English; see McMillion 2006 and van Gelderen 2011); the expansion of labile verbs is also well attested in the development of Greek (Lavidas 2009). Finally, the PIE causative morpheme **-eje/o-*, still present in Gothic (*jan-*verbs) and Old Church Slavonic (*i-causatives*), virtually disappears in this area leaving only minimal traces in the modern Germanic and Slavonic languages. This type of evolution, attested in the Western part of the IE area, might be called ‘syncretic’.

By contrast, the main tendencies in the development of voice in several Eastern IE languages, foremost in Indo-Iranian, can be summarized as follows (these tendencies are partly related to, but not entirely dependent on, one another). One of the main trends is the rise and development of specialized voice and valency-changing morphemes, cf. the Vedic causative suffix *-áya-* (see Jamison 1983) and the passive suffix *-yá-* (see Kulikov 2012). The specialized voice and valency-changing morphemes bring the language to a more overt morphological marking of transitivity oppositions and essentially run in parallel with the decay of labile patterning and the degrammaticalization of the middle voice. We observe these tendencies, in particular, in the evolution of: (a) the Indo-Aryan and Armenian markers of morphological passive, going back to the PIE suffix **-je/o-* as well as in the development and grammaticalization of free morphemes encoding some voice-related categories; (b) the Indo-Iranian reflexive pronouns *tanú-* (← ‘body’) and Indo-Aryan *ātmán-* (← ‘breath’) (for details, see Kulikov 2007b); and (c) the Indo-Iranian reciprocal pronouns (cf. Vedic *anyó’nyá-*; see Kulikov 2007a; 2014b).

Furthermore, morphological causatives become quite productive both in Armenian (causative marker *-uc'anem* based on the nasal present derived from a sigmatic aorist; see Kortlandt 1999) and Indo-Iranian (productive morphological causative suffixes *-áya-*, *-aiia-* going back to PIE **-eje/o-*). Note an interesting feature (isogloss) shared by several Eastern IE languages of the anti-syncretic type: the parallel development of the new morphological passive and the productive morphological causative. The PIE middle voice is degrammaticalized (that is, the active/middle opposition, albeit physically preserved in the paradigm, loses a large part of its grammatical content) and eventually disappears. The labile syntax, even if attested in some ancient languages of the Eastern branches, tends to disappear in the course of their history. One might call this type 'anti-syncretic'.

Thus, we observe two basic types of evolution, or evolutionary types, in the history of the system of transitivity oppositions and valency-changing categories in IE: the syncretic type found in many Western branches, including Greek, and the anti-syncretic type attested at least in some Eastern branches, in particular in Indo-Aryan. It is important to note that the degrammaticalization of the middle and the lexicalization of middle forms are supported by the grammaticalization of several new valency-changing categories (*-yá-*passives and *-áya-*causatives, reflexives with *ātmán-* and reciprocals with *anyó'nyá-*).

In the present article, we have focused on particular aspects of the status and the role of the non-specialized markers of the passive in PIE and their development, primarily in Vedic and Greek. The evidence discussed in this short survey supports the idea that, although passive morphology in the strict sense of the term (that is, regular encoding of the passive syntactic pattern by means of **specialized** verbal morpheme(s)) cannot be reconstructed for the proto-language, we have good reasons to reconstruct passive **syntactic patterns** associated with **non-specialized** verbal morphemes. In other words, we can posit the existence of the passive for the verbal system of PIE, despite the lack of the passive voice *sensu stricto*. The rise of specialized passive markers is arguably related to, and partly supported by, the development of other valency-changing categories; cf. (i) the active morphology that becomes the specialized marker of transitive syntactic patterns (after the loss of the auto-benefactive meaning of the mediopassive) in Greek from the Hellenistic Koine onwards or (ii) the development of productive causatives in Vedic.

References

- Allan, Rutger. 2003. *The Middle Voice in Ancient Greek: A Study of Polysemy*. Amsterdam: J. Gieben.
- Barðdal, Jóhanna. 2013. Construction-Based Historical-Comparative Reconstruction. *Oxford Handbook of Construction Grammar* ed. by Graeme Trousdale & Thomas Hoffmann, 438–457. Oxford: Oxford University Press.
- Barðdal, Jóhanna & Thórhallur Eythórsson. 2012. Reconstructing Syntax: Construction Grammar and the Comparative Method. *Sign-Based Construction Grammar* ed. by Hans C. Boas & Ivan A. Sag, 257–308. Stanford: CSLI Publications.
- Barðdal, Jóhanna & Valéria Molnár. 2003. The Passive in Icelandic — Compared to Mainland Scandinavian. *Structures of Focus and Grammatical Relations* ed. by Jorunn Hetland & Valéria Molnár, 231–260. Tübingen: Niemeyer.
- Benedetti, Marina. 2006. Mehr als Passiv: Über einige Verbalmorpheme in altindogermanischen Sprachen. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 3:2.91–110.
- Cennamo, Michela. 1993. *The Reanalysis of Reflexives: A Diachronic Perspective*. Napoli: Liguori.
- Cennamo, Michela. 1998. The Loss of the Voice Dimension between Late Latin and Early Romance. *Historical Linguistics 1997: Selected Papers from the 13th International Conference on Historical Linguistics* ed. by Monika Schmid, Jennifer Austin & Dieter Stein, 77–100. Amsterdam: John Benjamins. DOI: 10.1075/cilt.164.06cen
- Delbrück, Berthold. 1888. *Altindische Syntax*. Halle a. S.: Verlag der Buchhandlung des Waisenhauses.
- Delbrück, Berthold. 1897. *Vergleichende Syntax der indogermanischen Sprachen. Zweiter Teil*. Strassburg: Trübner.
- Di Giovine, Paolo. 1990–1996. *Studio sul perfetto indoeuropeo*, Vol. I–III. Roma: Il Calamo.
- Dowty, David. 1990. Thematic Proto-Roles and Argument Selection. *Language* 67:3.547–619. DOI: 10.1353/lan.1991.0021
- Ernout, Alfred. 1908–1909. Recherches sur l'emploi du passif latin à l'époque républicaine. *Mémoires de la Société de Linguistique* 15.273–333.
- Flobert, Pierre. 1975. *Les verbes déponents latins des origines à Charlemagne*. Paris: Les belles lettres.
- Friedrich, Johannes. 1967. *Hethitisches Elementarbuch*, 3rd edition, Vol. I–II. Heidelberg: Carl Winter.
- van Gelderen, Elly. 2011. Valency Changes in the History of English. *Journal of Historical Linguistics* 1:1.106–143. DOI: 10.1075/jhl.1.1.05van
- Geniušienė, Emma. 1987. *The Typology of Reflexives*. (= Empirical Approaches to Language Typology, 2.) Berlin: Walter de Gruyter. DOI: 10.1515/9783110859119
- George, Coulter H. 2005. *Expressions of Agency in Ancient Greek*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511482984
- Gildersleeve, Basil L. 1900. *Syntax of Classical Greek from Homer to Demosthenes*. New York: American Book Company.
- Givón, Talmy. 1981. Typology and Functional Domains. *Studies in Language* 5:2.163–193. DOI: 10.1075/sl.5.2.03giv
- Gonda, Jan. 1979. *The Medium in the R̥gveda*. Leiden: Brill.
- Goodwin, William W. 1978 [1894]. *A Greek Grammar*. London: St. Martin's Press.

- Gotō, Toshifumi. 1997. Überlegungen zum urindogermanischen “Stativ”. *Berthold Delbrück y la sintaxis indoeuropea hoy. Actas del Coloquio de la Indogermanische Gesellschaft* ed. by Emilio Crespo & José Luís, 165–192. Madrid: UAM; Wiesbaden: Reichert.
- Grosse, Hermann. 1889. *Beiträge zur Syntax des griechischen Mediums und Passivums*. Dramburg: Druck von Th. Kämpf.
- Haspelmath, Martin & Thomas Müller-Bardey. 2004. Valency Change. *Morphology: A Handbook on Inflection and Word Formation*, Vol. 2 ed. by Geert Booij, Christian Lehmann & Joachim Mugdan, 1130–1145. Berlin: Walter de Gruyter.
- Heidinger, Steffen. 2010. *French Anticausatives: A Diachronic Perspective*. Berlin: Walter de Gruyter. DOI: 10.1515/9783110251357
- Hoffmann, Karl. 1976. Die Aoristbildungen von ved. *vṛt*. In K. Hoffmann, *Aufsätze zur Indoiranistik. Bd. 2* ed. 562–569. Wiesbaden: Reichert.
- Hoffner, Harry A. & Craig H. Melchert. 2008. *A Grammar of the Hittite Language*. Winona Lake: Eisenbrauns.
- Hooker, James T. 1980. *Linear B: An Introduction*. Bristol: Classical Press.
- Humbert, Jean. 1945. *Syntaxe Grecque*. Paris: Librairie Klincksieck.
- Jamison, Stephanie W. 1983. *Function and Form in the -āya-formations of the Rig Veda and Atharva Veda*. (= KZ, Ergänzungsheft 31.) Göttingen: Vandenhoeck & Ruprecht.
- Jankuhn, Harald. 1969. *Die passive Bedeutung medialer Formen untersucht an der Sprache Homers*. (= KZ, Ergänzungsheft 21.) Göttingen: Vandenhoeck & Ruprecht.
- Jannaris, Antonius. 1968 [1897]. *An Historical Greek Grammar, Chiefly of the Attic Dialect, as Written and Spoken from Classical Antiquity down to the Present Time, Founded upon the Ancient Texts, Inscriptions, Papyri, and Present Popular Greek*. Hildesheim: Olms.
- Jasanoff, Jay. 1978. *Stative and Middle in Indo-European*. (= Innsbrucker Beiträge zur Sprachwissenschaft, 23.) Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Jasanoff, Jay. 2003. *Hittite and the Indo-European Verb*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249053.001.0001
- Kemmer, Suzanne. 1993. *The Middle Voice*. (= Typological Studies in Language, 23.) Amsterdam: John Benjamins. DOI: 10.1075/tsl.23
- Kortlandt, Frederik H.H. 1979. Toward a Reconstruction of the Balto-Slavic Verbal System. *Lingua* 49.51–70. DOI: 10.1016/0024-3841(79)90074-3
- Kortlandt, Frederik H.H. 1981. 1st sg. middle *-H2. *Indogermanische Forschungen* 86.123–136.
- Kortlandt, Frederik H.H. 1999. The Armenian Causative. *Annual of Armenian Linguistics* 20.47–49.
- Kortlandt, Frederik H.H. 2009. *Baltica & Balto-Slavica*. Amsterdam: Rodopi.
- Kortlandt, Frederik H.H. 2010. An Outline of Proto-Indo-European. In: F.H.H. Kortlandt, *Studies in Germanic, Indo-European and Indo-Uralic*, 37–45. Amsterdam: Rodopi.
- Krause, Wolfgang. 1968. *Handbuch des Gotischen*. 3rd ed. München: C.H. Beck'sche Verlagsbuchhandlung.
- Kühner, Raphael & Bernhard Gerth. 1963 [1898–1904]. *Ausführliche Grammatik der Griechischen Sprache*. Zweiter Teil: Satzlehre. Hannover: Hannsche Buchhandlung.
- Kulikov, Leonid. 1999. Split Causativity: Remarks on Correlations between Transitivity, Aspect, and Tense. *Tense-Aspect, Transitivity and Causativity: Essays in Honour of Vladimir Nedjalkov* (= Studies in Language Companion Series, 50) ed. by Werner Abraham & Leonid Kulikov, 21–42. Amsterdam: John Benjamins. DOI: 10.1075/slcs.50.06kul

- Kulikov, Leonid. 2006a. Passive and Middle in Indo-European: Reconstructing the Early Vedic Passive Paradigm. *Passivization and Typology: Form and Function* (= Typological Studies in Language, 68) ed. by Werner Abraham & Larisa Leisiö, 62–81. Amsterdam: John Benjamins. DOI: 10.1075/tsl.68.06kul
- Kulikov, Leonid. 2006b. The Vedic Mediopassive Aorists, Statives and their Participles: Reconsidering the Paradigm. *Themes and Tasks in Old and Middle Indo-Aryan Linguistics* ed. by Bertil Tikkani & Heinrich Hettrich, 45–63. Delhi: Motilal Banarsidass.
- Kulikov, Leonid. 2007a. Reciprocal Constructions in Vedic. *Reciprocal Constructions*, Vol. 2 (= Typological Studies in Language, 71) ed. by Vladimir P. Nedjalkov, 709–738. Amsterdam: John Benjamins. DOI: 10.1075/tsl.71.23kul
- Kulikov, Leonid. 2007b. The Reflexive Pronouns in Vedic: A Diachronic and Typological Perspective. *Lingua* 117:8.1412–1433. DOI: 10.1016/j.lingua.2006.05.009
- Kulikov, Leonid. 2009. Valency-Changing Categories in Indo-Aryan and Indo-European: A Diachronic Typological Portrait of Vedic Sanskrit. *Multilingualism: Proceedings of the 23rd Scandinavian Conference of Linguistics* (= Studia Linguistica Upsaliensis, 8) ed. by Anju Saxena & Åke Viberg, 75–92. Uppsala: Uppsala Universitet.
- Kulikov, Leonid. 2011a. Voice Typology. *The Oxford Handbook of Linguistic Typology* ed. by Jae-Jung Song, 368–398. Oxford: Oxford University Press.
- Kulikov, Leonid. 2011b. Passive to Anticausative through Impersonalization: The Case of Vedic and Indo-European. *Impersonal Constructions: A Cross-Linguistic Perspective* (= Studies in Language Companion Series, 124) ed. by Anna Siewierska & Andrej Malchukov, 229–254. Amsterdam: John Benjamins. DOI: 10.1075/slcs.124.08kul
- Kulikov, Leonid. 2012. *The Vedic -ya-presents: Passives and Intransitivity in Old Indo-Aryan*. (= Leiden Studies in Indo-European, 19.) Amsterdam: Rodopi.
- Kulikov, Leonid. 2014a. The decline of labile syntax in Old Indo-Aryan: A diachronic typological perspective. *Linguistics* 52:4 (special issue: *Typology of labile verbs: Focus on diachrony*). 1139–1165.
- Kulikov, Leonid. 2014b. Grammaticalization of reciprocal pronouns in Indo-Aryan: Evidence from Sanskrit and Indo-European for a diachronic typology of reciprocal constructions. *Journal of South Asian Languages and Linguistics* 1:2.117–156.
- Kümmel, Martin. 1996. *Stativ und Passivaorist im Indoiranischen*. (= HS, Ergänzungsheft 39.) Göttingen: Vandenhoeck & Ruprecht.
- Kümmel, Martin. 2000. *Das Perfekt im Indoiranischen*. Wiesbaden: Reichert.
- Kurylowicz, Jerzy. 1932. Les désinences moyennes de l'indo-européen et du hittite. *Bulletin de la Société de Linguistique de Paris* 30:1.1–4.
- Kurylowicz, Jerzy. 1935. *Études indoeuropéennes I*. Krakow: Gebethner & Wolff.
- Kurylowicz, Jerzy. 1964. *The Inflectional Categories of Indo-European*. Heidelberg: C. Winter.
- Lavidas, Nikolaos. 2009. *Transitivity Alternations in Diachrony: Changes in Argument Structure and Voice Morphology*. Newcastle: Cambridge Scholars Publishing.
- Leumann, Manu & Johann B. Hofmann. 1928. *Lateinische Grammatik, Bd. 2, Syntax*. München: Beck.
- Luraghi, Silvia. 2010a. The Extension of the Transitive Construction in Ancient Greek. *Acta Linguistica Hafniensia*. 42:1.60–74. DOI: 10.1080/03740463.2010.486919
- Luraghi, Silvia. 2010b. Transitivity, Intransitivity and Diathesis in Hittite. *Indo-evropejskoe jazыkoznanie i klassičeskaja filologija* 14.133–154.
- Luraghi, Silvia, Anna Pompei & Stavros Skopeteas. 2005. *Ancient Greek*. Munich: Lincom Europa.

- Macdonell, Arthur Anthony. 1910. *Vedic Grammar*. Strassburg: Trübner.
- McMillion, Alan. 2006. *Labile Verbs in English: Their Meaning, Behavior and Structure*. Stockholm: Stockholm University PhD dissertation.
- Meillet, Antoine. 1964 [1903]. *Introduction à l'étude comparative des langues indo-européennes*. Alabama: University of Alabama Press.
- Meillet, Antoine & Joseph Vendryes. 1924. *Traité de grammaire comparée des langues classiques*. Paris: Champion.
- Mel'čuk, Igor. 1993. The Inflectional Category of Voice: Towards a more Rigorous Definition. *Causativity and Transitivity* (= Studies in Language Companion Series, 23) ed. by Bernard Comrie & Maria Polinsky, 1–46. Amsterdam: John Benjamins. DOI: 10.1075/slcs.23.02mel
- Neu, Erich. 1968a. *Das hethitische Mediopassiv und seine indogermanischen Grundlagen*. (= Studien zu den Boğazköy-Texten, 6.) Wiesbaden: Harrassowitz.
- Neu, Erich. 1968b. *Interpretation der hethitischen mediopassiven Verbalformen*. (= StBoT, 5.) Wiesbaden: Harrassowitz.
- Renou, Louis. 1925. *La valeur du parfait dans les hymnes védiques*. (= Collection linguistique, 18.) Paris: Champion.
- Schaefer, Christiane. 1994. *Das Intensivum im Vedischen*. Göttingen: Vandenhoeck & Ruprecht.
- Schmidt, Karl Horst. 1963a. Präteritum und Medio-Passiv. *Sprache* 9.14–20.
- Schmidt, Karl Horst. 1963b. Zum altirischen Passiv. *Indogermanische Forschungen* 68.257–275.
- Schwyzler, Eduard. 1943. *Zum persönlichen Agens beim Passiv, besonders im Griechischen*. (= Abhandlungen der Preußischen Akademie der Wissenschaften. Philosophische-historische Klasse, Jahrgang 1942: Nr 10.) Berlin: Akademie der Wissenschaften.
- Schwyzler, Eduard & Albert Debrunner. 1950. *Griechische Grammatik*. Zweiter Band: Syntax und syntaktische Stilistik. (= Handbuch der Altertumswissenschaft, 2. Abt., 1. Teil, 2. Bd.) München: C.H. Beck.
- Shibatani, Masayoshi. 2006. On the Conceptual Framework for Voice Phenomena. *Linguistics* 44.217–269. DOI: 10.1515/LING.2006.009
- Sihler, Andrew L. 1995. *New Comparative Grammar of Greek and Latin*. Oxford: Oxford University Press.
- Siewierska, Anna. 1984. *The Passive: A Comparative Linguistic Analysis*. London: Croom Helm.
- Smyth, Herbert W. 1984 [1920]. *Greek Grammar*, 13th ed. Cambridge, MA: Harvard University Press. (rev. ed. by Gordon M. Messing.)
- Stahl, Johann M. 1907. *Kritisch-historische Syntax des griechischen Verbums der klassischen Zeit*. Heidelberg: Winter.
- Stang, Christian S. 1932. Perfektum und Medium. *Norsk tidsskrift for sprogvidenskap* 6.29–39.
- Szemerényi, Oswald. 1990. *Einführung in die vergleichende Sprachwissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft. (4., durchgesehene Auflage.)
- Šackov, Andrej B. 2008. Xettskie glagoly s suffiksom -nu- [Hittite Verbs with the Suffix -nu-]. *Indoeuropejskoe jazykoznanie i klassičeskaja filologija* 12.471–474.
- Wackernagel, Jacob. 1904. Studien zum griechischen Perfectum. *Programm zur akademischen Preisverteilung*, 3–24. Göttingen. [Reprinted in: Jacob Wackernagel. 1953. *Kleine Schriften*, 1000–1021. Göttingen: Vandenhoeck & Ruprecht.]
- Watkins, Calvert. 1969. *Indogermanische Grammatik, III: Formenlehre. 1. Teil. Geschichte der indogermanischen Verbalflexion*. Heidelberg: Winter.
- Whitney, William Dwight. 1885. *The Roots, Verb-Forms, and Primary Derivatives of the Sanskrit Language*. Leipzig: Breitkopf & Härtel.

Toward a syntactic phylogeny of modern Indo-European languages*

Giuseppe Longobardi^{a,b}, Cristina Guardiano^c,
Giuseppina Silvestri^{a,d}, Alessio Boattini^e and Andrea Ceolin^a
University of Trieste^a / University of York^b / University of Modena e Reggio
Emilia^c / University of Pisa^d / University of Bologna^e

The Parametric Comparison Method (PCM, Guardiano & Longobardi 2005, Longobardi & Guardiano 2009) is grounded on the assumption that syntactic parameters are more appropriate than other traits for use as comparanda for historical reconstruction, because they are able to provide unambiguous correspondences and objective measurements, thus guaranteeing wide-range applicability and quantitative exactness. This article discusses a set of experiments explicitly designed to evaluate the impact of parametric syntax in representing historical relatedness, and performed on a selection of 26 contemporary Indo-European varieties. The results show that PCM is in fact able to correctly identify genealogical relations even from modern languages only, performing as accurately as lexical methods, and that its effectiveness is not limited by interference effects such as ‘horizontal’ transmission. PCM is thus validated as a powerful tool for the analysis of historical relationships not only on a long-range perspective (as suggested by Longobardi & Guardiano 2009), but even on more focused, though independently well-known domains.

1. Introduction

The historical classification of Indo-European languages is traditionally based on the inspection of **lexical** entities (roots and grammatical morphemes) and sound laws connecting cognate words crosslinguistically. In contrast, syntactic structures have rarely been regarded as encoding relevant genealogical information.

* For useful suggestions and discussions we are especially indebted to F. Bernasconi, R. Lazzeroni, N. McCoy, C. Melchert, I. Roberts, D. Willis and two anonymous referees. The elaboration of this article has benefited from the ERC Advanced Grant n. 295733 (2012–2017) LANGEIN, awarded to G. Longobardi. This article is dedicated to the memory of Calvert Watkins (1933–2013), who kindly discussed some preliminary ideas pursued here with two of us.

Such supposed lack of parallelism between syntactic and lexical classification was most explicitly challenged in Longobardi & Guardiano (2009). That article used a new taxonomic technique crucially based only on syntax, namely the Parametric Comparison Method (PCM), which is built on the insights of modern generative theory. Its main claim is that even syntax codes for phylogenetic history once a sufficient number of characters is considered.

The possibility of a Parametric Comparison Method was first exposed in Longobardi (2003). In Guardiano & Longobardi (2005), it was shown that the analysis of a few languages in terms of PCM could satisfy certain conceptual conditions of adequacy (scattering of language distances, the Anti-Babelic principle). Longobardi & Guardiano (2009) argued that PCM provided language distances and preliminary taxonomies in good empirical agreement with some independent historical evidence, within and outside of the Indo-European family. Bortolussi et al. (2011) argued that the probability of many of the language distances that are empirically calculated by PCM is significant beyond chance, and hence calls for a historical explanation. Finally, Longobardi (2012) argued that some apparently paradoxical instances of parallel developments can be readily explained by the structure of the syntactic theory underlying PCM.

Thanks to its reliance on a detailed theory of Universal Grammar (in particular a universal list of parameters), PCM is in principle capable of comparing any pair or set of natural languages. Actually, it was devised in order to potentially address unsolved long-range genealogical questions; to fully assess and increase its adequacy, however, it is crucial to first test its performance in domains whose genealogy is already known. Furthermore, it is important to test PCM's ability to reconstruct chronologically deep phylogenies using exclusively modern language data, often the only available data outside Eurasia.

This article describes some experiments performed on a selection of 26 contemporary Indo-European varieties belonging to the Romance, Greek, Germanic, Celtic, Slavic, Indic and Iranian families. With respect to previous PCM experiments (cf. Longobardi & Guardiano 2009), the present article relies on:

- a. Nine additional contemporary IE languages analyzed parametrically: Sicilian, (Northern) Calabrese, Bovesse Greek, Danish, Icelandic, Slovenian, Polish, Farsi, and Marathi.
- b. A much more refined grid of nominal parameters. The list and a brief description (the first of its kind) of the 56 parameters used, defining the variability of a single module (the internal structure of nominal phrases), is contained in the electronic support material (see the Appendix).
- c. A wider range of statistical procedures (UPGMA trees, SplitsTree package, distribution and correlation tests, Structure, PAUP*).

The experiments produced two major results:

1. The current version of PCM identifies the main subfamilies of Indo-European strikingly well, even from modern languages only, performing genealogically as accurately as lexical methods.
2. ‘Horizontal’ transmission (interference) does not seem to limit the effectiveness of PCM to the extent of seriously undermining the correct representation of the main ‘vertical’ relations.

We contend, thus, that generative grammar, and more generally the bio-cognitive framework of which it is a salient part (Roberts 2012), is not only an insightful theory of mind (Chomsky 1975) and of its synchronic (Chomsky 1981 and following work) and diachronic (Lightfoot 1979) variability: it can also become a true historical science, capable of gaining insights into the actual (pre)history of human populations, no less than the successful historical-comparative enterprise of the 19th century. In turn, such historical success of generative parameters brings about new qualitative evidence that parameters are really at work in language transmission through time, hence, ultimately, in the actual processes of language acquisition.

This article is organized as follows: Section 2 provides an outline of the recent debate about the reconstruction of genealogical relationships within Indo-European on a quantitative/statistical basis; Section 3 discusses a class of experiments performed over the novel database of syntactic (parametric) data with the aid of a selection of distance-based taxonomic algorithms of biostatistic derivation, and compares the results with those provided by lexical taxonomic characters (i.e. the cognacy judgments provided in Dyen et al.’s 1992 database); Section 4 proposes two preliminary tests with character-based algorithms, and provides comments about the role of plausible interference effects (i.e. contact) on the taxonomies produced; Section 5 lists the relevant conclusions.

2. Indo-European classifications on a quantitative basis

2.1 Quantitative taxonomies

Many subfamilies of Indo-European, often identified long before the establishment of the classical comparative method, are undisputed. The hypotheses of intermediate units (aggregations across subgroups, position of isolates) are more debated, such as, for instance, the Balto-Slavic hypothesis (Schleicher 1861–1862, Szemerényi 1957, *contra* Meillet 1905, 1924) or the Italo-Celtic one (Meillet 1922, also cf. recently, e.g., Ringe et al. 2002). Even the model of Indo-Iranian as a secondary proto-language simply diverging into Indic and Iranian (and Nuristani),

reconstructed mainly from their earliest literary stages of the former two (Meillet 1922), has been occasionally called into question (Lazzeroni 1968).

In recent years, all such questions are being readdressed with the aid of statistical and computational methods. This quantification of relatedness across IE languages has still been mostly based on lexical datasets: the most frequently adopted is Dyen et al.'s (1992) 'Comparative Indo-European data corpus', where a 200-item Swadesh list (Swadesh 1950) is translated into 95 Indo-European languages, with (idealized) binary cognacy judgments provided for all pairs of synonyms.¹

Phylogenetic computational procedures derive mainly from those of molecular biologists and, according to the nature of the input, fall into two main types: distance-based and character-based. Character-based algorithms deal with matrices of finite-state characters, and the chronological distance of each single taxonomic unit from the root is calculated without assuming any uniform rate of evolution across the tree. Conversely, distance-based algorithms are more flexible with respect to the nature of the input (i.e., they can work with different typologies of data, *a priori* uniformed by the distance metrics), but are subject to rigid restrictions on the form of the output; for instance, most of them assume a constant rate of evolution across the tree, which might produce misleading effects on the topologies (cf. Section 3.3 below).

An extensive list of effective linguistic implementations of algorithms from the classic PHYLIP package (*Neighbour*, *Fitch* and *Kitsch*: Felsenstein 1993), all empirically relying on Dyen et al. (1992), is reported in McMahon & McMahon (2003, 2005), with the purpose of singling out those most appropriate for linguistic data; they suggest that distance-based algorithms are more viable than character-based ones. In the same vein, Nakhleh et al. (2005) discuss various experiments on six reconstruction methods (both character-based and distance-based) and on four different versions of an Indo-European lexical database (for further discussion cf. at least Lohr 1998, van Cort 2001, Rigon 2009, 2012, Barbañon et al. 2013).

In contrast, some other scholars have used algorithms directly, taking linguistic characters as input, e.g. works by Gray, Atkinson and their collaborators, and Ringe et al. (2002). Gray & Atkinson's (2003) input is based on Dyen et al.'s (1992) dataset; their experiments are more explicitly aimed at dating splits and testing the two most disputed hypotheses on the origin of Indo-Europeans, the 'Kurgan expansion' (Gimbutas 1973) and 'Anatolia farming' (Renfrew's 1987 'Neolithic Discontinuity Theory'). Ringe et al.'s (2002) work is notable because it develops a phylogenetic algorithm specifically conceived for the analysis of language

1. Two main versions of Swadesh's list of supposedly 'basic' (i.e. universal and culturally independent) meanings are available, with 100 and 200 items, respectively. For a detailed discussion of the impact of the list length in refining relatedness hypotheses, see Embleton (1986).

relationships, and stresses the practical difficulties in assigning discrete values to lexical correspondences. They adopt taxonomic characters of three types: root correspondences (2002: 333), sound correspondences (2002: 22) and correspondences in inflectional endings (2002: 15), all encoded as binary. Their experiments across 24 Indo-European ancient and modern languages suggest the possibility of constructing a stable evolutionary tree for Indo-European; their conclusions sustain the Italo–Celtic hypothesis with “respectable support” (Ringe et al. 2002: 112), and Greek–Armenian unity, though with “significantly poorer” evidence (Ringe et al. 2002: 102); no counterevidence is prompted against Balto-Slavic and Indo-Iranian.

2.2 Beyond classical trees

The methods described above are mainly used to produce evolutionary trees, essentially still in the tradition of Schleicher’s (1861–1862) *Stammbaum*: they often assume a root, and always one single possible path between each pair of leaves. As is known, such classical graphs only encode genealogical (i.e. **vertical**) transmission and cannot represent secondary convergence (i.e. **horizontal** transmission), to which lexical characteristics are particularly sensitive (Thomason & Kaufman 1988, Thomason 2001, 2004, 2010, among others). In biology, network graphs have been elaborated precisely to deal with “cases where a particular genetic sequence has more than one possible history” (McMahon & McMahon 2005: 140); networks collapse different trees into a single representation of all possible transmission paths. When different possible branchings are in conflict, a tree must choose one, while a network provides a reticulation, i.e. it depicts all the possible branches together. Thus, networks represent historical relations without favoring vertical over horizontal transmission. For this reason, network programs have been advocated in linguistics as well (cf., among many others, McMahon & McMahon 2005, McMahon et al. 2005, McMahon 2010, Noonan 2010, Nelson-Sathi et al. 2010), under the traditional (essentially since Schmidt’s 1872 *Wellentheorie*) idea that “the diversification of the Indo-European family must be modeled at least in part as a network rather than a tree” (Ringe et al. 2002: 110).

3. Indo-European classification on a syntactic basis

3.1 The rise of PCM

The computational attempts to analyze relations within Indo-European have been encouraging, though hardly uncontroversial. One difficulty, perhaps, is that

the programs adopted are only partially adequate to deal with language history (Nakhleh et al. 2005); but another may be that lexical characteristics are not unconditionally suitable for effective measurement, because many lexical correspondences are only arbitrarily reducible to discrete units. Longobardi & Guardiano (2009) list and briefly exemplify four subcases, in which relevant affinities between lexical items cannot be easily reduced to binary, or perhaps even discretely computable correspondences (partial correspondence of form, partial correspondence of meaning, correspondence of form with no correspondence of meaning, identical peculiar semantic shifts in the same cultural area with no correspondence of form). Many such difficulties are well represented in the attempt to compute relative distances among IE languages from, say, four unquestionable cognates like German *gieße* 'I pour', Italian *fondo* 'I melt', Sanskrit *juhomi* 'I pour', and Ancient Greek *χέω* 'I pour', which all variously differ from each other regarding reduplication, *Ablaut* grade, suffix, infixation, ending, or meaning.

Using sound laws (still a broad-sense lexical device, in our terminology, since they are based on the arbitrary distribution of sounds across the vocabulary) circumvents some of these problems, but others may arise. For instance, if a sound law has different subclauses (e.g. the three shifts of Grimm's law), how many identities/differences should be counted between two languages with respect to it? The problem is especially acute for, e.g., microvariation in the German second consonant shift, which is dialectally quite non-uniform.

In any case, some non-trivial arbitrary manipulation seems inevitable to turn lexical data into input characters for precise quantitative experiments. Therefore, the idea of considering taxonomic characters beyond the vocabulary has often been advocated (Nichols 1992, Longobardi 2003, Heggarty 2004, Dunn et al. 2005, among others). Longobardi (2003), developing some ideas from Roberts (1998), suggested that syntactic parameters are particularly suited to function as comparanda for phylogenetic reconstruction, because they require little arbitrary idealization, since they are endowed by definition with formal properties (discreteness and universality) that guarantee longest-range comparison, quantitative exactness, and no ambiguity of settings (Chomsky 1981, Lightfoot 1991, Baker 1996, Kayne 2000, Biberauer 2008, in a vast literature).

While parameter values can in principle measure syntactic distances within any set of languages, no single binary parameter can, of course, prove kinship between two languages; however, since each parametric comparison yields a clear-cut answer, one can calculate probabilistic thresholds (Bortolussi et al. 2011). Longobardi (2003) also suggested, as a strategy of realistic size, to study relatively many parameters in relatively many languages, at the acceptable cost of focusing on one compact module of grammar (Modularized Global Parametrization, MGP).

Among its advantages, MGP allows for a better identification of cross-parametric implications and for a probabilistically sounder sample/population ratio.

3.2 The parametric database

An updated database has been used in the parametric experiments presented in this article. It consists of 56 binary parameters, all, in agreement with MGP, defining properties of nominal structures, identified on the basis of the existing literature (Longobardi 1994, 2001, 2005, 2008, Plank 2003, Alexiadou et al. 2007, Ghomeshi et al. 2009, Keenan & Paperno 2012, among many others) and ongoing investigation. A salient feature of this type of datasets, stressed in Longobardi & Guardiano (2009), is that in many cases the value of a parameter is entirely predictable from those of other parameters (also cf. Baker 2001), and thus represents information to be completely disregarded for taxonomic purposes. Figure 1 presents the parameters and their settings, in the format introduced in Longobardi & Guardiano (2009). The alternative states are encoded as '+' and '-', while the neutralized states resulting from implications across parameters are coded as '0'; the conditions determining 0 for a parameter are indicated next to the name of that parameter in the second column. A few empirically uncertain states, indicated by '?', are also treated as 0s for the purposes of taxonomic computations. A brief description of the 56 nominal parameters, as well as the illustration of their implications, is available in the electronic support material (see the Appendix).

The phylogenetic experiments have been performed on two distinct sets of Indo-European languages. One includes 26 contemporary varieties (Sicilian: Sic, Northern Calabrese: Cal, Italian: It, Salentino: Sal, Spanish: Sp, French: Fr, Portuguese: Ptg, Rumanian: Rm, Bovesse Greek of Southern Calabria: BoG, Grieco, i.e. Greek of Salento: Gri, standard Greek: Grk, English: E, German: D, Danish: Da, Icelandic: Ice, Norwegian: Nor, Bulgarian: Blg, Serbo-Croatian: SC, Slovenian: Slo, Polish: Po, Russian: Rus, Irish: Ir, Welsh: Wel, Farsi: Far, Marathi: Ma, Hindi: Hi). The other consists of the subset of 21 such varieties overlapping with those of Dyen et al.'s (1992) database (which does not include the five Greek and Romance dialectal varieties of Southern Italy).

3.3 Phylogenetic algorithms

Among phylogenetic algorithms, character-based ones, though more precisely representing language history (Rigon 2009, Barbançon et al. to appear), are less suitable in our context for two reasons. First, the implications among parameters, very pervasive within the same module, violate their crucial assumption of character independence. Second, character-based algorithms produce more informative

TABLE A		Sic	Cal	It	Sp	Fr	Pq	Rm	Bog	Gk	E	D	Da	Ice	Nor	Big	SC	Stb	Pro	Rus	Ir	Wel	Far	Me	H
1	± gramm. person	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1
2	± gramm. number	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	2
3	± gramm. gender	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	3
4	± NP over D	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	4
5	± feature spread to N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	5
6	± nouns, on N (RM)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	6
7	± gramm. initial def	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	7
8	± gramm. initial def	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	8
9	± strong person	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	9
10	± free null partitive Q	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	10
11	± gramm. dist. art.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	11
12	± def-checking N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	12
13	± def-checking N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	13
14	± def spread to N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	14
15	± D-controlled inf. on N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	15
16	± plural spread from cardinals	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	16
17	± gramm. boundedness	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	17
18	± strong article	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	18
19	± bounded-checking N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	19
20	± null-N-licensing art	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	20
21	± structured APs	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	21
22	± feature spread to struct. APs	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	22
23	± feature spread to pred. APs	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	23
24	± D-controlled inf. on A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	24
25	± DP over relatives	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	25
26	± relative extrap.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	26
27	± free reduced rel	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	27
28	± N-raising with obl. pied-piping	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	28
29	± free Gen	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	29
30	± uniform Gen	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	30
31	± DP over free Gen	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	31
32	± GenO	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	32
33	± Gen-feature spread to N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	33
34	± D checking poss.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	34
35	± adjectival poss.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	35
36	± post-affix poss.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	36
37	± diffc. poss.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	37
38	± N-feat. spr. to pron. poss.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	38
39	± N-feature spread to free Gen	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	39
40	± adjectival Gen	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	40
41	± Poss-checking N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	41
42	± Strong partial Locality	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	42
43	± Strong Locality	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	43
44	± D Checking Dem	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	44
45	± N over cardinals	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	45
46	± N over ordinals	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	46
47	± N over M1 As	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	47
48	± N over M2 As	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	48
49	± N over As	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	49
50	± N over GenO	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	50
51	± N over ext. arg.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	51
52	± free MOD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	52
53	± class MOD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	53
54	± def on APs	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	54
55	± gramm. AP marker	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	55
56	± gramm. AP marker	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	56

Figure 1. Table A (56 nominal parameters and their settings in 26 contemporary IE languages)

results when the input is supplemented with additional information, especially regarding the directionality of change (e.g. the markedness theory assumed in Ringe's experiments) or the characters' stability (i.e. the possibility of assigning each character a weight, according to how resistant it appears to change). At the

[illegible]

Figure 2. Matrix of parametric (syntactic) distances

present stage, we cannot yet rely on a solid theory of parameter resetting, therefore no such information was included in the input. Thus, distance-based methods appear more readily applicable: they minimize the effects of internal dependencies across parameters, and allow phylogenetic hypotheses even in the absence of a refined theory of diachronic change.

Following Longobardi & Guardiano (2009), the number of parametric identities and differences for each pair of languages was first represented as an ordered pair of non-negative numbers $\langle i; d \rangle$ and then normalized to a monadic figure, the **Jaccard-Tanimoto** distance ($d/(d+i)$), to neutralize most drawbacks of the cross-parametric dependencies. Hence, two identically set languages will have distance 0, two with all opposite settings distance 1, all other cases falling in between. Such distances are represented in Figure 2.

The algorithms that turned out to be most effective with parametric data (Rigon 2009, Longobardi & Guardiano 2009), i.e. UPGMA (Sokal & Michener 1958) and *Kitsch* (from the PHYLIP package, Felsenstein 1993), have the further property of producing ‘ultrametric trees’, i.e. trees complying with the ‘molecular clock hypothesis’ (Bromham & Penny 2003, Felsenstein 1993, 2004): in the rooted trees produced, the length of each branch, that is the distance between the root and every single leaf, is uniform. This requires all the languages to be chronologically equidistant from the root; the method treats all taxonomic units as leaves, i.e. it does not place any language on non-terminal (ancestral) nodes; a tree from languages of different chronological attestation is only possible under the risky idealization that they can be all viewed as coeval. For the purpose of argument rigor, this idealization has not been pursued here; therefore, the five ancient varieties of Longobardi & Guardiano’s (2009) database (i.e., Latin, Classical Greek, New Testament Greek, Gothic and Old English) have not been included in the present experiments.

3.4 Distance-based trees and networks

The rooted trees in Figures 3–4 have been obtained from the syntactic distances through UPGMA and *Kitsch*, after bootstrapping the input with 1,000 resamples. Bootstrapping is a resampling procedure aimed at testing the solidity of the classification (the robustness of each single branching and of the tree as a whole): the data are resampled sequentially, i.e. a given percentage of the characters is removed and replaced by others randomly chosen within the same corpus, so that new matrices are produced and in turn used as input for further tree-experiments.

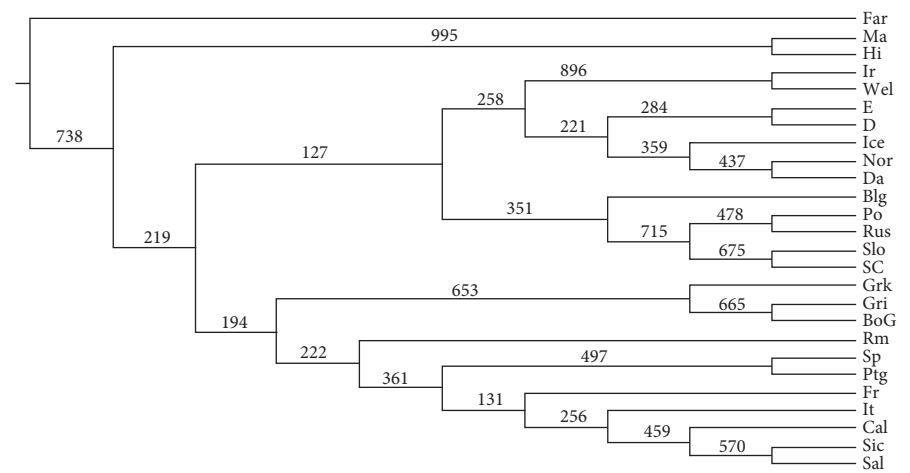


Figure 3. UPGMA (bootstrapped) tree from the syntactic distances in Figure 2

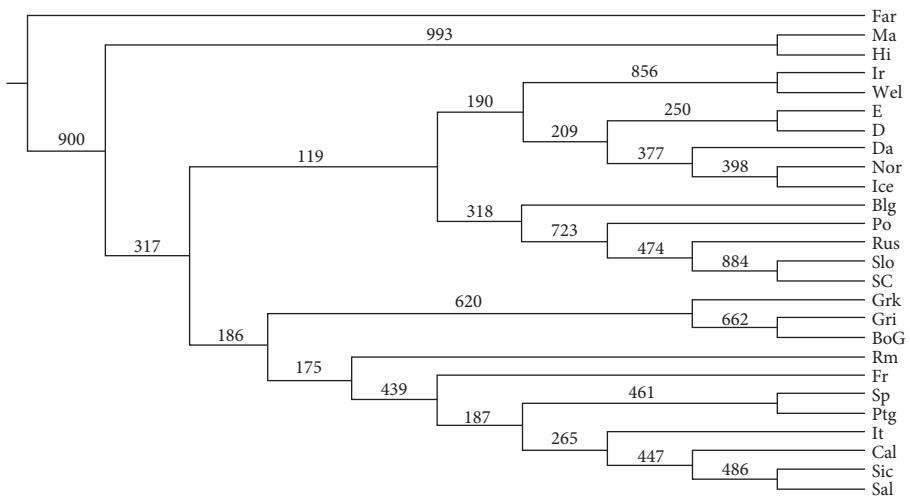


Figure 4. *Kitsch* (bootstrapped) tree from the syntactic distances in Figure 2

The procedure can be repeated thousands of times, according to the size of the database. The final result is a tree where each non-terminal node is labeled with the number of the iterations supporting that particular branching.

The corresponding networks (Figures 5–6) have been generated from the algorithms *ConsensusTree* and *Equalangle* (contained in the ‘SplitsTree’ package, Huson & Bryant 2006), using as an input the same 1,000 trees resampled through UPGMA and *Kitsch*, respectively.

The trees, as well as the networks, are remarkably consistent with each other, and with the suggestions of traditional scholarship.

The Romance languages have been grouped together in all experiments, though bootstrapping values in the trees are lower than 500 for most of the nodes.² Rumanian always appears as the outlier; Iberian unity is well-represented, and the Italian varieties are appropriately placed, with the three Southern dialects

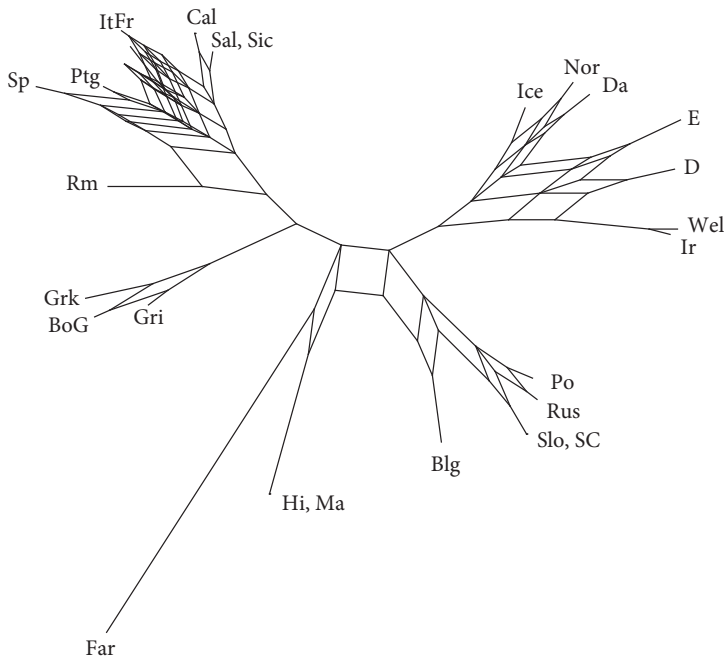


Figure 5. Network from 1,000 trees resampled by UPGMA

2. Bootstrapping values have turned out slightly lower than one might wish more generally (eight nodes exceed the threshold of 500 in UPGMA, seven in *Kitsch*). However this is certainly influenced by the relatively small number of relevant taxonomic characters. Furthermore, the bootstrapping procedure is not completely appropriate for treating parametric data, again because of the pervasive implicational structure, which is not taken into account in the perturbation of the matrix.

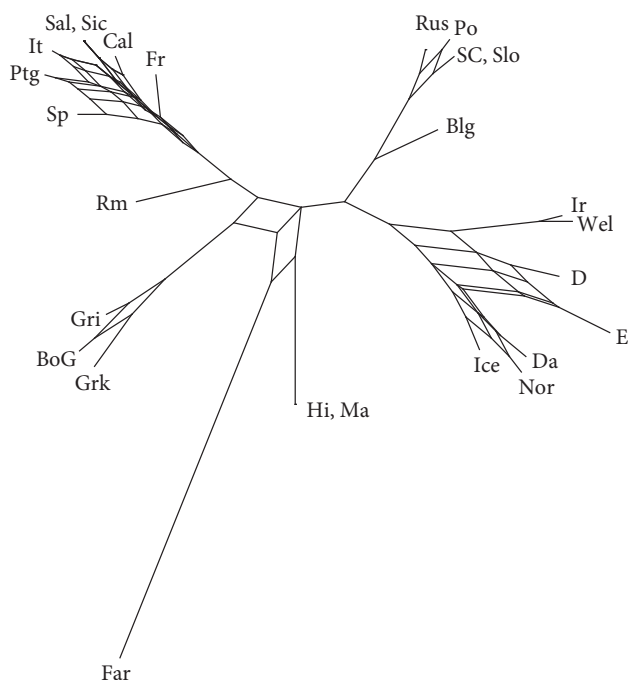


Figure 6. Network from 1,000 trees resampled by *Kitsch*

(Northern Calabrese, Sicilian and Salentino) clustered together, and their internal articulation reflecting the traditional separation (Pellegrini 1977) between ‘Extreme’ Southern dialects (here Salentino and Sicilian) and ‘Upper’ Southern ones (Northern Calabrese). Only the position of French is slightly variable: under several replications, in the trees it shows up as an outlier of the Italian group, of the Iberian cluster, or even of a larger cluster including both.³ Rumanian is never attracted towards any other group in particular; its isolation, and its relations with the other languages of the area, may have produced some divergence from the core of the ancestral family, but actual convergence with the surrounding Balkan languages is limited (cf. Section 4.3 below).⁴

The three Greek varieties (standard Greek, Grieco and Bovesese Greek) are included under one and the same node, with significant bootstrapping values (653 UPGMA, 620 *Kitsch*). The two of Southern Italy are clustered apart (665 UPGMA,

3. For the relations between Gallo-Romance and other subgroups, cf. at least Tagliavini (1972), Harris & Vincent (1988), Renzi & Salvi (1994), Lindenbauer et al. (1995), and Holtus et al. (1988–2005).

4. Divergence is one of the typical effects of the contact relations across the Balkan languages pointed out in Hock & Joseph (1996), among others.

662 *Kitsch*). Thus, the very obvious contact between Romance and Greek in Southern Italy (Guardiano & Stavrou to appear) is not overrepresented: the two Greek dialects are never misclustered with any of the Romance languages rather than with homeland Greek.

Slavic unity is also well recognized in trees and networks. One disparity between PCM and traditional evidence concerns its internal articulation, namely, only two of the three Southern varieties, i.e. Serbo-Croatian and Slovenian, are clustered together (with high bootstrapping values in both trees: 675 UPGMA, 884 *Kitsch*), while Bulgarian systematically shows up as the outlier of the whole family. This is a taxonomic **factual mistake** (Guardiano & Longobardi 2005: 166): given three languages X, Y and Z, X turns out to be closer to Y than to Z (or vice versa), while the opposite is true according to independently well-established taxonomies. The cluster of all the Slavic languages, excluding Bulgarian, is supported by a significant bootstrapping index (715 UPGMA, 723 *Kitsch*); on the contrary, the node including Bulgarian is associated with smaller values. Such a position of Bulgarian is also replicated in the two networks. Here, all the Slavic varieties are grouped together, while Bulgarian appears on a separate branch. Yet, precisely as with Rumanian, no evidence of systematic attraction from one external group is visible; indeed, no particular branch (or reticulate of branches) connects Bulgarian to a non-Slavic language. Actually, Bulgarian contrasts with the rest of Slavic in three parameters. In p7 (DGP) and p29 (GFR), Bulgarian exhibits a '+' (like many non-Slavic languages of the sample), while the other Slavic languages homogeneously select '-'.^{5,6} In p32 (GFO), Bulgarian exhibits a '-', while the rest of Slavic homogeneously selects '+'. Precisely two of the mentioned parameters (p7, DGP and p29, GFR) are responsible for intricate cascades of implications, so that their values in the Slavic languages other than Bulgarian trigger a high number of zeros. Now, as noted, the distance measure adopted here successfully normalizes the zeros in the assessment of similarities/differences, but cannot avoid their reducing the potential for comparison and the overall balance of differences and identities. In any case, contrary to some recent literature (for recent contributions, see at least Joseph 1999, Thomason 2000, Tomić 2003), which suggests that the grammar is

5. Exceptions: Farsi, Hindi and Marathi are -p7 (DGP); Rumanian, Bovese Greek, Grico and Greek are -p29 (GFR).

6. For proposals regarding the properties associated with the absence of grammaticalized definiteness in the Slavic languages, cf. Bošković (2008a, 2008b, 2010). Even if certain interesting hypotheses of such works turned out to be correct (e.g., if no language without grammaticalized definite articles could have prepositional genitives), it would not be the two parameters p7 (DGP) and p29 (GFR) that would be implicationally related in the current formulation of Figure 1, so that this would not reduce the distance between Bulgarian and other Slavic languages.

even more sensitive to contact than the lexicon within the Balkan *Sprachbund*, PCM misplaces none of the languages of the area outside its genealogical family.

The two Celtic languages are systematically clustered together, with remarkably high bootstrapping values (896 UPGMA, 856 *Kitsch*).

The Germanic group is recognized as well, and its internal articulation into Northern and Western varieties is acknowledged, though less significantly supported by bootstrapping values.

Celtic and Germanic are then clustered together in the two trees: even though such a higher cluster is not supported by significant bootstrapping values, it recurs in the experiments performed. The two groups also appear in one and the same region of the networks, which (especially the one produced through UPGMA) confirm the existence of the two families but also their closeness, particularly affecting the Western branch of Germanic. Although between English and Welsh convergence effects have unquestionably existed (Willis 2008), within Figure 1 only the value of parameter 23 (FSP) seems identical in Celtic and West Germanic to the exclusion of their neighbors or kin (it is only further shared with remote Farsi) and thus may represent a common secondary innovation; for the rest, the whole extent of Celto-Germanic similarity is not obviously traceable to particular points of our dataset. No experiments, on the contrary, provide any evidence for an Italo-Celtic unit.

The most interesting point concerns the Indic and Iranian languages. In the trees, the two Indic languages, displaying identical nominal syntax, correctly cluster under the same node, with the highest bootstrapping values (995 UPGMA, 993 *Kitsch*). Farsi, instead, is the outlier of the whole tree in both topologies, a position corroborated by high values of bootstrapping, especially in the tree produced by *Kitsch* (900; 738 UPGMA). Thus, no Indo-Iranian unity is recognized when a strict genealogical model is assumed. Their relative proximity, however, shows up moderately in the networks, where the three varieties are grouped in one separate section. Here, Farsi is on a branch longer than any other of the network, thus suggesting a higher number of mutations, in principle due either to isolation, which is historically implausible, or to mixing with external varieties. More forcefully, the relation between Farsi and Indic is suggested by the particular additional experiment attempted in Section 4.1. below.

However, it is most remarkable that the failure of the trees in perceiving Indo-Iranian as a genealogical unit is exactly shared with Dyen et al.'s (1992) lexicostatistical experiment. In their results, while the highest percentage of cognates exhibited by Iranian is indeed with the Indic cluster (18.1), Indic exhibits a higher percentage of cognates with the so-called 'Mesoeuropeic' set, i.e. Central European IE languages (18.6): as a consequence, Indo-Iranian fails their subgrouping meth-

od, which would require a cognacy percentage of 21.1 ($= 18.6 + 2.5$).⁷ Second, the percentage of cognates between an Indo-Aryan and an Iranian language ranges from 9.4 (Singhalese–Afghan) to 25.4 (Panjabi–Baluchi), indeed suggesting some relevance of areal closeness/distance.⁸ For Dyen et al., a range of 16 percentage points of internal variability is too high to characterize a well-defined group. They justify the discrepancy between their lexicostatistical classification and the traditional Indo-Iranian hypothesis with the self-imposed restriction to contemporary word lists: they suppose that the method, if applied to older stages of Indic (i.e. Sanskrit) and Iranian (i.e. Avestan, Old Persian) evidence, might perhaps identify an Indo-Iranian group. Meillet's (1922: 25–30) and others' conclusion that Indo-Aryan and Iranian continue a specific proto-language is based precisely on evidence from Old Persian, Avestan, and Sanskrit. Thus, the failure of the lexicostatistical (as well as the syntactic) experiments in confirming Indo-Iranian may result from the loss of evidence in the modern languages: "this erosion is conceivably great enough that contemporary languages fail clearly to reveal the group, whether the data used are lexicostatistical or those of common innovations" (Dyen et al. 1992: 48).

Speaking of secondary interference effects, the percentage of cognates has been considerably deflated by the massive borrowing into Iranian of lexical items coming from languages belonging to non-native populations that, since the first millennium AD, have dominated in the area, e.g. Arabic.

In syntactic terms, Farsi exhibits at least four cases of parameter values which are not shared with any other Indo-European variety of our sample, i.e. p3 (FGG), p19 (CCN), p46 (NOO) and p55 (AMO), plus two further parameters (p16, CPS, and p22, FFS), whose value (–), within Indo-European, is shared with Welsh (p16, CPS) and English (p22, FFS) only. Of these parameters, the values of p3 (FGG) and p16 (CPS) are shared by Uralic and Altaic languages: this could be taken as a marker of Central Asian influence. Finally, p46 (NOO) affects the position of nouns with respect to adjectives, and its value is partly shared with Semitic languages, as the possible correspondent to Arabic loanwords in the lexicon. However, only an analysis of more Iranian and Indic languages could sway syntactic evidence either in favor or against skepticism about Indo-Iranian.

7. In their **subgrouping method**, 2.5 is a safety coefficient called **critical difference** (c.f. Dyen et al. 1992: 26).

8. By itself, this kind of evidence does not help one choose between the hypothesis of secondary contact as disrupting a primeval genealogical relatedness or as overstating an original looser kinship of the two groups. The hypothesis that many isoglosses shared by ancient Indic and Iranian literary varieties derive in fact from cultural sharing in contiguous areas was suggested by Lazzeroni (1968, 1998), after a detailed exploration of their exact chronology.

3.5 Syntax and lexicon

The parallelism of syntactic and lexical analyses in a critical case (the Indo-Iranian question) suggests the opportunity of pursuing such comparison more generally, refining an enterprise inaugurated in Longobardi & Guardiano (2009).

With Dyen et al.'s (1992) database, distances can be computed as the complement of the percentage of cognates shared between any two languages in a pair. By way of illustration, the percentage of cognate forms between Italian and French amounts to 80.3%; thus, if measured on a scale ranging from 0 to 100, their distance turns out to be 19.7 (0.197 on a scale ranging from 0 to 1). As mentioned, 21 languages of our sample overlap with those of Dyen et al. (which does not include the five Greek and Romance dialectal varieties of Southern Italy), allowing for comparisons.

First, the two distance matrices (Figure 7) reveal that syntactic distances are significantly smaller than lexical ones. For most pairs, the ratio between the two, attended around 3:4, is considerably smaller, varying between 1:3 and 1:4 (mean = 1:3.6).⁹ Such a difference hints at the formal demonstration that the com-

SYNTACTIC DISTANCES																					
	It	Sp	Fr	Ptg	Rm	Grk	E	D	Da	Ice	Nor	Blg	SC	Slo	Po	Rus	Ir	Wei	Far	Ma	Hi
It	0	0.0682	0.0476	0.0233	0.103	0.225	0.15	0.125	0.128	0.2	0.128	0.15	0.233	0.233	0.194	0.226	0.211	0.243	0.433	0.2	0.2
Sp	0.212	0	0.0714	0.0233	0.103	0.25	0.175	0.15	0.154	0.2	0.154	0.2	0.267	0.267	0.226	0.258	0.184	0.216	0.467	0.233	0.233
Fr	0.197	0.266	0	0.0488	0.162	0.289	0.184	0.158	0.162	0.184	0.162	0.184	0.276	0.276	0.233	0.267	0.194	0.229	0.483	0.241	0.241
Ptg	0.227	0.126	0.291	0	0.105	0.256	0.15	0.125	0.128	0.179	0.128	0.154	0.233	0.233	0.194	0.226	0.189	0.222	0.433	0.2	0.2
Rm	0.34	0.406	0.421	0.371	0	0.2	0.211	0.25	0.162	0.205	0.132	0.205	0.241	0.241	0.233	0.233	0.257	0.286	0.5	0.214	0.214
Grk	0.822	0.833	0.843	0.833	0.843	0	0.3	0.256	0.263	0.293	0.282	0.25	0.188	0.188	0.182	0.182	0.257	0.286	0.393	0.267	0.267
E	0.753	0.76	0.764	0.76	0.773	0.838	0	0.0732	0.0714	0.122	0.0714	0.19	0.219	0.219	0.242	0.242	0.111	0.139	0.3	0.212	0.212
D	0.735	0.747	0.756	0.753	0.751	0.812	0.422	0	0.093	0.0909	0.093	0.167	0.147	0.147	0.147	0.176	0.111	0.118	0.31	0.242	0.242
Da	0.737	0.75	0.759	0.75	0.763	0.817	0.407	0.293	0	0.0889	0.0435	0.14	0.156	0.156	0.188	0.188	0.167	0.176	0.345	0.176	0.176
Ice	0.755	0.763	0.772	0.763	0.777	0.802	0.454	0.409	0.221	0	0.0435	0.133	0.118	0.118	0.147	0.147	0.132	0.139	0.3	0.176	0.176
Nor	0.754	0.761	0.77	0.761	0.786	0.821	0.452	0.367	0.146	0.191	0	0.136	0.156	0.156	0.188	0.188	0.167	0.176	0.345	0.176	0.176
Blg	0.769	0.782	0.791	0.781	0.798	0.811	0.772	0.769	0.76	0.775	0.773	0	0.0882	0.0882	0.147	0.118	0.222	0.25	0.333	0.147	0.147
SC	0.755	0.768	0.772	0.766	0.778	0.821	0.766	0.764	0.749	0.768	0.772	0.291	0	0	0.0571	0.0286	0.222	0.259	0.321	0.188	0.188
Slo	0.76	0.772	0.782	0.781	0.79	0.821	0.751	0.733	0.733	0.763	0.762	0.385	0.316	0	0.0571	0.0286	0.222	0.259	0.321	0.188	0.188
Po	0.764	0.772	0.781	0.776	0.784	0.837	0.761	0.754	0.749	0.758	0.762	0.369	0.32	0.367	0	0.0278	0.214	0.25	0.345	0.219	0.219
Rus	0.761	0.769	0.778	0.773	0.781	0.832	0.758	0.755	0.74	0.754	0.758	0.365	0.325	0.386	0.266	0	0.214	0.25	0.345	0.219	0.219
Ir	0.8	0.805	0.812	0.817	0.837	0.859	0.817	0.806	0.817	0.755	0.836	0.818	0.796	0.809	0.8	0.782	0	0.0263	0.31	0.31	0.31
Wei	0.793	0.813	0.81	0.804	0.812	0.867	0.841	0.82	0.825	0.82	0.849	0.838	0.821	0.838	0.822	0.818	0.645	0	0.276	0.321	0.321
Far	0.859	0.86	0.859	0.854	0.864	0.875	0.86	0.86	0.86	0.862	0.874	0.853	0.842	0.849	0.849	0.844	0.888	0.899	0	0.345	0.345
Ma	0.833	0.839	0.839	0.834	0.838	0.869	0.855	0.849	0.855	0.856	0.859	0.811	0.822	0.817	0.827	0.817	0.879	0.893	0.834	0	0
Hi	0.818	0.819	0.824	0.813	0.827	0.874	0.854	0.853	0.843	0.855	0.852	0.801	0.805	0.8	0.799	0.8	0.878	0.876	0.815	0.436	0
LEXICAL DISTANCES																					

Figure 7. Matrix of syntactic and lexical distances (21 IE languages)

9. Some statistical procedures were implemented in Bortolussi et al. (2011) on a corpus of roughly 10,000 fictitious languages, randomly generated, though according to the constraints imposed by the implicational structure of the parameter list. The experiment revealed that the highest value attended for syntactic distances cannot exceed 0.75 rather than 1 (cf. Guardiano and Longobardi's 2005: 161–2, Anti-Babelic Principle for precisely an expectation of this kind). Distance 1 must instead be attended for the vocabulary, given that two languages with no recognizable common etymology at all do precisely attain this value.

position of the basic vocabulary has changed more rapidly than (at least nominal) syntax in the history of these 21 Indo-European languages.

The distribution of syntactic distances in the matrix can be assimilated to a Normal: this is shown in the histogram in Figure 8. On the contrary, lexical distances (Figure 9) display a Bimodal distribution, with a first peak for the interval 0.3–0.4 and a second at 0.8–0.9. In other words, it seems that lexical distances are able to clearly detect taxonomic signals up to 0.5; on higher values, they tend to saturate, i.e. to become uninformative because they are no longer able to scatter on different degrees of relationship. Notice that all syntactic distances distribute within 0.5 (perhaps in part also as a reflex of the Anti-Babelic principle, cf. fn. 9).

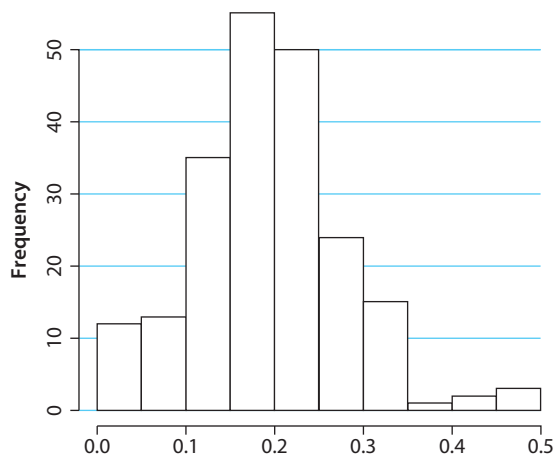


Figure 8. Histogram representing the distribution of the syntactic distances in Figure 7

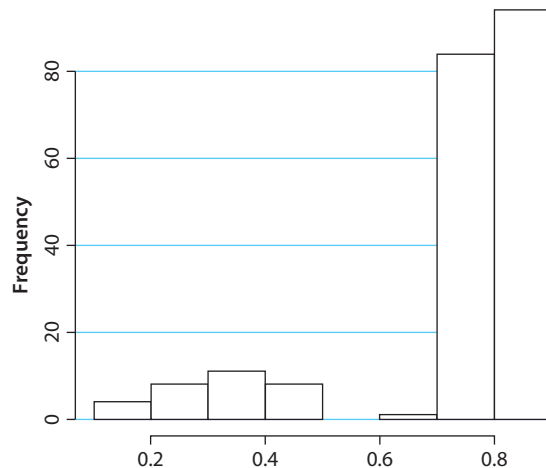


Figure 9. Histogram representing the distribution of the lexical distances in Figure 7

A correlation test between the two distance matrices (Mantel test) shows that lexical and syntactic distances are highly correlated ($r=0.7285$; $p<0.001$; 9,999 permutations), thus proving wrong the hypothesis of orthogonality between lexicon and syntax. However, this does not provide any evidence that their relationship is linear, i.e. that they return fully overlapping information. The proof against the overlapping hypothesis is well visible in Figure 10, where lexical distances are plotted against syntactic ones. The graph neatly identifies two clouds: the bottom-left one shows an essentially proportional increase of the two distances, thus a linear covariation; the top cloud has a more compact distribution: lexical distances show

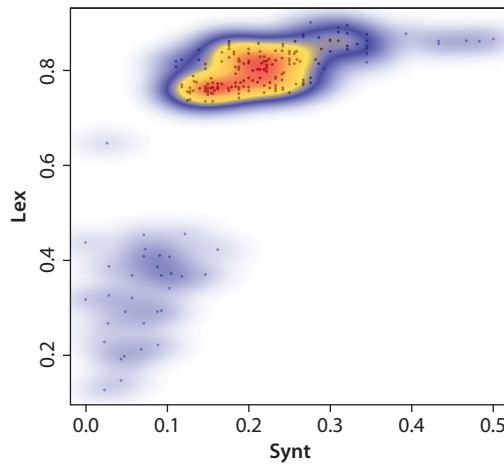


Figure 10. ScatterPlot of the syntactic and lexical distances in Figure 7

little variation, as expected the more they approximate to 1, while syntactic distances remain more scattered. The language pairs corresponding to the dots on the first cloud are all and only those where both languages belong to the same subfamily. This suggests that lexical distances minimize the internal relations within the same subgroup, while maximizing those across distinct subgroups.

In any case, the tree topologies in Figures 11–12, drawn from the lexical and the syntactic distance matrices respectively, largely overlap, with a few low-level differences.¹⁰ Within Romance, the position of French is variable: the lexical tree suggests some unity with Italian, while the syntactic tree separates them. Within Germanic, both trees recognize the Northern group (with Icelandic as the outlier), though the syntactic tree identifies the Western unit as well, not detected in the lexical one. The superordinate node that groups Germanic and Celtic together only

10. Both trees have been computed from a distance-matrix using *Kitsch*. A character matrix of the lexical data is not available; therefore, bootstrapping procedures cannot be implemented here.

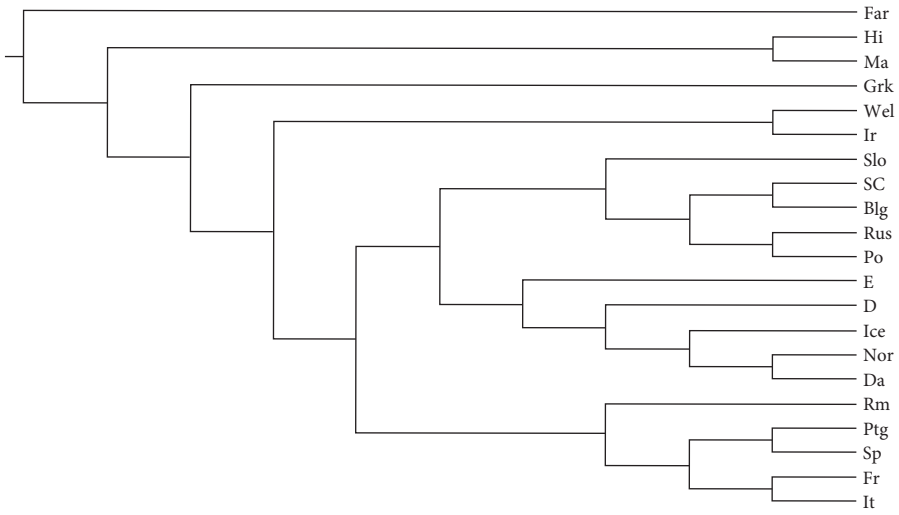


Figure 11. *Kitsch* tree from the lexical distances in Figure 7

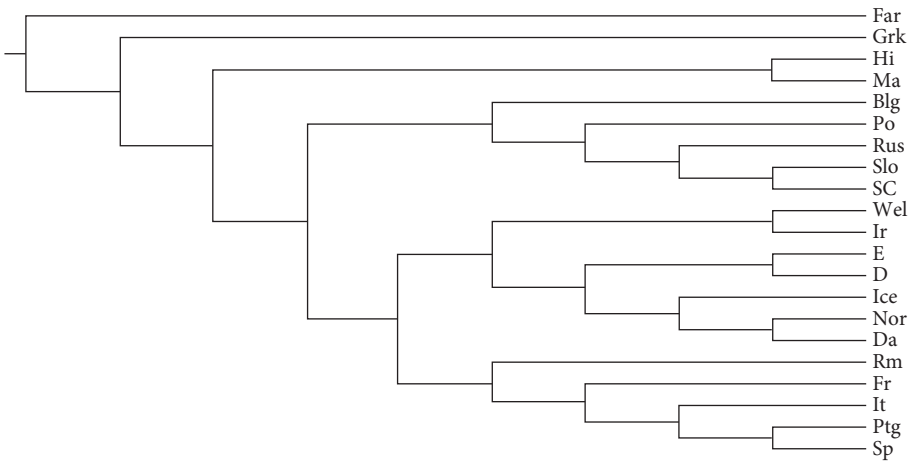


Figure 12. *Kitsch* tree from the syntactic distances in Figure 7

appears in the syntactic tree: the lexical one separates the two groups. Within Slavic, the lexical tree captures some relation between Bulgarian and Serbo-Croatian, but, again, does not properly recognize Southern Slavic unity, as Slovenian shows up as the outlier of the whole group. The position of Russian and Polish, instead, varies slightly in the two trees.

4. Further testing

4.1 Multiple ancestors: A character-based experiment

In the past years, biostatisticians have developed clustering methods that are neither based on phylogenetic models nor assume tree-like representations. These methods are of great interest for molecular anthropologists, because they allow taxonomic classifications not dependent on predefined evolutionary paths. The major difference with tree-like representations is that taxonomic units are treated as sets of elements coming from different ancestors, whose impact can be evaluated separately, rather than as single-fathered entities. One such algorithm, *Structure* (Pritchard et al. 2000), has recently been argued to be appropriate for the purposes of language classification (Reesink et al. 2009). It implements a Bayesian clustering technique that assumes a model with a number (call it *K*) of unspecified or unknown ancestral populations, each characterized by a salient set of properties (e.g. allele frequencies at each DNA locus). As a result, individuals, on the basis of their genotypic profile, are assigned either to one, or jointly to two or more, of the *K* populations; in the latter case, they are considered as the result of admixture between different ancestral populations.

Structure is character-based. As such, it assumes characters to be unlinked to each other, and thus cannot simply use the parametric matrix as an input, because of the noted pervasive interdependencies across parameters. Thus, to explore how syntactic data behave in such a non-phylogenetic model, we manipulated the parameter grid, trying to minimize the impact of the implications and treat parameters as ideally independent: the subset of parameters exhibiting less than ten blanks ('0' or '?') across all languages was singled out.¹¹ In this way, the amount of information suitable for the experiment was inevitably reduced: only 37 out of the 56 parameters could indeed be selected. Notice that the reduction of the comparanda obviously makes taxonomic hypotheses less stable, increasing the probability of mistakes, as already experienced in studies on DNA polymorphisms.¹² Furthermore, our selection was not based on qualitative criteria of genealogical stability, adding to the risk of accidentally overestimating or underestimating

11. For details on the selection cf. the support material in the Appendix.

12. Colonna et al. (2009) show that genetic structures are clearly detectable even within individuals from the same geographic domain (Europe) by using 239 STRs (Short Tandem Repeat) loci, while comparisons become uninformative when the number of loci is reduced to 36. Based on three continental populations (Afro-American, Asian and European), Turakulov & Eastaer (2003) conclude that for the correct assignment of an individual to the continent of origin with a mean accuracy of at least 90%, a minimum of 65–100 Single Nucleotide Polymorphisms (SNPs) are needed, and the same holds for other genetic markers as STRs.

differences and identities. In fact, any attempt to adapt the parameter probe to character-based algorithms involves violating some constraints, and this could undermine the reliability of the results. Ideally, one should be able to devise an appropriate character-based algorithm able to deal with the complex implicational structure of the whole set of parameters, i.e. of truly treating zeros as implied information rather than as missing values. No such situation seems to have arisen with molecular-genetic databases, so that no biostatistical program has so far been designed for these purposes.

To test the basic reliability of the subset, we first built phylogenetic trees from the corresponding distances (Figures 13–14) and checked them against the ones based on all the parameters. Beyond less relevant readjustments in the subarticulation of Romance and Slavic, there are no differences with the trees generated from the full dataset. The trees were generated without bootstrapping, owing to the inevitably small number of replaceable characters in the reduced dataset.

We ran *Structure* for K values from 2 to 8, K=5 being the one with the lowest average log-likelihood value (hence the ‘best’, Figure 15). The results are represented in the form of barplots, each bar corresponding to a language, and colors representing the contribution of each of the supposed K ancestral ‘populations’. The experiments with K=5 were replicated ten times, and all the replications produced the same pattern. Such a result independently attests the statistical solidity of the subset of parameters selected.

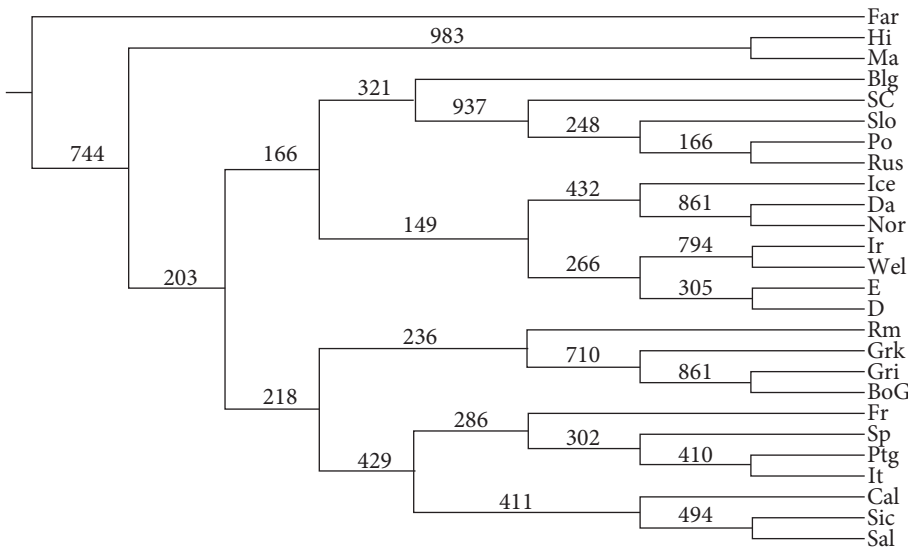


Figure 13. UPGMA tree from the reduced parameter grid (37 parameters)

The red component dominates in the Romance languages, confirming their recent common ancestry; the sole exception is indeed Rumanian, which displays significant manifestations of the green, orange, and blue components. Thus, Rumanian is weakly assigned to the varieties characterized by the red component, but is not distinctly associated to any of the other potential ancestors; this is coherent with the information provided in the networks obtained from the complete dataset.

The three Greek varieties are all overwhelmingly marked by the green component. The red (Romance) component does not show up significantly in the two varieties of Southern Italy, in any case no more there than in Modern Greek; this confirms that contact with Romance in Southern Italy has not obscured their genealogical origin.

The Slavic varieties, with the exception of Bulgarian, show an almost complete absence of contributions from other components. In Bulgarian, besides the common Slavic component (i.e. the yellow one), others are detectable (though, as in Rumanian, not necessarily those prevailing in its closest neighbors), so that here Bulgarian is hardly identified as a member of the Slavic family. However, at least the proportion of the other components in its structure does not allow for more salient alternative hypotheses of grouping.

The expected 'forced' aggregation concerns Germanic and Celtic, in which the same blue component dominates, so that they cannot be plausibly attributed to separate intermediate ancestors. Actually, the closeness of Germanic and Celtic languages was overstated in the networks as well.

The Indic varieties are characterized by the same orange component that appears conspicuously in the bar of Farsi (and rather less saliently also in some other bars). This result is more in line with the topology of the networks. Both experiments suggest some Indo-Iranian affinity, not equally recognized in any of the trees, i.e. when a stricter genealogical model is assumed. One should conclude that, within the sample, the strongest ancestral component is shared by Farsi with the two Indic languages, although this does not necessarily speak for a genealogical relationship, at least from the evidence of modern languages. In any case, the graph of *Structure* confirms that it was not interference from other Indo-European components of our sample that has syntactically detached Farsi from Indo-Aryan. This matches the particular length of the Farsi branch exhibited by the networks.

To summarize, the outcomes of this first implementation of *Structure* on parametric data do not contradict any of the hypotheses which emerged from the corresponding phylogenetic experiments, but also in some cases (the composite structure of Bulgarian and Rumanian) directly suggest plausible effects of admixture.

4.2 Preliminary character-based phylogenies

Once a parameter subset, relatively informative in spite of its limits (low number of variable sites, high amount of missing data, non-independence of many characters), was selected for use with *Structure*, it became tempting to test it with a character-based algorithm of a strictly phylogenetic type. Here, the present lack of directionality and weighting information is an even more serious bias. Indeed we chose a parsimony (the ‘maximum parsimony method’ was first discussed in Edwards & Cavalli-Sforza 1963, cf. also Felsenstein 2004) program, PAUP* (Swofford 2003), that does not require any explicit model of character evolution. Parsimony algorithms search for best trees assuming that the lowest amount of change occurred between the nodes and a common ancestor; their output is a series of trees tied at their best parsimony score (i.e. which share the minimum number of changes).¹³ With lexical data, parsimony-based methods have been claimed to yield the best performances (Barbançon et al. to appear).

As the trees produced are unrooted, a certainly remote outlier language, Wolof (from the West Atlantic family of Niger-Congo), was included in order to root them. After an analysis through a branch-and-bound search, the program was able to produce 59 different trees, which require a total of 50 changes.

Then, a consensus tree (Figure 16) was built through the ‘extended majority rule’.¹⁴ Despite the noted limitations, it is consistent with *Structure*. Farsi is, once more, the outlier of Indo-European. Most of the genealogical subgroupings are correctly recognized, and display high consensus values: the two Celtic languages are always grouped together, like the two Indic ones and the compact group consisting of all Slavic languages but Bulgarian; most of the trees (55 out of 59) acknowledge the Greek group and its internal articulation, as well as Romance unity (53 out of 59) including Rumanian. As for Bulgarian, the composite nature emerged from *Structure* seems to affect its position in the tree: it is not included in the Slavic cluster, although it shows up close to it. Finally, the program expectedly fails to acknowledge the Germanic cluster; the overstated closeness between Celtic and West Germanic, which biases the reduced dataset also with *Structure*, emerges here as well; however, the cluster including Germanic and Celtic displays consensus values consistently lower than the others, thus resulting as the weakest of the whole tree.

13. An anonymous referee correctly notices that phylogenetic classification is traditionally guided by the principle of shared innovations. This principle cannot be implemented in a distance-based method but is automatically embodied in a parsimony system, which is designed precisely to minimize homoplasy.

14. Consensus trees built through an extended majority rule retains **all** best nodes, even if they score lower than 50%. On the contrary, a majority rule retain **only** branchings with a consensus score higher than 50%.

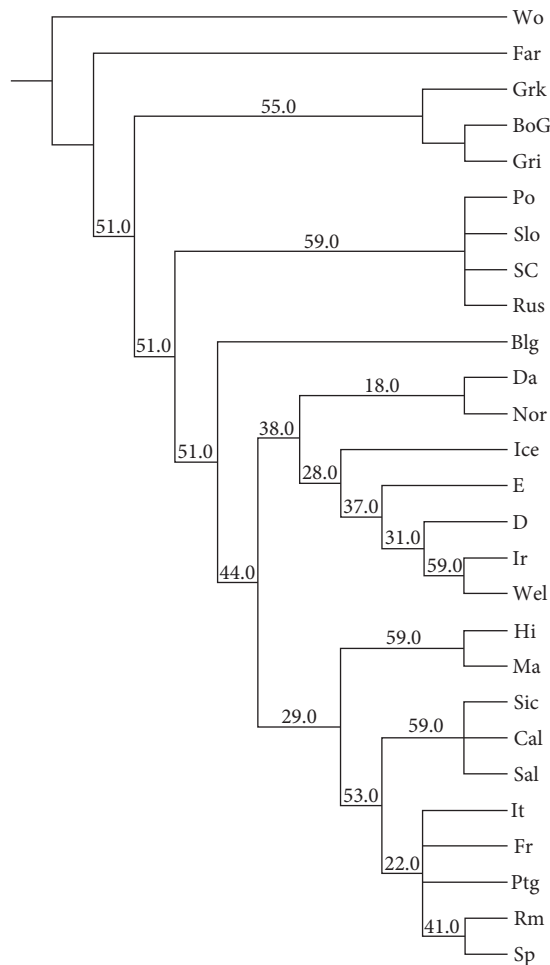


Figure 16. PAUP* consensus tree, from the reduced parameter grid (37 parameters).

4.3 Some remarks on language contact

In general, all algorithms and datasets agree in presenting obvious points of taxonomic failure only in a modest number of cases; rather interestingly, they are all cases where geography and known history independently suggest the possibility of areal contact: the questionable positions of Farsi and of Bulgarian in the full dataset with 56 parameters and the overstated Celtic-West Germanic relation, in the reduced one.

Even if such plausible effects of horizontal transmission may help explain some of these perturbations of genealogical expectations, on the other side they are in any case limited enough and relatively controllable. First, programs such as *Structure*

turned out to be able to warn us in any case about the possible mixed syntactic character of some Balkan languages, such as Rumanian and Bulgarian, yet interference is not so strong as to pull them away from Romance and Slavic, respectively (except for Bulgarian in a subset of the experiments with the reduced dataset).

Second, in other cases of obvious contact, like that of unbalanced interference affecting Southern Italian Greek, the expected genealogy is not disrupted by such effects under any program used.¹⁵

Third, secondary convergence may have contributed to the attraction between Celtic and West Germanic so visible in experiments with the reduced dataset, but is far from having caused it. Actually, one of the advantages of the parametric approach is that it allows one to define precisely a notion of **potentially borrowed characters** as follows: potentially borrowed characters are those parameter values shared by a language supposedly targeted by interference (e.g. Grico) with its neighbors (e.g. Italo-Romance), but crucially not with its closest genealogical relatives (e.g. homeland Greek). This provides a rough tool for measuring the plausibility of contact-based explanations. In this sense, the maximal set of characters potentially borrowed by Grico from any one of the three Romance dialects of Southern Italy in our sample is four out of 38 comparable parameter values (i.e. 10.5%); it reduces to 2/37 in Bovesè (5.4%). Recall that all methods have been still able to connect Grico, Bovesè, and homeland Greek to a common ancestry.

In the same vein, the perturbation of vertical expectations for Rumanian and Bulgarian can be correctly foreseen to be only marginally detectable, though possibly overstated by the combined contact sources of the area: Bulgarian may have maximally borrowed from Greek two parameter values (p7, DGP and p8, DGR) out of 40 (5%), and four out of 39 (less than 10.3%) from Rumanian (p7, DGP, p8, DGR, p12, DCN, and p32, GFO). This is expectedly not enough to remove it steadily from Slavic (only in PAUP* does it occur slightly out of its family, even if it is not attracted by any language in particular), though it may reinforce its surprising peripheral position in the family, because p7 (DGP) and p8 (DGR) are very consequential in terms of the neutralizing force their ‘-’ value exerts on other parameters.¹⁶ Rumanian, in turn, may have maximally borrowed three parameter values out of 40 (7.5%) from Greek (p29, GFR, p35, APO, and p43, TSL) and, indeed, shares one relevant (p12, DCN) out of 39 with Bulgarian (about 2.5%). With

15. The Greek varieties are associated with ‘minority’ communities and, traditionally, have not had any systematic exposure to their homeland variety (cf. the sources reviewed in Guardiano & Stavrou 2014).

16. Cf. Section 3.4 above. For the impact of the value ‘-’ of p7 (DGP) and p8 (DGR) on the overall implicational structure of Table A, cf. also the support material in the Appendix.

such values, it is not surprising that only the reduced dataset displays questionable taxonomic choices.

As noted above, no convergence of equal size and solidity can be identified in our parametric database for Celtic and West Germanic. This fact, compared to the Southern Italian and Balkan situation, suggests that some small amount of homoplasy rather than simple interference is likely to bias the groupings of Germanic and Celtic within the reduced dataset, which, precisely because it is smaller in size, is more exposed to such chance factors.

Pending further research, it appears that neither horizontal transmission of syntax nor homoplasy, for that matter, are such factors to severely impair the effectiveness of PCM, when applied to a robust number of parameters.

5. Summary

The application of various quantitative tools to parametric data already provides support to traditionally well-substantiated genealogical groupings, and confirms instability for taxonomies that previous comparative methods have equally challenged. It must be remarked that, presently, no available phylogenetic method seems capable of handling the implicational structure that characterizes parametric data; the application of those already existing inevitably implies some loss of important information. Therefore, in order for parameter data to become even more informative for phylogenetic purposes, it would be necessary to pursue the implementation of dedicated algorithms and methods.

In particular, the use of character-based algorithms is for now weakened by the absence of reliable models of markedness and parameter change/evolution. However, it is important that parametric comparison already yields acceptable genealogical outcomes in advance of, and unbiased by, any such hypotheses and of the inevitable idealizations and manipulations associated with them.

On the whole, only two points in our taxonomic results obtained from 56 parameters hint at some divergence from genealogies of traditional scholarship, which have been crucially assisted, unlike ours, by evidence from ancient varieties: the outlier position of Bulgarian within Slavic and that of Farsi. In the first case, simply importing grammaticalized definiteness from Greek and/or Rumanian may have overstated the position. In the second case, the intervention of minor areal interference from outside Indo-European is a plausible hypothesis (somewhat detectable through the further experiments); in addition, the problem partly finds parallelism in the outcome of vocabulary comparisons. Further local issues, such as the excessive interconnection of Celtic and West Germanic, seem to only occur as a byproduct of the reduced dataset.

In our view, once the rather moderate effects of horizontal transmission are properly factored out, the evidence of the refined experiments presented here appears to substantially back Longobardi & Guardiano's (2009) claim that parametric syntax encodes a genealogical signal of strength comparable to the lexical one.

References

- Alexiadou, Artemis, Liliane Haegeman & Melita Stavrou. 2007. *Noun Phrase in the Generative Perspective*. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110207491
- Baker, Mark C. 1996. *The Polysynthesis Parameter*. Oxford: Oxford University Press.
- Baker, Mark C. 2001. *The Atoms of Language*. New York: Basic Books.
- Barbançon, François G., Steven N. Evans, Luay Nakhleh, Donald Ringe & Tandy Warnow. 2013. An Experimental Study Comparing Linguistic Phylogenetic Reconstruction Methods. *Diachronica* 30:2, 143–170. DOI: 10.1075/dia.30.2.01bar.
- Biberauer, Theresa, ed. 2008. *The Limits of Syntactic Variation*. Amsterdam: John Benjamins. DOI: 10.1075/la.132
- Bortolussi, Luca, Giuseppe Longobardi, Cristina Guardiano & Andrea Sgarro. 2011. How Many Possible Languages Are There? *Biology, Computation and Linguistics* ed. by Gemma Bel-Enguix, Verónica Dahl & Maria Dolores Jiménez-López, 168–179. Amsterdam: IOS Press.
- Bošković, Željko. 2008a. What Will You Have, DP or NP? *Proceedings of the North East Linguistic Society 37* ed. by Emily Elfner & Martin Walkow, 101–114. Amherst: University of Massachusetts Press.
- Bošković, Željko. 2008b. The NP/DP Analysis and Slovenian. *Proceedings of the Novi Sad Generative Syntax Workshop* ed. by Ljiljana Subotić, 53–73. Novi Sad: Filozofski fakultet u Novom Sadu.
- Bošković, Željko. 2010. On NP and Clauses. Ms. University of Connecticut.
- Bromham, Lindel & David Penny. 2003. The Modern Molecular Clock. *Nature Reviews Genetics* 4:216–224. DOI: 10.1038/nrg1020
- Chomsky, Noam. 1975. *Reflections on Language*. New York: Pantheon.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Colonna, Vincenzo, Teresa Natile & Ronald R. Ferrucci. 2009. Comparing Population Structure as Inferred from Genealogical Versus Genetic Information. *European Journal of Human Genetics* 17:1635–1641. DOI: 10.1038/ejhg.2009.97
- van Cort, Tracy. 2001. *Computational Evolutionary Linguistics: Tree-based Models of Language Change*. Pomona, CA: Harvey Mudd College PhD Dissertation.
- Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley & Stephen C. Levinson. 2005. Structural Phylogenetics and the Reconstruction of Ancient Language History. *Science* 309:5743.2072–2075. DOI: 10.1126/science.1114615
- Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. An Indo-European Classification: A Lexicostatistical Experiment. *Transactions of the American Philosophical Society* 82:5.1–132. DOI: 10.2307/1006517
- Edwards, Anthony W.F., & Luigi Luca Cavalli-Sforza. 1963. The Reconstruction of Evolution. *Annals of Human Genetics* 27:105–106.
- Embleton, Sheila M. 1986. *Statistics in Historical Linguistics*. Bochum: Brockmeyer.

- Felsenstein, Joseph. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington.
- Felsenstein, Joseph. 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.
- Ghomeshi, Jila, Ileana Paul & Martina Wiltschko, eds. 2009. *Determiners: Universals and Variation*. Amsterdam: John Benjamins. DOI: 10.1075/la.147
- Gimbutas, Marija. 1973. The Beginning of the Bronze Age in Europe and the Indo-Europeans 3500/2500 B.C. *Journal of Indo-European Studies* 1.163–214.
- Gray, Russel D. & Quentin D. Atkinson. 2003. Language Tree Divergences Support the Anatolian Theory of Indo-European Origin. *Nature* 426.435–439. DOI: 10.1038/nature02029
- Guardiano, Cristina & Giuseppe Longobardi. 2005. Parametric Comparison and Language Taxonomy. *Grammaticalization and Parametric Variation* ed. by Montserrat Batllori, Maria-Lluïsa Hernanz, Carme Picallo & Francesc Roca, 149–174. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199272129.003.0010
- Guardiano, Cristina & Melita Stavrou. 2014. Greek and Romance in Southern Italy: History and Contact in Nominal Structures. *L'Italia Dialettale* LXXV, 121–147.
- Harris, Martin & Nigel Vincent, eds. 1988. *The Romance Languages*. London: Routledge.
- Heggarty, Paul. 2006. Interdisciplinary Indiscipline? *Can Phylogenetic Methods Meaningfully be Applied to Language Data — And to Dating Language? Phylogenetic Methods and the Prehistory of Languages* ed. by James Clackson, Peter Forster & Colin Renfrew, 183–194. Cambridge: McDonalds Institute for Archaeological Research.
- Hickey, Raymond, ed. 2010. *The Handbook of Language Contact*. Oxford: Wiley-Blackwell. DOI: 10.1002/9781444318159
- Hock, Hans H. & Brian D. Joseph. 1996. *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Mouton De Gruyter.
- Holtus, Günter, Michael Metzeltin & Christian Schmitt, eds. 1988–2005. *Lexikon der Romanistischen Linguistik (LRL)*. Tübingen: Niemeyer.
- Huson, Daniel H. & David Bryant. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23:2.254–267. DOI: 10.1093/molbev/msj030
- Joseph, Brian D. 1999. Romanian and the Balkans: Some Comparative Perspectives. *The Emergence of the Modern Language Sciences: Studies on the Transition from Historical-Comparative to Structural Linguistics in Honour of E.F.K. Koerner*, Vol. 2 ed. by Sheila Embleton, John E. Joseph & Hans-Josef Niederehe, 218–235. Amsterdam: John Benjamins. DOI: 10.1075/z.emls1
- Kayne, Richard. 2000. *Parameters and Universals*. New York: Oxford University Press.
- Keenan, Edward L. & Denis Paperno, eds. 2012. *Handbook of Quantifiers in Natural Language*. (= Studies in Linguistics and Philosophy, 90.) Dordrecht: Springer. DOI: 10.1007/978-94-007-2681-9
- Lazzeroni, Romano. 1968. Per una definizione dell'unità indo-iranica. *Studi e Saggi Linguistici. Supplement to L'Italia Dialettale* 8.131–159.
- Lazzeroni, Romano. 1998. Sanskrit. *The Indo-European Languages* ed. by Anna Giacalone Ramat & Paolo Ramat. London: Routledge.
- Lightfoot, David. 1979. *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.
- Lightfoot, David. 1991. *How to Set Parameters*. Oxford: Blackwell.
- Lindenbauer, Petrea, Michael Metzeltin & Margit Thir. 1995. *Die romanischen Sprachen: Eine einführende Übersicht*. Wilhelmsfeld: Egert.

- Lohr, Marisa. 1998. *Methods for the Genetic Classification of Languages*. Cambridge: Cambridge University PhD dissertation.
- Longobardi, Giuseppe. 1994. Reference and Proper Names. *Linguistic Inquiry* 25:4.609–665.
- Longobardi, Giuseppe. 2001. How Comparative is Semantics? A Unified Parametric Theory of Bare Nouns and Proper Names. *Natural Language Semantics* 9.335–369.
DOI: 10.1023/A:1014861111123
- Longobardi, Giuseppe. 2003. Methods in Parametric Linguistics and Cognitive History. *Linguistic Variation Yearbook* 3.101–138. DOI: 10.1075/livy.3.06lon
- Longobardi, Giuseppe. 2005. A Minimalist Program for Parametric Linguistics? *Organizing Grammar: Linguistic Studies for Henk van Riemsdijk* ed. by Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz & Jan Koster, 407–414. Berlin: Mouton de Gruyter.
- Longobardi, Giuseppe. 2008. Reference to Individuals, Person, and the Variety of Mapping Parameters. *Essays on Nominal Determination* ed. by Alex Klinge & Henrik Høeg Müller, 189–211. Philadelphia: John Benjamins. DOI: 10.1075/slcs.99.11lon
- Longobardi, Giuseppe. 2012. Convergence in Parametric Phylogenies: Homoplasy or Principled Explanation? *Parameter Theory and Language Change* ed. by Charlotte Galves, Sonia Cyrino, Ruth Lopes, Filomena Sandalo & Juanito Avelar, 304–319. Oxford: Oxford University Press.
DOI: 10.1093/acprof:oso/9780199659203.003.0016
- Longobardi, Giuseppe & Cristina Guardiano. 2009. Evidence for Syntax as a Signal of Historical Relatedness. *Lingua* 119:11.1679–1706. DOI: 10.1016/j.lingua.2008.09.012
- McMahon, April. 2010. Computational Models and Language Contact. In Raymond Hickey, ed., 31–47. Oxford: Wiley-Blackwell.
- McMahon, April, Paul Heggarty, Robert McMahon & Natalia Slaska, 2005. Swadesh Sublists and the Benefits of Borrowing: An Andean Case Study. *Quantitative Methods in Language Comparison* (= Transactions of the Philological Society, 103:2) ed. by April McMahon, 147–169. Oxford: Blackwell.
- McMahon, April & Robert McMahon. 2003. Finding Families: Quantitative Methods in Language Classifying. *Transaction of the Philological Society* 101:1.7–55.
DOI: 10.1111/1467-968X.00108
- McMahon, April & Robert McMahon. 2005. *Language Classification by Numbers*. Oxford: Oxford University Press.
- Meillet, Antoine. 1905. Avertissement. *Abrégé de grammaire comparée des langues indo-européennes d'après le Précis de grammaire comparée de K. Brugmann et B. Delbrück* ed. by Antoine Meillet & Robert Gauthiot, i–v. Paris: Klincksieck.
- Meillet, Antoine. 1922. *Les dialectes indo-européens*. Paris: Champion.
- Meillet, Antoine. 1924. *Le slave commun*. Paris: Champion.
- Nakhleh, Luay, Tandy Warnow, Don Ringe & Steven N. Evans. 2005. A Comparison of Phylogenetic Reconstruction Methods on an IE Dataset. *Transactions of the Philological Society* 3:2.171–192. DOI: 10.1111/j.1467-968X.2005.00149.x
- Nelson-Sathi, Shijula, Johann Mattis List, Hans Geisler, Heiner Frangerau, Russel D. Gray, William Martin & Tal Dagan. 2010. Networks Uncover Hidden Lexical Borrowing in Indo-European Language Evolution. *Proceedings of the Royal Society, Biological sciences* 1713.1794–1803.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press. DOI: 10.7208/chicago/9780226580593.001.0001
- Noonan, Michael. 2010. Genetic Classification and Language Contact. In Raymond Hickey, ed., 48–65. DOI: 10.1002/9781444318159.ch2

- Pellegrini, Giovan Battista. 1977. *Carta dei dialetti d'Italia*. Pisa: Pacini.
- Plank, Frans, ed. 2003. *Noun Phrase Structure in the Languages of Europe*. Berlin: Mouton de Gruyter.
- Pritchard, Jonathan K., Matthew Stephens & Peter Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155.945–959.
- Reesink, Ger, Ruth Singer & Michael Dunn. 2009. Explaining the Linguistic Diversity of Sahul Using Population Models. *PLoS Biology* 7:11. e1000241. DOI: 10.1371/journal.pbio.1000241
- Renfrew, Colin. 1987. *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Jonathan Cape.
- Renzi, Lorenzo & Gianpaolo Salvi. 1994. *Nuova Introduzione alla filologia romanza*. Bologna: Il Mulino.
- Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and Computational Cladistics. *Transactions of the Philological Society* 100:1.59–129. DOI: 10.1111/1467-968X.00091
- Rigon, Gabriele. 2009. *A Quantitative Approach to the Study of Syntactic Evolution*. Pisa: Università di Pisa PhD dissertation.
- Rigon, Gabriele. 2012. An Evolutionary Perspective on Diachronic Syntax. *Studi e Saggi Linguistici* 50:2.31–95.
- Roberts, Ian. 1998. Review of Historical Syntax in Cross-Linguistic Perspective (= Cambridge Studies in Linguistics, 74) by Alice Harris & Lyle Campbell (Cambridge: Cambridge University Press, 1995). *Romance Philology* 51.363–370.
- Roberts, Ian. 2012. *The Mystery of the Overlooked Discipline: Modern Syntactic Theory and Cognitive Science*. Available at <http://ling.auf.net/lingbuzz/001611>.
- Schleicher, August. 1861–1862. *Compendium der vergleichenden Grammatik der indo-germanischen Sprachen: Kurzer Abriss einer Laut- und Formenlehre der indo-germanischen Ursprache*. Weimar: Böhlau.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau.
- Sokal, Robert R. & Charles D. Michener. 1958. A Statistical Method for Evaluating Systematic Relationships. *University of Kansas Science Bulletin* 38.1409–1438.
- Swadesh, Morris. 1950. Salish Internal Relationships. *International Journal of American Linguistics* 16.157–167. DOI: 10.1086/464084
- Swofford, David L. 2003. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*. Sunderland, MA: Sinauer Associates.
- Szemerényi, Oswald. 1957. The Problem of Balto-Slav Unity. *Kratylos* 2.97–123.
- Tagliavini, Carlo. 1972. *Le origini delle lingue neolatine. Introduzione alla filologia romanza*. Bologna: Pàtron.
- Thomason, Sarah G. 2000. Linguistic Areas and Language History. *Languages in Contact* ed. by Dicky Gilbers, John Nerbonne & Jos Schaeken, 311–327. Amsterdam: Rodopi.
- Thomason, Sarah G. 2001. *Language Contact: An Introduction*. Washington, DC: Georgetown University Press.
- Thomason, Sarah G. 2004. Rule Borrowing. Ms., University of Michigan.
- Thomason, Sarah G. 2010. Contact Explanations in Linguistics. In Raymond Hickey, ed., 31–47.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Tomić, Olga M. 2003. The Balkan Sprachbund Properties. *Topics in Balkan Syntax and Semantics* ed. by Olga M. Tomić, 1–58. Amsterdam: John Benjamins. DOI: 10.1075/la.67.04mis

- Turakulov, Rust & Simon Eastal. 2003. Number of SNPs Loci Needed to Detect Population Structure. *Human Heredity* 55.37–45. DOI: 10.1159/000071808
- Willis, David. 2008. Celtic and English Language Contact and Shift. 2nd Historical Sociolinguistics Network Summer School. University of Bristol, 7–13 August 2008.

Appendix

Please see the appendix online at <http://dx.doi.org/10.1075/jhl.3.1.07lon.additional>

Subject index

A

actor 102-104, 108, 109
 anti-syncretic (type of development) 101, 102, 119, 120
see also syncretic (type of development)
 aorist 73, 107, 110-113, 115, 117n8, 118, 120
 auto-benefactive 108-120

B

Balto-Slavic hypothesis 127
 bootstrap 134-138, 135n2, 142n10, 145

C

causative 14, 17, 19, 102, 119, 120
 character-based (algorithms) 127, 128, 131, 144, 145, 148, 151
see also distance-based (algorithms)
 clitic climbing 83- 88, 93, 99
 complementizer 3, 31, 37-39, 43, 46, 47, 67n6, 82, 86, 91, 98
 control 44, 79, 80, 85, 89, 94, 98
 cycle 29, 33, 46, 47, 61, 62

D

diachronic typology 4
see also typology
 diathesis 22, 24, 103
 Discourse Projection 35, 36, 42
 distance-based (algorithms) 127, 128, 133, 134, 146, 148n13
see also character-based (algorithms)

F

finite 2, 3, 30, 31, 34-41, 43, 44, 51-54, 54n2, 55, 58n3, 68-72, 79, 80, 87-89, 93, 98, 109, 112

G

generative
 grammar 127
 parameters 127
 theory 126
 Greek
 Classical Greek 29, 32, 37-39, 41, 43, 45, 48, 107-110, 133
 Non Literary Koine Greek 32, 44
 Pre-Classical Greek 29, 32, 37, 41
 Post-Koine Greek 32, 36, 45, 46

H

horizontal transmission 125, 127, 129, 149, 151, 152
 hōs 29, 38, 41, 43-47
 hōti 29, 37-41, 39n2, 43-47
see also pleonastic hōti

I

Indo-European 125-129, 131, 139, 141, 147, 148, 151
 Indo-Iranian hypothesis 139
 infinitival complements 33, 40, 42, 43, 80, 87
 infinitive 2, 36, 40-44, 46, 55, 68, 85
 Italo-Celtic hypothesis 129

K

Kitsch 128, 133-138, 135n2, 142n10, 143, 146

L

labile 117, 119, 120
 lexical distance 140-143

M

mediopassive 11n2, 102, 105, 107-112, 119, 120
 middle (voice) 2, 7-12, 14, 15, 17, 19-22, 24, 25, 109, 117, 119-122
 mood shift (T₃) 32, 38, 42, 44, 46

N

network 129, 135-138
 non-specialized (voice markers) 101, 102, 108, 110, 111, 113, 116, 120
see also specialized (voice markers)

O

object 1, 2, 13-17, 19, 20, 30, 83n4, 84-87, 90, 102-104, 102n2, 108-110
 OID 42- 44, 47
 optative 29, 38, 39n1, 42, 44, 73
 of indirect discourse 42, 44

P

parameters 125-127, 130-132, 137, 139, 144-146, 149-151
 particle -z 8-12, 12n4, 14, 19, 20, 21, 25
 passive 2, 3, 8, 10, 19, 20, 23, 101-120, 102n1, 106n3, 107n5
 PCM 125-127, 129, 137, 138, 151
 perfect 7, 8, 23, 24, 39, 73, 81n3, 105, 106, 110, 111, 111n6, 113, 115-118
 person shift (T₂) 32, 45

phylogenetic 126, 128, 130,
131, 133, 144-148, 148n13, 151
phylogeny 126, 148
PIE verbal system 22
pleonastic *hóti* 29, 46
see also *hóti*
pro-drop 81
pronominal strategy 21, 25
see also verbal strategy
prosodic 3, 51, 52, 69n8, 72

R

raising 79, 80, 84-88, 94-98
reciprocal 15, 18, 118, 119
reconstruction 1-4, 9, 22n16,
51, 52, 52n1, 53, 54, 58, 66,
68, 69, 72, 101, 106n4, 116,
117, 125, 127, 128, 130
reflexive 7-21, 8n1, 21n15, 23,
24, 25, 92, 105, 107, 107n5,
108, 118, 119
relative 2, 3, 31, 37, 46, 51-56,
54n2, 58-62, 58n3, 61n4,
61n5, 65-68, 72, 73
relative-correlative 52, 53, 55,
56, 58, 60, 62, 64, 66-68,
72, 73
restructuring 2, 3, 79, 80, 83-
89, 93, 96, 98
RIDE 34-36
root characteristics 43

S

scrambling 84, 90
second-position clitic 79, 81
see also Wackernagel clitic

serial verb construction 80,
82, 83, 83n4, 83n5, 93
see also verb serialization
Southern Italy 131, 136, 137,
140, 147, 150
specialized (voice markers)
101, 102, 106, 110, 111, 113,
116, 118-120
see also non-specialized
(voice markers)
stative 2, 22-25, 101, 111, 111n6,
113-115, 117
Structure 126, 144-149
subject 1-3, 10, 13, 15, 17, 19,
20, 22, 25, 54n2, 65, 66,
80, 84, 85, 87-98, 102-104,
102n2, 106, 108-110
clitic 81, 81n2, 81n3, 82,
92, 93
subordinating conjunction
(T_1) 32, 37, 39, 43, 45
subordination traits 30, 32,
36, 39
syncretic (type of
development) 102, 118-120
see also anti-syncretic (type
of development)
syntactic distance 130, 133,
134, 140-143
syntactic pattern 101-105, 107,
113-116, 119, 120

T

T-cluster 32, 47
tense shift (T_4) 32, 38, 46

transitive 13, 14, 16, 17, 19, 20,
25, 98, 103, 104, 108, 109,
112, 113, 115, 117, 119, 120
typological 2, 3, 7, 12, 12n4, 21,
36, 52-54, 58, 62, 68, 72, 73,
102n1, 117
typology 4, 9, 11, 51-54, 58, 68,
72, 102, 103, 128
see also diachronic typology

U

unaccusative 81, 83n4
undergoer 102, 103, 108-110
Universal grammar 126
UPGMA 126, 133-138, 135n2,
145

V

verb-final(ity) 2, 51-53, 69,
69n8, 70, 72, 73
verb serialization 79, 80,
82, 93
see also serial verb
construction
verbal strategy 7, 10-12, 21, 25
see also pronominal strategy
voice 2, 4, 7-15, 17, 19, 20-22,
24, 25, 41, 101, 104, 105, 107-
109, 116-120

W

Wackernagel clitic 11, 21,
92, 93
see also second-position
clitic
word order 1-3, 30, 37, 39, 43,
52, 53, 70, 72, 73, 81, 93

Although for some scholars the very possibility of syntactic reconstruction remains dubious, numerous studies have appeared reconstructing a variety of basic elements of Proto-Indo-European syntax based on evidence available particularly from ancient and/or archaic Indo-European languages. The papers in this volume originate from the Workshop "PIE Syntax and its Development" (Thessaloniki 2011), which aimed to bring together scholars interested in these problems and to shine new light on current research into ancient Indo-European syntax. Special attention was paid to the development of the hypothetical reconstructed features within the documented history of Indo-European languages.

The articles in this volume were originally published in the *Journal of Historical Linguistics* 3:1 (2013).

ISBN 978 90 272 4263 1



9 789027 242631

John Benjamins Publishing Company