

# Interrogatives as relativization markers in Indo-European

Sandra Auderset

University of California, Santa Barbara | MPI for the Science of Human History

The use of interrogative pronouns as relative clause markers is often mentioned as a typical feature of European languages. This study presents an empirical approach to the distribution of interrogative pronouns as relative clause markers in time and space in the Indo-European language family. Based on a comprehensive sample of ancient and modern Indo-European languages, it is shown that interrogative-marked relative clauses are present in all stages of Indo-European within and outside of Europe. An analysis by branch suggests that this constitutes a case of parallel innovations subsequently spreading via language contact. The study also shows that interrogatives are used as relative clause markers independently of whether they are inflected pronouns or invariable markers.

**Keywords:** diachronic typology, interrogatives, relative clauses, Indo-European languages, areal features, morphosyntactic change

## 1. Introduction

European languages have featured prominently in the literature on relative clauses and, conversely, relative clauses have featured prominently in Indo-European linguistics. There is general consensus that relative clauses in European languages, and perhaps more broadly in Indo-European, are cross-linguistically unusual. Most often mentioned in this context are: (a) the use of an inflected clause-initial pronoun that indicates the role of the head noun in the relative clause, and (b) the use of an interrogative pronoun to mark relative clauses. This study provides a comprehensive overview of the latter feature, including the areal and diachronic distribution of the use of interrogative pronouns as relative clause markers within the Indo-European language family. The study also engages with the broader methodological significance of the pattern in question

for typology and areal linguistics. Despite the prominence of the topic, there are few dedicated comparative studies and consequently we are still facing open questions. These concern important aspects such as the reconstruction of relative clauses to proto-Indo-European, the stages and steps that take interrogatives from questions to subordinate clauses modifying nouns, and why interrogatives are used as relative clause markers in the first place. The present study cannot address all of these complex, interrelated questions, but it will provide a firm basis for further investigations. Here I will focus on answering the question of when and where interrogative-based relative clauses are found in Indo-European and whether or not the use of interrogative pronouns as such correlates with other features.

Including a broad sample instead of focusing on a few well-documented cases contributes to a more nuanced view of the relationship between the diachrony of morphemes (i.e., etymology) and that of larger units, such as phrases and clauses (i.e., morphosyntactic reconstruction). At this point, it is worth mentioning that the study is restricted to finite relative clauses: non-finite relative clauses in Indo-European languages typically do not consist of a separate clause boundary marker and thus there is no possibility for an interrogative pronoun to be used in such a construction.

The use of interrogative pronouns as relative clause markers is frequently mentioned together with the relative pronoun strategy, as for example by Haspelmath (2001a: 1494) discussing relative pronouns in Standard Average European: “Furthermore, in most SAE languages the relative pronoun is based on an interrogative pronoun (this is true of all Romance, all Slavic and some Germanic languages, Modern Greek, as well as Hungarian and Georgian)”. While this is not explicit – and perhaps not intended by the author – such statements evoke a connection between the two phenomena. In principle, however, they are independent of each other – or, rather, not all of the variables involved in either construction are inherently correlated with each other. The relative pronoun strategy found in many European languages, and occasionally outside of this area, displays two salient properties: (i) it is marked by a pronoun inflected for case indicating the role of the head noun in the relative clause, and (ii) this pronoun appears at the beginning of the relative clause (Comrie 1989: 149). An interrogative pronoun can be used as a relativization marker independently of these properties: the marker in question need not be case-marked nor appear clause-initially. Since this study is focused on Indo-European languages, many relative clause markers discussed will have one or both of these properties, but it is important to note that these are not diagnostics for identifying the feature in question. In order to explore the relationship of the relative pronoun strategy and interrogative-marked relative clauses, I investigate potential correlations with morphosyntactic features of

the markers, such as case, number, and noun class (both gender and inflectional class) marking.

In addition, the study includes an analysis of the use of relative clause markers as general complementizers. This strand of inquiry derives from the one proposed historical scenario concerning the rise of interrogatives as relative clause markers. In their discussion of the phenomenon in European languages – the most comprehensive treatment of the topic at hand – Heine & Kuteva (2006) conceptualize this development as a case of contact-induced grammaticalization (i.e., a series of grammaticalization processes triggered by influence from other languages). The direction of the development is from interrogative to relative clause (RC) marker via two intermediate stages in which the element is used first in indefinite, and then in definite complement or adverbial clauses, presented here with English illustrations (Heine & Kuteva 2006: 209–210):

1. Marking word questions – “*WHO came?*”
2. Introducing indefinite complement or adverbial clauses – “*I don’t know WHO came.*”
3. Introducing definite complement or adverbial clauses – “*You also know WHO came.*”
4. Introducing headed relative clauses – “*Do you know the woman WHO came?*”

These stages are not discrete entities, because there are intermediate constructions that can be interpreted as belonging to either of two subsequent stages. This is illustrated by Heine & Kuteva (2006: 210) with an English example: ‘What he wrote was not of much help to us’ – which is ambiguous between stage 2 and 3. It could mean that we do not know what he wrote and it does not matter, because it was not helpful (stage 2). The clause could also refer to a piece of writing we are familiar with (e.g., a letter), which was not helpful (stage 3). Note, however, that stages 2 and 3 also represent the two stages between which no major restructuring takes place: from stage 1 to 2, the construction transitions from monoclausal to biclausal, while stage 4 introduces a head noun. The difference between stage 2 and 3 is not structural, but depends on the referential status of the head.<sup>1</sup> The distribution of Interrogative-marked relative clauses (IRCs) is summarized as follows: “it is found in all Romance and Slavic languages, as well as in some Germanic languages, in

---

1. In his seminal book on relative clause typology, Lehmann (1984: 325–329) devotes a small chapter to the discussion of the commonalities and differences between interrogative and relative clauses that ultimately lead to them being marked the same way in some languages. He points out that many languages that do not use interrogatives in relative clauses nevertheless use them in indirect questions or so-called free relatives. This observation is congruent with and in fact probably underlies Heine & Kuteva’s (2006) stages of grammaticalization.

Modern Greek, Hungarian and Georgian.” (Heine & Kuteva 2006: 205 following Haspelmath 2001a: 1494). The authors propose that the polysemy arose independently in Italic and Slavic and then spread to many other languages via language contact. It is not based on inheritance, because it is found in many non-Indo-European languages of Europe such as Balkan Turkish and Basque. In other words, there were two centers of innovation and dispersal. This shows that while interrogative-marked relative clauses are considered a typically European phenomenon, the property spreads easily to other languages. As will be demonstrated in the following sections, including a wider range of languages in the picture suggests that Iranian might constitute a third center of innovation and that the construction is found outside of Europe as well.

Lastly, I provide a preliminary overview of the distribution of IRCs with respect to spoken vs. written language. Fiorentino (2007) points out that the relative pronoun strategy first arose in written language and the written origin explains its cross-linguistic uniqueness. She finds that the relative pronoun strategy is rare in spoken languages – even in Europe. This leads to the question of whether the same is true for IRCs or not. If so, this would be an argument in point to propose a connection between IRCs and the relative pronoun strategy; if not, it is a further indication of the independence of the two phenomena. There are two asymmetries present in the data that render any such assessment preliminary: Ancient Indo-European languages are necessarily only documented in written form, while for many modern (and recently extinct) languages information on both the spoken and written form is available. Furthermore, there is the issue of standard languages, i.e., separate forms of a language primarily used in formal, written contexts vs. the spoken/dialectal forms of the language. Ideally, all of them should be included in the sample (e.g., Standard Italian and all regional languages), but this is beyond the scope of this paper. Even so, including the variable in the study will help identify first tendencies and avenues for further research and I return to this point in §5.

As mentioned above, Indo-European languages have featured prominently in studies on relative clauses, especially in those focused on diachrony. This is due to their well documented history – something that is largely unavailable for other language families – but I suspect also to their dominant status in academia. While it is not the goal of this paper to provide a comprehensive reconstruction of the morphosyntax of Proto-Indo-European (PIE) relative clauses, the findings gathered here might still be able to shed some light on this. PIE is well reconstructed thanks to ample documentation of ancient languages, but there is no consensus on the reconstruction of the syntax or morphological marking of relative clauses in PIE, nor on whether PIE had relative clauses at all. This is partially attributable to the difficulties surrounding syntactic re-construction in general. Scholars

assuming that relative clauses did exist in PIE have put forward various proposals as to what the marker could have looked like (cf. Luján 2009 who provides a more in-depth overview of different proposals along with relevant references):

1. a construction marked by the PIE interrogative  $*k^{wi-}/k^{wo-}$
2. a construction marked by the PIE deictic/relative particle  $*yo-$
3. two constructions, one marked by  $*k^{wi-}/k^{wo-}$  and the other by  $*yo-$
4. an unmarked construction

All of these proposals assume that neither  $*k^{wi-}/k^{wo-}$  nor  $*yo-$  were originally relative pronouns. Rather, they had grammaticalized into such already in PIE. For the current study, proposals (2) and (4) imply that interrogative-marked relative clauses are an innovation of certain Indo-European (IE) languages or branches that took place after they split off from PIE. Conversely, proposal (1) implies that interrogative-marked relative clauses are inherited, i.e., non-IRCs represent the innovation. In proposal (3) both interrogative-marked RC and RC marked by  $*yo-$  would be inherited, and other strategies would constitute innovations. As it turns out, which of these scenarios fits the data collected here best depends to some degree on one's take on subgrouping and the etymology of the element  $*yo$ . I will return to this issue in §4.

Hendery (2012) presents a diachronic typological study of relative clauses, in which she discusses the sources of relative markers, among them interrogatives. About the distribution of IRCs, she says the following: "In terms of distribution, interrogative-based relative clause markers are more restricted than might be expected, given how well documented they are. (...) so the generalisation could be restated as interrogative-based pronouns *in general* being a European phenomenon." (Hendery 2012: 51). However, I think the relationship is rather the opposite: since IRCs are common in European languages, and European languages are well documented, it follows that IRCs are well documented, too. Her discussion mostly follows Heine & Kuteva (2006) and concludes with the assessment that contact is the driving factor behind the distribution of IRCs and that genealogy only plays a minor role, if any.

After this brief review of some of the previous research, I will first explain how the data was collected and analyzed in §2, before presenting a summary of the data by branch in §3. Section 4 discusses the distributions with a focus on diachrony and areality. Section 5 explores the correlations or absence thereof with other features of the relative clause markers, and the main findings are summarized in §6.

## 2. Data collection and analysis

### 2.1 Sample

The sample consists of 150 relative clause markers from 99 Indo-European languages and covers ancient and modern languages from each branch, excluding direct daughter languages of PIE like Messapian and Phrygian, for which we have limited records. I aimed at including all ancient IE languages for which enough documentation exists. For modern languages (i.e., languages spoken today), I aimed at a proportional coverage relative to the size of the branch as presented in Glottolog (Hammarström et al. 2019), cf. Table 1. Many of the branches that consist of only a few languages, such as Greek or Albanian, are slightly over- or underrepresented, all of which is due to the availability of relevant material. The greatest deviation is found for Indo-Iranian languages, which are underrepresented, and Balto-Slavic languages, which are overrepresented. As mentioned in §1 Slavic has been identified as one of the centers of innovation of IRCs, and thus a broad diachronic and genealogical coverage of this branch is crucial to assessing this claim. The opposite is true of Indic languages, which have been said to only exhibit IRCs in a few scattered cases (Heine & Kuteva 2006:205) – a finding which is supported by the current study. Lastly, Iranian languages are slightly overrepresented because Old and Middle Iranian languages are relatively well attested and described and thus offer a more detailed glimpse into diachrony than many other branches could provide. A complete overview of the languages surveyed and the sources used for data collection can be found in Appendices I and II, archived at <https://zenodo.org/record/4071195>. The geographical coverage is presented in Figure 1.

### 2.2 Coding

For each language, I collected the forms that can be used to mark relative clauses. These forms were then compared to content interrogatives in the same language to assess whether the marker is used in this function as well. They were also compared to complementizers (if such exist in the language) with respect to the same question of formal identity. For each relative clause marker,

I also coded for the following information: number of cases (if any), number of noun classes (if any), presence or absence of number marking, genre (written vs. spoken vs. both). I also note to which proto-Indo-European form the marker can be traced back, as far as this information is available. For reasons of scope, I did not attempt to provide the exact etymology, such as including the case form or

**Table 1.** Number of languages and percentages per branch (compared with Glottolog (Hammarström et al. 2019))

Branch	Subbranch	No. of Languages	% in Glottolog
Anatolian		3	1.7
Tocharian		2	0.3
Armenian		2	0.5
Greek		3	1.7
Albanian		2	0.7
Indo-Iranian		38	55.2
	Indic	(17)	(38.0)
	Iranian	(20)	(16.1)
	Nuristani	(1)	(1.0)
Balto-Slavic		13	3.9
	Slavic	(9)	(3.4)
	Baltic	(4)	(0.5)
Germanic		19	18.0
Italic		13	14.6
Celtic		4	2.4
Total		99	100

the addition of other morphemes. A more in-depth discussion of the coding decisions along with an annotated example can be found in Appendix II.

2.3 Working definitions and concepts

This study uses definitions tailored to the question at hand, to the specific characteristics of Indo-European languages, and with the available material in mind. The goal of this study is to better understand the distribution of IRCs in space and time. Even though Indo-European languages display a range of constructions and markers when it comes to relativization, only a subset of those will be considered here. For example, in some languages there are unmarked relative clauses, but since interrogative pronouns are never zero, these cases are excluded. Furthermore, while this is not true of all the world’s languages, most IE languages have a finite construction for expressing relativization which facilitates comparison. As mentioned above, the data is collected primarily from grammars and grammar

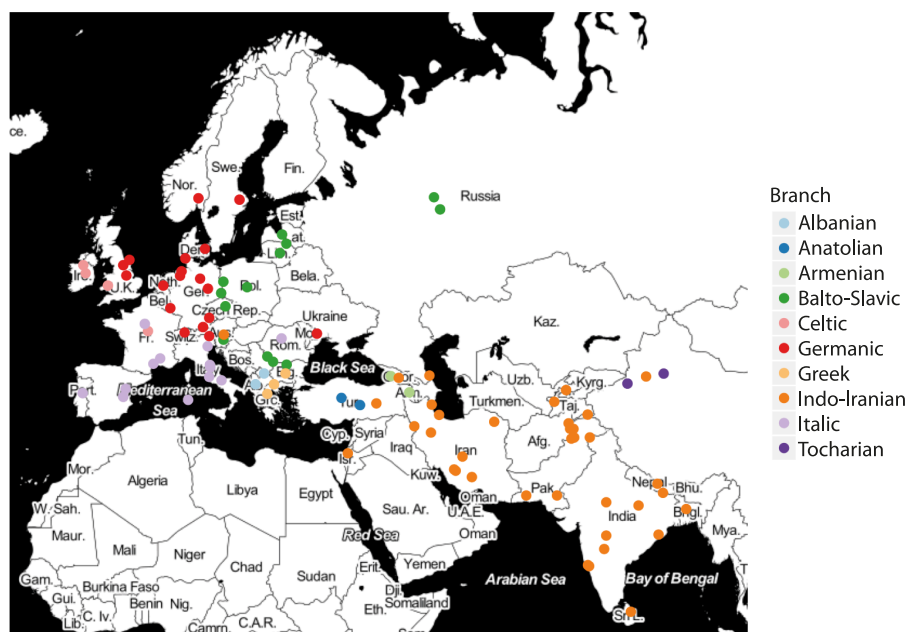


Figure 1. Geographic location and affiliation of the sample languages<sup>2</sup>

sketches and thus necessarily limited to the descriptions and examples provided there. Focusing on a subset of RC constructions allows for a broader sample.

### *Relativization and its markers*

Relativization is defined here as finite clausal noun modification, or in other words, noun modification that includes a finite predicate and its argument(s). This definition works well for Indo-European languages, in which word class distinctions are quite rigid. It is clear that such a definition could not serve as the basis for a more extensive cross-linguistic study. For practical and theoretical reasons, the present study is restricted to relative clauses with an overtly expressed head noun (i.e., to stage 4 in Heine & Kuteva's (2006) proposal), as in (1a) and (1b).

- (1) a. *The woman* [ *WHO is sitting at the table* ] is my sister.  
 b. *Do you have the pencil* [ *THAT I gave you this morning* ]?

Headless relative clauses are difficult to delimit from other types of dependent clauses, such as complement clauses (cf. (2a)) including indirect questions (cf.

2. This and all the following maps in the paper have been created in R with the ggmap package (Kahle & Wickham 2013).



(2b)). Note that (2b) can also be interpreted as a headless relative, i.e., it is indeterminate between the two interpretations. Headless relative clauses are also often used for generic statements that include some marking or construction specific to indefiniteness, as in (2c). Furthermore, headed relative clauses are easier to identify in texts and examples if there is no syntactic description (which is common for ancient IE languages, see below), and thus there is more material available for comparison.

- (2) a. *I heard [ WHAT you said].*
- b. *I wonder [ WHERE my pencil is].*
- c. *[ WHOEVER told you this] is a liar.*

This study does not systematically investigate the syntax of the relative clauses in question. This is primarily due to the lack of description and materials on the syntax of ancient IE languages, which means that syntactic information such as the exact construction, could only be reliably collected from (most) modern languages. I thus focus on the form and morphology of the marker(s) involved.

The majority of IE languages have finite relative clauses, either as the primary or at least as one of the available strategies. Finite in this context simply means that the verbal predicate takes the same inflection it would in a main clause. Other strategies present in many IE languages include the use of participles, i.e., verbal predicates that have taken some properties otherwise associated with nominals, such as case and number marking. In a few languages surveyed here, this is the main strategy for relative clauses. These participial or less finite constructions, however, do not have additional marking of the clause boundary, and thus are not systematically included.

### *Interrogative*

Any form that is used to form a content question (as opposed to a polar question), such as ‘what?’ and ‘where?’ in English. Note that I did not collect all the interrogatives of a given language, but only those that overlap with a relativization marker. In the languages that have case marking, there may be a division in the morphosyntax of interrogatives used for core roles and those used for adjuncts/obliques: the former are usually inflected for case (and number and/or noun class), while the latter are not.

### *Complementizer*

A form that marks a complement clause, i.e., a clause that functions as an argument of a verb, as in (3a). The form only signals that there is a clausal argument, and does not have any additional semantic content. Conjunctions or interrogatives used in

indirect questions, such as *why* in (3b), are thus not considered complementizers for the purposes of this study.

- (3) a.  $I_A$  know [ *that* she is coming ] O .  
 b.  $I_A$  wonder [ *why* she didn't come ] O .

### 3. Overview by branch

In this section, I briefly summarize the distribution and, where possible, the source of interrogatives as RC markers in each primary branch of Indo-European. A detailed description of each language with examples and paradigms as well as additional maps can be found in Appendix II, and the coding for each language is provided in Appendix I.

Regarding the subgrouping, I follow Hammarström et al. (2019), which is largely based on and congruent with the subgrouping presented in Fortson IV (2010:10). It differs from other proposals in that it only groups major branches together in cases where there is sufficient evidence from shared innovations such that the groupings are, for the most part, uncontroversial. These two subgroups are Indo-Iranian and Balto-Slavic (cf. also the discussion in Ringe 2017).

The order of presentation of each branch below roughly follows an east to west trajectory, starting with the eastern-most Tocharian and ending with the western-most Celtic languages.

#### 3.1 Tocharian

The two attested Tocharian languages, referred to as A and B (or eastern and western, respectively), are attested around 600 CE and ceased to be spoken around 900 CE, after the area was invaded by Turkic peoples. Due to limited documentation and time depth, the relative chronology of sound changes, as well as other aspects of grammar are not as well understood as in other IE languages.

Close affinities, some of which can be interpreted as early loans, with Germanic, Italic, and Celtic indicate that those speakers were in contact before the Tocharian people migrated east to the Tarim Basin (Gamkrelidze & Ivanov 1990). Later contact is documented through numerous loanwords from Sanskrit, the Prakrits, and various Old and Middle Iranian languages, which appear in the translations of Buddhist texts, the primary source through which Tocharian is attested. With respect to Iranian languages, most loanwords can be traced back to Old Iranian, Bactrian and Northeast Iranian languages (Kim 1999).

In both Tocharian A and B, interrogative pronouns are used to mark relative clauses but not complement clauses. Historically, the pronoun ‘who, what’ is a combination of the root *ku-* and inflected forms of the proximal demonstrative *su*. The former goes back to the PIE interrogative *\*k<sup>w</sup>i-/k<sup>w</sup>o-*, while the latter continues the PIE demonstrative *\*so-/to-*, but these etymologies are no longer synchronically transparent. Although the interrogative-relative pronouns do appear in other types of subordinate clauses, this is a relatively late development in both languages and remains rare. Moreover, they almost never appear in complement clauses (Hackstein 2012: 120). The sources of the attributive pronouns ‘which’ are unclear, cf. Table 2.

Given the contact scenario outlined above, it seems plausible if not likely that Tocharian acquired IRCs through contact with other IE languages. Interrogatives are not attested in Bactrian, but in the neighboring northwestern Prākṛit Ghāndārī we find specialized relative pronouns like in other Indic languages. I have summarized the information on Ghāndārī and closely related languages attested in a similar time period in Table 3. We can see that IRCs are absent from all three Indic contact languages, as well as from the earlier Iranian languages. They are, however, present in all the Central Middle Iranian languages, which are the closest relatives of Bactrian, and in the other Middle Iranian languages of the sample as well. This suggests that Tocharian IRCs are the result of contact with Middle Iranian languages. Of course, it is also possible that proto-Tocharian already had IRCs and we are thus dealing with a case of retention. In the absence of older documents and other sister languages, this question will have to remain open for now.

Table 2. Tocharian relative clause markers

Language	Attestation BP	RC Marker	INT	COMP	Origin
Tocharian A	1415–1115	<i>kus=ne</i>	yes	no	KW+TO
		<i>äntsam</i>	yes	no	NA
		<i>kuse</i>	yes	no	KW+TO
Tocharian B	1415–1115	<i>intsu</i>	yes	no	NA
		<i>mäksu</i>	yes	no	NA

3.2 Indo-Iranian

The Indo-Iranian branch is the largest within Indo-European, featuring approximately 320 languages to date (Hammarström et al. 2019). It is further divided into the Indic, Iranian, and Nuristani subbranches, which exhibit different ten-

**Table 3.** Relative clause markers in Indo-Iranian languages in contact with Tocharian \*

Language	Subbranch		Attestation BP	RC Marker	INT	COMP	Origin
Sanskrit	Indic		3015–2615	<i>yáh, yá, yát</i>	no	yes	YO
Pāli	Indic	Bihari	2515–2115	<i>yo, ya, yaṃ</i>	no	NA	YO
Ghāṇḍārī	Indic	C. Zone	2315–1515	<i>yo</i>	no	NA	YO
Avestan	Iranian		3215–2815	<i>hiiat, yaṭ</i>	no	yes	YO
Old Persian	Iranian	SW I.	2615–2315	<i>haya, hayā, taya</i>	no	yes	YO
Parthian	Iranian	C I.	2260–1790	<i>kē</i>	yes	no	KW
Khwarezmian	Iranian	C I.	2515–815	<i>ki, ci</i>	yes	no	KW
Sogdian	Iranian	C I.	2115–1015	<i>ke, ču</i>	yes	no	KW
Khotanese	Iranian	Sakha	2115–1015	<i>kye/i, ce/i, cu</i>	yes	NA	KW
Middle Persian	Iranian	SW I.	2215–1115	<i>kē, čē</i>	yes	no	KW

\* Additional abbreviations used in the table: C = Central, I = Iranian, SW = Southwestern.

dencies with respect to the use of interrogatives as relative clause markers: while it is almost completely absent from Indic, it is quite common in Iranian. Nuristani languages are not well documented, but in the one language included here, Wai-gali, there are no IRCs. The Nuristani branch will thus not be further discussed (but for details on Waigali, see Appendix II).

Proto-Indo-Iranian most probably had a correlative construction with a relative pronoun \*yo- and a co-referential demonstrative \*so-/to- in the matrix clause – this is the situation we find in the earliest attested Indic and Iranian languages alike. Indic languages basically remained this way up until today, although in most cases there was a significant reduction of the case system of the relative and correlative pronouns. Iranian on the other hand almost completely switched to relative constructions marked by interrogatives.

### 3.2.1 *Indic*

As mentioned above, in the majority of Indic languages, the relative pronoun or marker is not an interrogative but rather a dedicated one only used in relative clauses. Table 4 provides an overview of the presence and absence of IRCs. In the following, I provide an example of the two strategies and their marking most commonly found in Indic and discuss the three cases in which we do find an interrogative as relative clause marker.

In most Indic languages, relativization can be expressed either by a correlative construction or by a non-finite verb form, although in a few cases only one of these two strategies is used. A correlative construction consists of a relative clause

**Table 4.** Relative clause markers in Indic languages\*

Language	Subbranch	Attestation BP	RC Marker	INT	COMP	Origin
Sanskrit		3015–2615	<i>yáh, yá, yát</i>	no	no	YO
Pāli	Bihari	2515–2315	<i>yo, ya, yaṃ</i>	no	NA	YO
Maithili	Bihari		<i>je</i>	no	yes	YO
Ghāṇḍārī	Central Z.	2315–1515	<i>yo</i>	no	NA	YO
Domari	Central Z.		<i>illi</i>	no	no	L
Hindi	Central Z.		<i>jo</i>	no	no	YO
Vlax Romani	Central Z.		<i>kaj</i>	yes	yes	KW
Sinhala	Divehi-Sinhala		<i>NA</i>	no	yes	NA
Bengali	Eastern Z.		<i>ye(-), ya(-)</i>	no	yes	YO
Oriya	Eastern Z.		<i>je, jaha</i>	no	yes	YO
Nepali	Northern Z.		<i>jo, je</i>	no	NA	YO
Dameli	NW Z.		<i>ki, kyaa</i>	yes	no	KW
Sindhi	NW Z.		<i>jo, jā</i>	no	NA	YO
Kashmiri	NW Z.		<i>yus, yʊsɪ, yi</i>	no	no	YO
Old Marathi	Southern Z.	1015–715	<i>yo, ye, yeṃ</i>	no	NA	YO
Marathi	Southern Z.		<i>jo</i>	no	no	YO
Goan Konkani	Southern Z.		<i>khanco</i>	yes	no	KW
Goan Konkani	Southern Z.		<i>jo, ji, je</i>	no	no	YO

\* Additional abbreviations used in the table: NW = Northwestern, Z = Zone.

and a matrix clause, cf. (4). The head noun is present in both clauses, but can also be expressed with a pronoun in the matrix clause, and the relative clause has a dedicated relative pronoun not used for any other function. The coreferential pronoun in the matrix clause is either a distal demonstrative or a dedicated correlative pronoun. For this reason, correlative constructions are often analyzed as head-internal, but since the relative clause is not a syntactic argument of the matrix clause, they might be better analyzed as adjoined (Comrie 1989: 146). This can be seen in (4), which more literally translates as: “The pen he gave to me, that one is red”. The main clause is a complete clause in Nepali, i.e., the relative clause is co-referential with the demonstrative, but is not a syntactic argument in the main clause.

The other main strategy is the use of non-finite verb forms (variously referred to as participles, nominalizations, attributive verb forms, etc. in grammars), illustrated in (5). In most cases – and this holds for either of the two strategies – the

relative clause is placed before the matrix clause or head noun. This is another difference from most other Indo-European languages, which strongly favor post-nominal relative clauses.

(4) Nepali (Indic; Nepal)

[ *jun* *kalam* *us-le* *ma-lāi* *di-yo* ], [ *tyo rāto cha* ].

REL.DIR PEN 3SG-ERG 1SG-DAT give-PST.3SG DEM.DIST.DIR red be.3SG

“The pen which he gave to me is red.”

(Paudyal 2009: 12)

(5) Oriya (Indic; India)

[ *mo maa* *pəṭha-iba* *au* *ame kha-ith-iba* ] *pīṭha bōhut bhoḷo*.

my mother send-PTCP and we eat-PRF-PTCP cake very good

“The cake that my mother sent and that we ate was very good.”

(Neukom & Patnaik 2003: 329)

The only three Indic languages of the sample that use interrogatives as relative markers are Dameli, Vlax Romani, and one dialect of Konkani. In all of these cases, this is a result of long-standing and intensive language contact.

Vlax Romani is the language of traditionally nomadic people and is currently spoken in South-eastern Europe (roughly from Hungary to Greece). Relative clauses are marked by the interrogative *kaj* ‘where?’. The same interrogative is also used for marking complement clauses. While it might seem that this is pattern borrowing from Modern Greek, which also uses ‘where’ as a relative clause marker, this is unlikely because the same situation is present in many other varieties of Romani that are not currently in contact with Greek (Iglā 1996: 175). However, the Modern Greek construction might have reinforced the Vlax Romani one. The use of *kaj* ‘where’ as an RC marker was present already in Early Romani, from which all modern Romani varieties descend. In Early Romani, *kaj* is the most frequent RC marker. Matras (2004: 113) states that the use of interrogatives in relative clauses can either be attributed to Proto-Romani as an areal feature of northwest India or later to the Balkanization of Early Romani. In the former case, it could be connected to the IRCs present in Dameli.

Dameli is spoken in the Domel valley in Pakistan where it has been in close and longstanding contact with Pashto such that nowadays almost all speakers of Dameli are bilingual in Pashto (Perder 2013: 7). Even though the RC marker in Pashto is not formally identical (anymore) with an interrogative, it goes back to one (Morgenstierne et al. 2003: 17). This could suggest that Dameli acquired IRCs from Pashto at an earlier point in time. Apart from Pashto, Dameli is surrounded by other smaller Indic and Iranian languages, for which there is currently little documentation. In addition, the history of the Dameli people and language is not well understood and thus it is difficult to say precisely what the source for the IRCs in Dameli is. Based on the geographical location and the data collected in

this study, the most plausible sources are Pashto or other Iranian languages, possibly at earlier points in time (i.e., around the Middle Iranian period).

Konkani presents an interesting case, because one dialect, namely the Karnataka Saraswat dialect, acquired IRCs from Kannada, a Dravidian language which itself acquired IRCs through contact with Indic (Nadkarni 1975). Kannada used to only have participial relative clauses, but started adopting correlative constructions from Indic around the 10th century CE. However, since Dravidian languages do not have dedicated relative pronouns, Kannada used interrogatives to mark the relative clause. This is the construction that was subsequently borrowed into Karnata Saraswat Konkani, such that this dialect has relative clauses introduced by relative pronouns and others introduced by interrogatives (for more details and examples see Appendix II). As the previous discussion has shown, the three cases of IRCs in Indic arose each in a very specific setting of intensive language contact. The Dameli and Vlax Romani cases might be connected geographically, but this requires further research.

The use of an RC marker as a complementizer is found in about a third of the Indic languages for which this information is available. The languages that show a formal identity of relativizer and complementizer are predominantly found on the eastern edge of India and adjacent territories (cf. Figure 4 in Appendix II) and in the subbranches present in that area. This could indicate that this is an areal pattern, but further research is needed to address this question. As noted above, most Indic languages do not have IRCs, but even excluding the ones that do, the distribution remains the same. This suggests that the use of one form for marking both relative and complement clauses is not limited to interrogative-based markers, but rather is an independent phenomenon.

### 3.2.2 *Iranian*

The situation in Iranian differs greatly from that of its sister branch: the use of interrogatives in relative clauses is frequent and there is a clear diachronic pattern emerging from the data.

The earliest attested Iranian languages, Avestan and Old Persian, both show a marker going back to PIE *\*yo-* and do not use interrogatives. The RC construction and marking are thus essentially the same as in the ancient Indic languages, suggesting that correlatives marked by *\*yo-* go back to proto-Indo-Iranian. By the Middle Iranian period, however, the use of interrogatives has become the most common strategy: it is present in all the Middle Iranian languages included in the sample, cf. Table 5, even though they belong to different subbranches. This situation continues into the modern languages, although in some cases the interrogative and relative are no longer completely identical. This can often be attributed to the attrition of the case system, which led to invariable forms for most modern

Iranian languages. Even in the languages that do not have IRCs, the relative clause marker in most cases looks like it could be related to the interrogatives, but there is not enough historical work to prove (or disprove) this. Although some details are missing, we can summarize the situation within the Iranian branch as follows: While absent from the earliest Iranian languages, IRCs emerged in Middle Iranian and came to replace the older markers completely. The modern languages continue this for the most part.

Iranian languages are also different from Indic with respect to the use of RC markers as complementizers. About two thirds of Iranian languages use an RC marker as a complementizer, and it is especially common in the languages spoken today. Interestingly, all the languages that do not exhibit IRCs still have the complementizer-RC overlap, and conversely, half of the languages with IRCs also have this overlap. This adds further evidence to the independence of the two phenomena, also because there is no clear areal distribution in this case (see Figure 6 in Appendix II).

### 3.3 Anatolian

Anatolian is often considered to be the first branch to split off from PIE, and Hittite is the earliest attested IE language. Anatolian languages were in contact predominantly with non-IE languages such as Sumerian (isolate) and later Akkadian (Semitic). In Sumerian, relative clauses were formed with a nominalizing suffix (Alster 2002), and consequently there was no relative pronoun strategy or interrogative involved. Akkadian did have an inflected relative pronoun in the earlier stage of the language, but it indicated the role of the head noun in the main clause and not in the relative clause and was not used as an interrogative (Deutscher 2002). Anatolian languages became extinct around the 6th century CE as a result of a gradual language shift to Greek.

In all the Anatolian languages included in the study, relative clauses are marked by the interrogative for ‘who, what, which’, which goes back to the PIE interrogative *\*k<sup>w</sup>i-/k<sup>w</sup>o-*. This is illustrated in (6a) and (6b). In Hittite, the best documented Anatolian language, the neuter singular nominative-accusative form of the interrogative-relative is also used as a complementizer, as in (6c). Luwian also has IRCs, but it is not possible to say whether complement clauses were marked by a form of an interrogative pronoun as well. Although there is not enough material to say with certainty, IRCs were probably present in Lydian and Lycian as well.

#### (6) Hittite (Anatolian; Turkey)



Table 5. Iranian relative clause markers \*

Language	Subbranch	Attestation BP	RC Marker	INT	COMP	Origin
Avestan		3215–2815	<i>hiiat, yaŋ</i>	no	no	YO
Wakhi		o	<i>tsə(y)</i>	related	yes	NA
Wakhi		o	<i>ki</i>	related	yes	L
Parthian	Central I.	2260–1790	<i>kē</i>	yes	no	KW
Khwarezmian	Central I.	2515–815	<i>ki, ci</i>	yes	no	KW
Sogdian	Central I.	2115–1015	<i>ke</i>	yes	no	KW
Sogdian	Central I.	2115–1015	<i>ču</i>	yes	no	KW
Northern Kurdish	Central I.	o	<i>ku</i>	yes	yes	KW
Dimli	Central I.	o	<i>ki</i>	related	yes	KW
Gilaki	Central I.	o	<i>ki, čī</i>	yes	yes	KW
Ossetic	Central I.	o	<i>či/cy, ka/ci</i>	yes	no	KW
Balochi	Central I.	o	<i>ki</i>	related	yes	KW
Talysh	Central I.	o	<i>ki, ke</i>	yes	yes	KW
Parachi	O-P	o	<i>če</i>	yes	yes	KW
Pashto	Pashto	o	<i>čə</i>	related	yes	KW
Khotanese	Sakha	2115–1015	<i>kye/i, ce/i, cu</i>	yes	NA	KW
Old Persian	SW I.	2615–2315	<i>haya, hayā, taya</i>	no	yes	TO+YO
Middle Persian	SW I.	2215–1115	<i>kē, čē</i>	yes	no	KW
Farsi	SW I.	o	<i>ke</i>	yes	yes	KW
Tajik	SW I.	o	<i>ki</i>	yes	yes	KW
Angali (SW Fars)	SW I.	o	<i>ke</i>	yes	yes	KW
Judeo-Tat	SW I.	o	<i>ho</i>	no	yes	NA

\* Additional abbreviations used in the table: I = Iranian, O-P = Ormuri-Parachi, SW = Southwestern.

- a.

kuīt

iya-nun?<sup>3</sup>

what do.PST-1SG

“What did I do?”

(Hoffner & Melchert 2008: 350)
- b.

nu

NUMUN.HLA

[kue

human

šanhu-ta]

so seed.PL

REL.NOM.PL

all

roast.PST-3SG

“so the seeds which were all roasted”

(Hoffner & Melchert 2008: 425)

3. The source does not provide glossing or morpheme segmentation. The glosses and segmentation are mine. The clitic boundaries indicated with = are from the source material, as is the transliteration.

- c. maḥḥan=ma LÚ.MEŠ<sup>URU</sup> Aššur auēr [URU.DIDLÍ.ĤI.A BÀD=KAN  
 when=and person.PL Assyrian see.PST.3PL city.PL fortified=DM  
 kuit zaḥḥiyaz katta daškeuwan teḥḥ-un]  
 COMP battle.ABL down take.SUP put.PST-1SG  
 “when the Assyrians saw that I had begun to capture fortified cities in bat-  
 tle” (Hoffner & Melchert 2008: 426)

As summarized in Table 6, the use of an interrogative pronoun as a relative clause marker is found in all the Anatolian languages, all going back to the PIE interrogative. It is unlikely that IRCs arose in each Anatolian language due to contact, since there are no obvious sources for such a scenario. This leads to the conclusion that they existed already in proto-Anatolian. It is possible that the use of the neutral form as a complementizer also goes back proto-Anatolian, or it could be a Hittite innovation.

### 3.4 Armenian

Armenian is attested from about the 5th century CE on, and much of its earlier history is unclear. Armenian has been in contact with many IE languages for a long time, primarily with Greek, Latin, and Iranian. The latter had considerable influence on Armenian on every level of the language. Particularly influential was the Middle Iranian language Parthian (Meyer 2017: 5).

**Table 6.** Anatolian relative clause markers

Language	Subbranch	Attestation BP	RC Marker	INT	COMP	Origin
Hittite		3715–3215	<i>kuiš, kuit</i>	yes	yes	KW
Cuneiform Luwian	Luvic	3665–3215	<i>kuiš, kui</i>	yes	NA	KW
Hieroglyphic Luwian	Luvic	3115–2715	<i>kwis, kwadi</i>	yes	NA	KW

In Classical as well as Modern Eastern Armenian, the interrogative *or* ‘which’ is used as an RC marker. In Modern Eastern Armenian the relative clause marker is also used as a complementizer, but this a more recent development absent from Classical Armenian. The etymology of *or* is unclear – it could come from either PIE *\*yo-* or *\*k<sup>w</sup>i-/k<sup>w</sup>o-*. Based on this study, the latter seems more plausible, because if it goes back to *\*yo-* we have to assume a change in use to interrogative first, something which has not taken place in any other IE language to my knowledge.

It is uncontroversial that the use of an interrogative for relative clause marking in Armenian is the result of contact, but the contact ultimately cannot be attrib-

uted to a single language. Most likely, it is the result of considerable influence from Iranian languages, and might have been reinforced later on by Greek and Latin.

### 3.5 Greek

Relative clause marking in Greek shows an interesting pattern of replacement with different forms over time. Ancient Greek had a specialized relative pronoun – derived from PIE \*yo- and cognate with the Sanskrit (and many other Indic) forms – but this fell out of use in Medieval Greek. The subsequent period saw a range of relativization markers, one of which was the interrogative pronoun. This pattern survives into Modern Greek, but the interrogative is different. Relative clauses are marked with the invariable interrogative *pou* ‘where’, whereas Medieval Greek used *tís, tí* ‘who, what’. Only in Modern Greek is the relativizer also used as a complementizer.

It is likely that this development is due to or was reinforced by contact with other IE languages that already had this strategy. As in most cases of contact, it is difficult to pinpoint the exact source. It is well known that speakers of Greek and Latin were in close contact for centuries, with a good amount of bilingual inscriptions (and thus bilingual speakers). The use of IRCs could well have been introduced by native Latin speakers writing in Greek, cf. Leiwo 2002: 175. It is thus plausible that Greek acquired IRCs from Latin.

### 3.6 Albanian

In the Albanian branch, there are no IRCs in the spoken language, but in the written form the attributive interrogative *që* ‘which’ is used to mark RCs. In both spoken and written languages, the invariable *që* is also used as a general complementizer. Albanian languages are attested only from the 16th century on and we have limited knowledge of earlier stages. The historical facts are further complicated by the unclear etymology of *që*, and the heavy influence of Latin on Old Albanian. It is possible that *që* is a direct borrowing from Latin *que*, but it could also be derived from a form of PIE \*k<sup>w</sup>o- (Orel 2000: 245–246).

The situation is quite comparable to that of Armenian, except that the new strategy with the interrogative is restricted to the written form of the language. The restriction to the written form of the language and the considerable influence of Latin in this domain suggest that IRCs in Albanian developed under the influence of a contact language, most probably Latin.

Table 7. Slavic relative clause markers \*

Language	Subbranch	Attestation		INT	COMP	Origin
		BP	RC Marker			
Old East Sl.	East Sl.	1115–515	<i>iže, yaže, yeže</i>	no	no	YO
Russian	East Sl.		<i>čto</i>	yes	yes	KW
Russian	East Sl.		<i>kotoryj, kotoraja, kotoe</i>	yes	no	KW
Old Church Sl.	South Sl.	1215–915	<i>iže, ježe, jaže</i>	no	no	YO
Bulgarian	South Sl.		<i>deto</i>	related	yes	KW
Bulgarian	South Sl.		<i>kojto, kojato, koeto</i>	related	no	KW
S-C-B	South Sl.		<i>koji, koja, koje</i>	yes	no	KW
S-C-B	South Sl.		<i>što/šta</i>	yes	yes	KW
Slovenian	South Sl.		<i>ki</i>	no	no	NA
Slovenian	South Sl.		<i>katēri, katéra, katére</i>	yes	no	KW
Czech	West Sl.		<i>který, která, které</i>	yes	no	KW
Czech	West Sl.		<i>co</i>	yes	yes	KW
Czech	West Sl.		<i>jenž, jenže</i>	no	no	YO
Polish	West Sl.		<i>który, która, które</i>	yes	no	KW
Polish	West Sl.		<i>jaki, jaka, jakie</i>	yes	no	NA
Polish	West Sl.		<i>co</i>	yes	no	KW
Upper Sorbian	West Sl.		<i>štož, štóž</i>	yes	no	KW
Upper Sorbian	West Sl.		<i>kotryž</i>	yes	no	KW

\* Additional abbreviations used in the table: S-C-B = Serbian-Croatian-Bosnian, Sl = Slavic.

### 3.7 Balto-Slavic

#### 3.7.1 Slavic

In the Slavic branch there is a clear shift between the earlier attested languages and the modern ones. Old Church Slavic and Old East Slavic use a form of PIE \*yo- as a relative clause marker and not an interrogative pronoun. In all the modern languages, the relative marker is either formally identical or clearly related to an interrogative, although some languages like Czech retain an option derived from \*yo-. As can be seen in Table 7, many Slavic languages have multiple RC markers,

although in many cases all of them are interrogatives or related to such. Slavic languages are only attested relatively late, and we have documentation of early Slavic languages only for two branches, i.e., East Slavic and South Slavic, which means that it is difficult to assess the historical development of IRCs in this branch.

Note also that RC markers are not commonly used as complementizers – over three quarters of Slavic languages do not show it. It is absent from languages that do not have IRCs, and accounts for under half of those that do. Like Indo-Iranian, the Slavic branch adds to the impression that IRCs and RC-marked complement clauses are not linked in any way.

### 3.7.2 *Baltic*

All the Baltic languages included in the sample use interrogatives as relative clause markers. It is thus possible that this was the strategy used in proto-Baltic, but it is difficult to say with certainty because of the late attestation of Old Lithuanian and Old Prussian from the 13th and 14th century CE on, respectively. In Modern Lithuanian the same form is also used as a complementizer, but this seems to be a more recent development that is absent from Old Lithuanian and the sister language Latvian. Not only are Baltic languages closely related to Slavic ones, they have also been in longstanding contact with the latter, such that IRCs can be attributed to Slavic influence.

## 3.8 *Italic*

In the Italic branch, interrogatives and relatives are clearly related from the earliest attestation in the Sabellic languages on. The paradigms of the interrogative and relative pronouns are partially differentiated for the two functions, with some case forms coinciding and others not (for details see Appendix II). This is still the case in Classical Latin, but in the early Romance languages like Old Occitan and Old Spanish, the two paradigms have merged into one. This can be attributed to the loss of the case system that took place in the formation of Romance (cf. Manzini & Savoia 2014, among others).

In Latin, the use of the neutral form of the relative pronoun *quod* as a complementizer is attested, but it is a very marginal phenomenon. Later on, RC-markers used as complementizers become more common, accounting for about half of the Italic languages. They seem to be more common in the languages that also have IRCs, but this might be a byproduct of IRCs being overall very common in this branch. The rise of RC-marked complement clauses can be linked to the decline of the case system, too, since complementizers are often invariable.

### 3.9 Germanic

The Germanic branch stands out from the rest with respect to two aspects: in many Germanic languages, there are multiple relativization markers that have a wide range of sources, and it is the only branch in which the use of demonstratives as RC markers is common.

The use of interrogatives in relative clauses is completely absent from earlier stages of Germanic – although it is attested as early as Old English in headless relative clauses. In the modern languages, however, it is quite common, if not as the primary strategy then at least as one of the options. It is clear that this situation can be attributed to contact, although in some cases this might mean propagating an already existing strategy, while in others it might mean introducing a new strategy. As Heine & Kuteva (2006) observe, Germanic is ‘sandwiched’ between Romance and Slavic, i.e., between the two assumed centers of innovation, and it is thus hardly surprising that these languages acquired IRCs over time.

For the most part, relative clause markers are not used as complementizers, i.e., those two types of subordinate clauses are kept (at least) formally distinct from each other. On the surface, this is reminiscent of the situation in Slavic, but the distribution with respect to IRCs is quite different: None of the languages that have IRCs have it, but of those that do not, about a third use the same marker as a complementizer. As mentioned above, many Germanic languages use a demonstrative to mark RCs – and often the complementizer also goes back to a form of the demonstrative. It is thus not surprising, that Germanic languages with IRCs do not use these markers as complementizers.

### 3.10 Celtic

With respect to relative clauses, most Celtic languages share neither constructions nor markers with the other branches. There are reflexes of \*yo- as invariable relative marker, so for example in Transalpine Gaulish (cf. §11.1 in Appendix II), a fragmentarily attested continental Celtic language.<sup>4</sup> It is thus possible that proto-Celtic relative clauses were marked by \*yo-, but since Transalpine Gaulish and the better attested Insular Celtic languages belong to the same subbranch (according to Eska 2009), this could also be an innovation of that subbranch. More importantly, there is no trace of IRCs in Celtic, which makes it the only branch from which they are completely absent. Celtic languages will thus not be further discussed in the subsequent sections.

---

4. I would like to thank an anonymous reviewer for bringing this to my attention.

Table 8. Italic relative clause markers \*

Language	Subbranch		Attestation BP	RC Marker	INT	COMP	Origin
Umbrian	Sabellic		2715–1915	<i>poe, NA, NA</i>	related	NA	KW
Oscan	Sabellic		2715–1915	<i>pui, púd, paí</i>	related	NA	KW
Classical Latin	L-F		2090–1715	<i>quí, quae, quod</i>	related	no	KW
Old Occitan	L-F	Western Rom.	1325–615	<i>qui, que</i>	yes	yes	KW
Old Spanish	L-F	Western Rom.	1115–515	<i>qui, quien, que</i>	yes	yes	KW
Old Spanish	L-F	Western Rom.	1115–515	<i>cuyo, cuya</i>	yes	yes	KW
Italian	L-F	Western Rom.		<i>che</i>	yes	yes	KW
Italian	L-F	Western Rom.		<i>quale</i>	yes	no	KW
Venetian	L-F	Western Rom.		<i>che</i>	yes	yes	KW
French	L-F	Western Rom.		<i>qui</i>	yes	yes	KW
French	L-F	Western Rom.		<i>dont</i>	no	no	NA
Occitan	L-F	Western Rom.		<i>que</i>	yes	yes	KW
Spanish	L-F	Western Rom.		<i>que, quien</i>	yes	yes	KW
Spanish	L-F	Western Rom.		<i>cuyo, cuya</i>	related	no	KW
Portuguese	L-F	Western Rom.		<i>que, quem</i>	yes	yes	KW
Portuguese	L-F	Western Rom.		<i>cujo, cuja</i>	related	no	KW
Romanian	L-F	Eastern Rom.		<i>care</i>	yes	no	KW
Romanian	L-F	Eastern Rom.		<i>de</i>	no	no	NA
Campidanese	L-F	Southern Rom.		<i>chi</i>	yes	yes	KW
Sardinian							

\* Additional abbreviations used in this table: L-F = Latino-Faliscan, Rom.= Romance.

4. Moving through time and space

In this section I discuss IRCs primarily from a diachronic perspective while also taking into account areal aspects of their distribution. I begin with the earliest attestations of IRCs and then move through time to the present-day distribution, ending with a discussion of the implications for the status of IRCs in PIE.

There are several caveats that have to be taken into account in the assessment of this section both with respect to genealogy and geography. First of all, the attestation of a language does not necessarily correspond to the time period during which it was actually spoken, although in most cases we can assume considerable overlap between the two, at least with respect to the endpoint. Second, even in branches in which earlier languages are relatively well attested, these are in many cases not the direct ancestors of the languages we know today. And even if they are, there are still gaps of attestation during which we do not necessarily know

Table 9. Germanic relative clause markers \*

Language	Subbranch		Attestation BP	RC Marker	INT	COMP	Origin
Gothic	East G.		1815–1015	<i>izei, sei</i>	no	no	NA
Gothic	East G.		1815–1015	<i>ei, þei</i>	no	yes	NA
Gothic	East G.		1815–1015	<i>sa-ei, þat-ei, sō-ei</i>	no	no	TO
Old Norse	North G.	West Scand.	1315–715	<i>er, sem</i>	no	no	ONE
Swedish	North G.	West Scand.		<i>som</i>	no	no	ONE
Swedish	North G.	West Scand.		<i>vars, vilkas</i>	yes	no	KW
Swedish	North G.	North Scand.		<i>vilken, vilket</i>	yes	no	KW
Norwegian	North G.	North Scand.		<i>som</i>	no	no	ONE
Norwegian	North G.	North Scand.		<i>hvem, hva, hvilket</i>	yes	no	KW
Afrikaans	West G.	Franconian		<i>wat, wie</i>	yes	no	KW
Afrikaans	West G.	Franconian		<i>waar</i>	yes	no	KW
Afrikaans	West G.	Franconian		<i>welke</i>	yes	no	KW
Dutch	West G.	Franconian		<i>die, dat</i>	no	no	TO
Dutch	West G.	Franconian		<i>wie</i>	yes	no	KW
Luxembourgish	West G.	Franconian		<i>wou, wat</i>	yes	no	KW
Luxembourgish	West G.	Franconian		<i>deen, dat, déi</i>	no	yes	TO
German	West G.	Franconian		<i>der, die, das</i>	no	no	TO
German	West G.	Franconian		<i>wo, wie, warum</i>	no	no	KW
Old High German	West G.	High German	1315–965	<i>der, daz, diu</i>	no	yes	TO
Middle High German	West G.	High German	965–665	<i>der, diu, daz</i>	no	yes	TO
Middle High German	West G.	High German	965–665	<i>dâ, dar, dannen</i>	no	no	NA
Middle High German	West G.	High German	965–665	<i>so, und</i>	no	no	NA
Swiss German	West G.	High German		<i>wo</i>	yes	no	KW
Western Yiddish	West G.	High German		<i>vos</i>	yes	no	KW
Western Yiddish	West G.	High German		<i>velken</i>	yes	no	KW
Old Saxon	West G.	North Sea G.	1515–815	<i>the, thiu, that</i>	no	yes	TO
Old English	West G.	North Sea G.	1365–715	<i>se, þæt, seo</i>	no	no	TO
Old English	West G.	North Sea G.	1365–715	<i>þe</i>	no	yes	TO
Middle English	West G.	North Sea G.	865–515	<i>wich þat, who þat</i>	yes	no	KW+TO
Middle English	West G.	North Sea G.	865–515	<i>þat</i>	no	yes	TO
English	West G.	North Sea G.		<i>that</i>	no	yes	TO
English	West G.	North Sea G.		<i>who, what, which</i>	yes	no	KW
Old Frisian	West G.	North Sea G.	1215–415	<i>thī, thiu, thet</i>	no	yes	TO
Old Frisian	West G.	North Sea G.	1215–415	<i>thēr</i>	no	yes	TO
Eastern Frisian	West G.	North Sea G.		<i>däi</i>	no	no	TO



Table 9. (continued)

Language	Subbranch		Attestation BP	RC Marker	INT	COMP	Origin
Northern Frisian	West G.	North Sea G.		<i>dit</i>	no	no	TO
Northern Frisian	West G.	North Sea G.		<i>wat</i>	yes	no	KW
Northern Frisian	West G.	North Sea G.		<i>hur</i>	yes	no	KW
Northern Frisian	West G.	North Sea G.		<i>diar</i>	no	no	NA

\* Additional abbreviations used in the table: G = Germanic, Scand = Scandinavian

what took place. If a construction is already attested in the first records of a language, we cannot determine with certainty at what point in time it arose in that language. Thus, the dates of attestation given in Table 10 roughly represent the date of the earliest records we have of this language and are an approximation of the attestation of IRCs in that language. Third, there are obvious asymmetries in the genre between modern and extinct languages: in the latter case, we do not have access to spoken language and often the register of the written documents is quite formal. For present-day languages, the opposite can be true of languages that do not have a long written tradition and/or that have not been documented until recently. With respect to contact and location, similar difficulties arise. The location of ancient languages is often not precisely known, and thus determining which languages precisely they have been in contact with can be challenging. Although there is often evidence for contact through loanwords or ethnographic materials, proving such early contacts is difficult and remains to some degree speculative. All of this should not discourage studies of this kind; there is still much insight to be gained from the wealth of attested materials in Indo-European.

The earliest attestations of IRCs are found in Anatolian (Hittite & Luwian) around the second millennium BCE in modern-day Turkey. About a thousand years after the earliest attestation, IRCs are attested in Italic, namely in the two Sabellic languages Oscan and Umbrian that were spoken in present-day Italy. Only a short time later, IRCs surface in Middle Iranian, first in the Central Iranian languages Khwarezmian and Parthian, then in Middle Persian, Sogdian, and Khotanese. If we assume that IRCs were already present in PIE, then the Anatolian and Italic cases are best seen as a retention. The same assumption should then be made for Iranian, although this requires an additional assumption that the absence of IRCs in the two earliest attested Iranian languages Avestan and Old Persian is not reflective of the situation in proto-Iranian. What complicates this proposal is Middle Persian, which would show the retention of a feature absent from its precursor Old Persian. I thus see the Iranian case as one of innovation, probably one that took place after the Old Iranian period.

Before moving into the common era, IRCs next appear in Classical Latin. Up to then, IRCs are attested in only three branches, namely Anatolian, Italic, and Iranian. It is noteworthy that two of those, Anatolian and Iranian, were located outside of Europe, cf. Figure 2. In fact, it seems that they have spread predominantly eastwards at this point, except for the Italic languages. Note also that Slavic languages are completely absent from these early attestations.<sup>5</sup>

Most languages with IRCs attested in the first millennium CE are likely to have acquired this construction through contact (indicated by gray shading in the table), but note that in the alternative scenario of assuming IRCs in PIE, these would be attributed to retention (at least in the case of Tocharian and Armenian, and probably also in the two Baltic languages). The two early Romance languages probably have IRCs due to inheritance, either directly from proto-Italic or from proto-Romance. As illustrated in Figure 3, IRCs have moved further westwards in this millennium, now reaching all the way to the Baltic area and present-day Spain. Although IRCs are undoubtedly well represented in Europe at this point (around 1500 CE), they are by no means restricted to that area.

The distribution across IE languages spoken today broadly reflects earlier statements (e.g., by Heine & Kuteva 2006 and Haspelmath 2001a), but with small deviations – which is expected, since they refer to languages, while this study focuses on markers. Italic, previously identified as one of the centers of innovation from which IRCs spread to other IE languages, indeed has one of the highest proportions of IRCs, cf. Table 11, surpassed only by Balto-Slavic, the other proposed center of innovation. They are followed by Germanic and Indo-Iranian. A closer look at the sub-branches reveals, however, that the low value in Indo-Iranian is attributable to the already noted absence of IRCs in Indic. Iranian languages are on a par with Slavic and Italic, if one includes cases in which the RC marker is clearly related to an interrogative, even if they are not identical (anymore). In fact, Iranian shows the lowest percentage of languages without interrogative-based RCs, even lower than Italic and Slavic. This is also reflected in the geographical distribution of IRCs in modern Indo-European languages, shown in Figure 4. There is undoubtedly a high concentration in Europe, but it is by no means restricted to that area, due to many Iranian languages exhibiting the construction.

---

5. It is possible that Old Church Slavonic had IRCs already – an analysis that is suggested by Heine & Kuteva (2006: 221–222), but they also say that interrogative pronouns only developed into proper relative pronouns in Middle Bulgarian. This is something that needs further investigation, but would not change the overall picture much, since Old Church Slavonic is attested from about 800 CE on.

On a descriptive level, these findings mean that referring to the use of interrogatives as relative clause markers as a typical feature of present-day European languages is not adequate, since they are clearly found outside of this area, all the way to Central Asia. Historically, it is accurate if a convincing argument can be made for attributing the IRCs in Iranian to Slavic contact. As I have mentioned before, prehistoric contact is very difficult to assess, and thus there might be such an argument by scholars more familiar with the detailed history of the Iranian languages and peoples. Based on the data gathered here, my conclusion is rather that Iranian is a third center, if not of innovation then at least of propagation, of the use of interrogatives as RC markers: IRCs are attested in Iranian much earlier than in most other branches including Slavic, and Iranian languages are the likely sources of IRCs in at least Tocharian and Armenian.

**Table 10.** Attestations of IRC ordered by earliest time of attestation \*

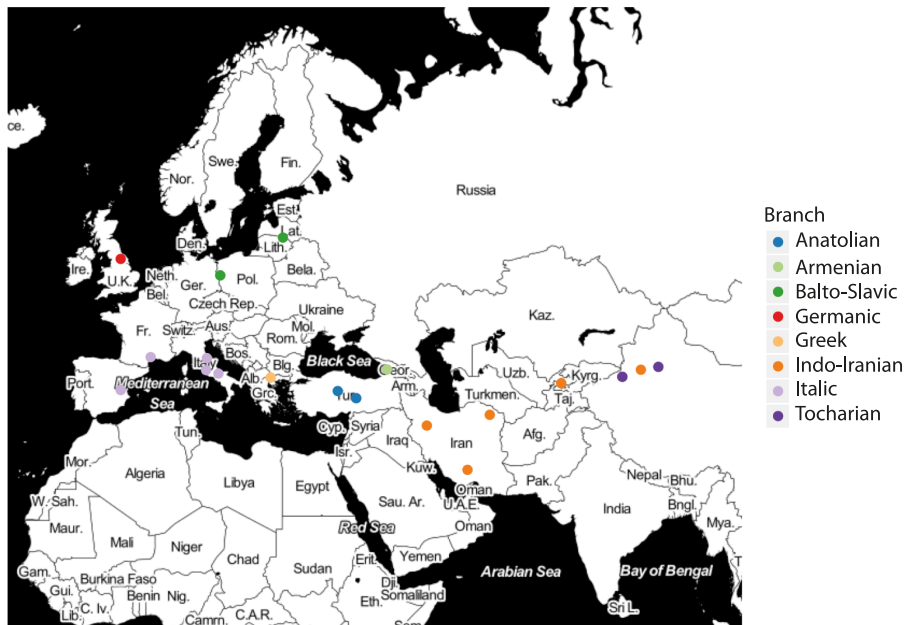
Language	Branch	Subbranch		Attestation	INT
Hittite	Anatolian			-1700	yes
Luwian	Anatolian	Luvic		-1650	yes
Umbrian	Italic	Sabellic		-700	related
Oscan	Italic	Sabellic		-500	related
Khwarezmian	Indo-Iranian	Iranian	Central I.	-500	yes
Parthian	Indo-Iranian	Iranian	Central I.	-245	yes
Middle Persian	Indo-Iranian	Iranian	Southwestern I.	-200	yes
Sogdian	Indo-Iranian	Iranian	Central I.	-100	yes
Khotanese	Indo-Iranian	Iranian	Sakha	-100	yes
Classical Latin	Italic	L-F		-75	related
Classical Armenian	Armenian			400	yes
Medieval Greek	Greek	East Greek	Koineic Greek	600	yes
Tocharian A	Tocharian			600	yes
Tocharian B	Tocharian			600	yes
Old Occitan	Italic	L-F	Western Rom.	690	yes
Old Spanish	Italic	L-F	Western Rom.	900	yes
Old Prussian	Balto-Slavic	Baltic		1369	yes
Old Lithuanian	Balto-Slavic	Baltic	Eastern Baltic	1500	yes

\* The minus indicates BCE, the other numbers are CE.



Figure 2. Geographic location of the languages with IRC attested before the year 0

The interpretation of what factors (or combinations thereof) best explain the observed distribution of IRCs over space and time is to some degree contingent on the model of subgrouping one follows and on the status of relative clauses one assumes for PIE. In §1 I briefly mentioned some proposals about the marking of RCs in PIE, and I will return now to the element *\*yo*. There are several instances of an invariable subordinator going back to *\*yo* found in Celtic (Jasanoff 1999), archaic Latin (Penney 2011), and possibly Luwian (Yakubovich 2008; Melchert 2012), which has led to another proposal regarding its origin: namely that *\*yo* was already used as an RC marker in PIE, albeit as an uninflected one. Based on this observation, an anonymous reviewer suggests that the absence of IRCs and the predominance of *\*yo* in the earliest stages of ‘Core Indo-European’ (defined as consisting of Indo-Iranian, Greek, Armenian, and Phrygian) should be seen as an innovation of this subgroup. Following a brief scenario outlined in Gamkrelidze & Ivanov (2010:339), they argue that the most economical scenario is to assume that PIE had both IRCs and RCs marked by the relative particle *\*yo*. In the ‘Core Indo-European’ subgroup, *\*yo* at first came to win out over IRCs and in this process acquired an inflectional paradigm like that of other pronouns. Other branches retained IRCs and largely phased out *\*yo*-marked relative clauses. This would mean that the IRCs in Anatolian, Tocharian, and Albanian could or should be seen as a retention rather than innovation or contact, respectively.



**Figure 3.** Geographic location of the languages with IRC attested up to but excluding present-day languages

While this scenario is certainly a plausible one, it is not the only plausible one. I do not adopt it here, because it rests on a few assumptions that are not uncontroversial, such as a ‘Core Indo-European’ subgroup or the reconstruction of \*yo as a relative particle. The subgrouping proposal followed in this study does not include a ‘Core Indo-European’ subgroup, and consequently the ‘winning’ of \*yo at the expense of IRCs would not be seen as one instance of change, but rather several instances of parallel change. Furthermore, in the reviewer’s scenario it is not clear whether the appearance of IRCs in Middle Iranian should be taken as a re-introduction due to contact, a later innovation, or a retention that happened to be absent from the attested Old Iranian languages. The same question arises for Medieval Greek and Armenian. Due to these issues, I conclude that this alternative scenario is not more economic nor does it explain the data better than my proposal above.

**Table 11.** Percentage of IRCs per branch in present-day IE languages

REL=INT	Balto-Slavic	Baltic	Slavic	Germanic	Italic	Indo-Iranian	Iranian	Indic
yes	80	100	75	58	70	38	57	21
related	10	0	12	0	15	17	36	0
no	10	0	12	42	15	45	07	79

**Table 12.** Presence and absence of IRC in present-day IE languages of minor branches

	Celtic	Albanian	Greek	Armenian
REL=INT	absent	partially present	fully present	fully present

No matter what the status of IRCs was in PIE, we observe a tendency in much of Indo-European for them to become the dominant relativization strategy or to replace other relativization strategies. And this is true for languages spoken in Europe as well as outside of it (and also independently of other variables, as will be shown in §5). In other words, there is something about interrogatives that makes them a suitable means to mark relative clauses in this language family and beyond (given its spread to non-Indo-European languages as well). Whether this happened for the first time in PIE already or as separate innovations later on has to remain an open question, since its assessment depends on other factors, viz. one's take on subgrouping and possible contact scenarios.

## 5. (Non-)Correlations with other features

This section explores the presence and absence of correlations of IRCs with other features of the markers, such as morphological marking. In a nutshell, there is no correlation between the presence of inflection and the use of an interrogative as a relative clause marker, and this holds even when considering each of the inflectional categories separately. Figure 6 shows a correlation matrix of each inflectional category and the types of relative markers.<sup>6</sup> We can see that there is a very slight negative correlation with class and number marking, which means that interrogative-based RC markers have a slight tendency not to be inflected for these features. This is probably a direct consequence of their interrogative nature, since interrogative pronouns in IE often exhibit a reduced paradigm compared to other pronouns. With case marking, there is no correlation whatsoever, cf. Table 13 showing the raw numbers.

In §1, I mentioned that IRCs might be linked with the relative pronoun strategy – or that some remarks seem to make this assumption. The relative pronoun strategy only applies if the pronoun in question indexes the role of the head noun

6. The correlation matrix was produced with the `ggcorr` function in R. All the variables were converted into numerical ones: the absence of an inflectional feature was coded as 0, its presence as 1. Relative markers that are not interrogatives were coded as 0, those that are related to an interrogative as 1, and those that are also interrogatives as 2. Occurrence in written language was coded as 2, written and spoken as 1, and spoken as 0.

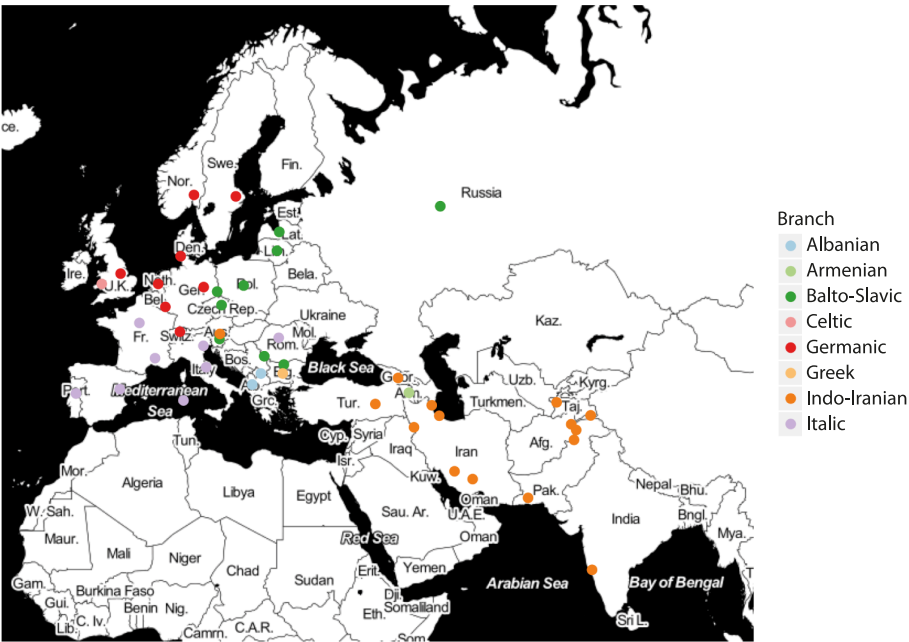


Figure 4. IRCs in present-day Indo-European languages

in the relative clause, i.e., is inflected for case. If the two phenomena are linked (for example diachronically or areally), then we would expect IRCs to be more common with inflected markers (especially with those inflected for case), but neither the raw numbers and nor the correlation matrix confirm this. As can be seen in Table 13, half of the RC markers are case-marked no matter whether the marker in question is an interrogative, related to one, or a different form.<sup>7</sup> This suggests that the two phenomena are not only independent in principle, but that the factors driving their respective distributions over space and time are actually different – which also means that they should be investigated independently from each other.

I have also mentioned the use of relative clause markers as complementizers throughout §3 to explore the possibility that IRCs extend to other subordinate clauses or pass through a stage at which they mark complement clauses before

7. This discrepancy might be partially due to differing definitions of case: in the present study, I only considered morphologically marked case, i.e., varying forms that are part of a paradigm, but with respect to the relative pronoun strategy, adpositions are often included in case (cf. Comrie 2006). The limitation to morphologically marked case is based on the consideration that morphologically marked case is almost always mentioned in grammars, while the use of a form with adpositions might not be (cf. Appendix II for more details).

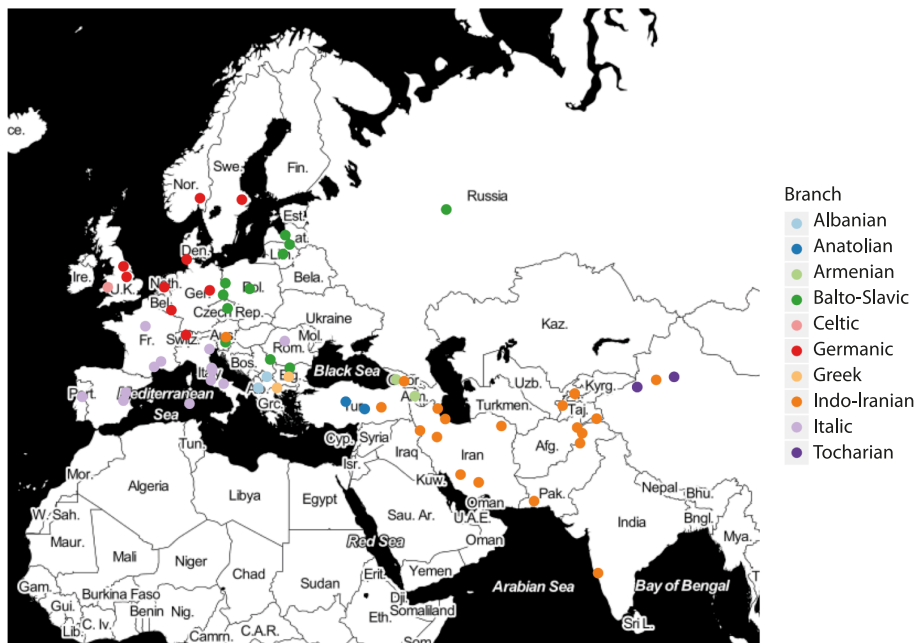


Figure 5. IRCs in extinct and present-day Indo-European languages

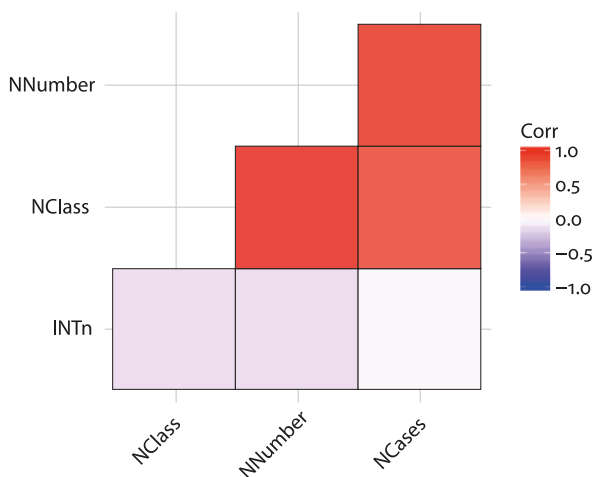


Figure 6. Correlation matrix with inflectional features

extending to relative clauses. Note that this cannot directly address Heine & Kuteva's (2006) claim that interrogatives first mark complement and/or adverbial clauses before relative ones (cf. §1). As mentioned in §2.3 I only included general



**Table 13.** (Non-)Correlations of IRCs with inflectional categories of RC markers (in counts)

REL=INT	Inflected	Invariable	Case	No case	Class	No class	Number	No number
yes	49	29	39	39	35	43	30	48
related	9	7	7	9	8	8	8	8
no	28	25	27	26	26	27	26	27

complementizers for practical reasons, while Heine & Kuteva (2006) include a wide range of subordinate clauses in their scenario. Nevertheless, it could be expected that IRCs show a higher proportion of markers that also function as complementizers than non-IRC. As can be seen from Table 14, there is no relationship between IRCs and having one marker cover both relatives and complementizers: the percentages are almost identical (corroborated by a correlation coefficient of  $-0.005$ ). As with inflection, this result should be interpreted as a possible sign of independence between the two phenomena that warrants further research.

**Table 14.** Relative markers as complementizers

REL=INT	REL=COMP	%	REL≠COMP	%
yes	25	(34)	49	(66)
related	10	(64)	4	(36)
no	16	(33)	33	(67)

**Table 15.** Relative markers in spoken and written language

REL=INT	Written	Written & spoken	Spoken
yes	30	19	30
related	6	2	5
no	28	5	21

Finally a comment is in order concerning the relationship between IRCs and spoken vs. written language. In §1, I mentioned that the relative pronoun strategy has its origins in written language and that we might expect to see a skewing towards written language for IRCs as well. To the extent that the data is available, we do not find such a tendency, as IRCs are equally common with spoken as with written languages, cf. Table 15. If anything, there is a slight tendency for IRCs to be associated with spoken language (cf. the correlation coefficient of  $-0.052$ ). This

adds support to the finding that IRCs are independent from the relative pronoun strategy, since the latter is more common (and often even restricted to) written language (cf. Fiorentino 2007), and suggests that IRCs are not tied to a shared written culture in the same way. Anecdotal evidence for this is also found within different varieties of one language: While Standard German does not have IRCs, the colloquial, spoken variety in Southern Germany does have them, as do all Swiss German varieties.

All of the (non-)correlations mentioned above are summarized in the overall correlation matrix in Figure 7. We see that IRCs exhibit either no or very weak negative correlations with each of the variables investigated in this study. As a nice side effect, the correlation matrix illustrates the use of portmanteau morphemes for case, class, and number across much of IE relative markers: all of of these features are highly correlated with each other. It also shows that written language has a positive correlation with each of the inflectional variables, adding support to Fiorentino's (2007) claims. The absence of strong correlations with any of the variables might also explain why the construction is so easily borrowed into non-IE languages with distinct typological profiles. If IRCs were constrained for example by the presence or absence of case marking this would be expected to be reflected in the languages that borrowed this construction. Although this has not been systematically explored in the present study, it does not seem to hold.

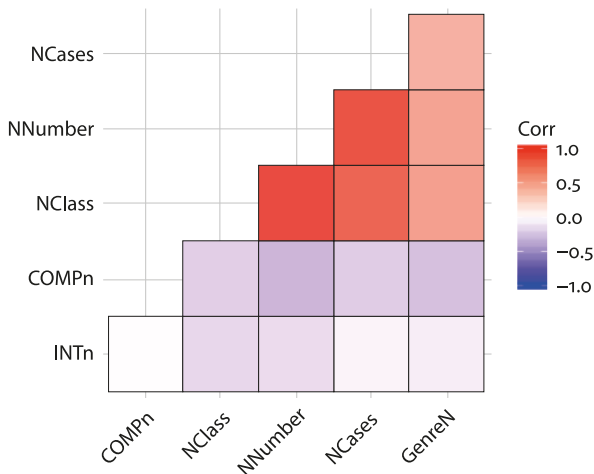


Figure 7. Correlation matrix across all variables

## 6. Conclusion

There are two key findings that emerge from the previous sections. The first is that interrogative-marked relative clauses are not, and never have been, confined to Europe. This means that taken by themselves, they are not an areal feature of Europe because they are common in adjacent areas as well. The second finding is that they are independent of the relative pronoun strategy: they are not tied to inflected markers and do not show the same restriction to written language and a shared written cultural tradition. The data collected here also suggest that the development of a form that is used in both relative and complement clauses is independent from IRCs as well. There is no indication that one development is diachronically related to the other.

Regarding the history of IRCs, I argue that a single origin of IRCs is unlikely given that they are absent from most of the early attested languages, such as Sanskrit or Ancient Greek. The occurrence in the Anatolian branch is thus best explained as an early innovation not connected to the other centers of innovation. These facts suggest that relative clauses in PIE were not marked by interrogatives. It remains an open question whether the Iranian branch constitutes a separate center of innovation, or whether its development of IRCs can be argued to be the result of Slavic contact. With respect to other languages, it seems uncontroversial that Slavic and Italic are indeed the centers from which the innovation spread. There is, however, a competing view in which IRCs were already present in PIE and retained as such in branches not belonging to Core-IE. Further investigations will have to decide which of these proposals better accounts for the diachronic and synchronic facts of IE languages overall.

Another question worth raising is whether the development is adequately characterized as contact-induced grammaticalization. In most cases, languages that acquired IRCs through contact already had relative clauses – they were just marked by another form. At least in some cases then, the development might be more straightforwardly described as a case of pattern borrowing (cf. the Konkani case). Another argument in favor of this analysis is that since the languages in question are all related, they share many syntactic characteristics, which should make it easier to copy patterns from one language to another. This brings us to another important question already mentioned in §1, namely that of why interrogatives should be used to mark relative clauses in the first place. Heine & Kuteva's (2006) grammaticalization scenario suggests that it has something to do with indefiniteness and indirect questions. It has been noted that indefinite pronouns often derive from interrogatives (Haspelmath 2001b; Diessel 2003), which could mean that indefiniteness is the crucial link between interrogatives and relatives. To demonstrate this, one would have to show that in each of the cases in which

an interrogative became a RC marker, the form also was or is used as an indefinite pronoun.<sup>8</sup> I have not conducted such a study, but it is quite likely that this is true. I believe that the most fruitful advances to answer this question will come from detailed studies of the semantics of the interrogatives apart from their interrogative function. As has been shown above (§5), morphological properties cannot be invoked for an explanation. Thus, focusing on the semantic properties of the elements involved could provide a better answer to this question in the future. A step in this direction would be to conduct a follow up study on headless relative clauses and indirect questions marked by interrogatives. This would help refine or discard the preliminary statements made here and would also shed more light on the diachronic stages. Furthermore, it will elaborate on the questions of what the drivers of this change are, which in turn would demand including more detailed information on the syntax of the constructions each marker can be found in as well as a more detailed etymological account. Other obvious avenues of further research would include the syntax of the construction IRCs are found in, as well as their exact function, for example as core arguments or adverbials.

Apart from expanding the study qualitatively it would also be worthwhile to approach the issue with Bayesian phylogenetics. Such methods have proven insightful for similar studies in the past (cf. Widmer et al. 2017) and can provide insights into branch dynamics (i.e., changes over time along branches) which are difficult to assess with qualitative methods alone.

Since it is well known that IRCs have also been borrowed into non-IE languages, a closer look at those languages could tell us more about the presence or absence of potential correlations with other features. In this study, I hope to have shown that diachronic typology can adequately address such questions, even in related and well studied language families like Indo-European.

## Acknowledgements

I would like to thank Bernard Comrie, Marianne Mithun, Eric W. Campbell, Claire Bowerman and two anonymous reviewers for their comments on earlier drafts. Their contributions have much improved this paper. All responsibility for remaining errors is mine.

---

8. Lehmann (1984: 327) notes that indirect questions are not necessarily marked by interrogatives, but can also be modeled after relative constructions with other markers, such as nominalizers.

# Supplementary material

Supplementary materials are available from <https://zenodo.org/record/4071195>

## Abbreviations

BP	before present	DEM	demonstrative
IE	Indo-European	DIR	direct case
IRC	interrogative-marked relative clause	DIST	distal
KW	PIE interrogative *k <sup>w</sup> i-/k <sup>w</sup> o-	DM	discourse marker
L	loanword, borrowed form	ERG	ergative
PIE	Proto-Indo-European	INT	interrogative
RC	relative clause	NOM	nominative
TO	PIE demonstrative *so-/to-	PL	plural
YO	PIE *yo-	PRF	perfect
1	first person	PST	past
3	third person	PTCP	participle
ABL	ablative	REL	relative
COMP	complementizer	SG	singular
DAT	dative	SUP	supine (non-finite verb form)

## References

- Alster, Bendt. 2002. Relative clauses and case relations in Sumerian. *Wiener Zeitschrift für die Kunde des Morgenlandes* 92: 7–31.
- Comrie, Bernard. 1989. Relative clauses. In *Language universals and linguistic typology* 138–164. 2nd edn. Chicago: University of Chicago Press.
- Comrie, Bernard. 2006. Syntactic typology: Just how exotic ARE European-type relative clauses. In Ricardo Mairal & Juana Gil (eds.), *Linguistic universals*, 130–154. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511618215.007>
- Deutscher, Guy. 2002. The Akkadian relative clauses in cross-linguistic perspective. *Zeitschrift für Assyriologie und vorderasiatische Archäologie* 92(1): 86–105. <https://doi.org/10.1515/zava.2002.92.1.86>
- Diessel, Holger. 2003. The relationship between demonstratives and interrogatives. *Studies in Language* 27(3): 635–655. <https://doi.org/10.1075/sl.27.3.06die>
- Eska, Joseph F. 2009. The emergence of the Celtic languages. In *The Celtic languages*, 36–41. London: Routledge.
- Fiorentino, Giuliana. 2007. European relative clauses and the uniqueness of the relative pronoun type. *Italian Journal of Linguistics* 19(2): 263–291.
- Fortson IV, Benjamin W. 2010. *Indo-European language and culture: An introduction*. 2nd edn. Malden, MA: Wiley-Blackwell.

- Gamkrelidze, Thomas V. & Vyacheslav V. Ivanov. 1990. The early history of Indo-European languages. *Scientific American* 262(3): 110–117.  
<https://doi.org/10.1038/scientificamericano390-110>
- Gamkrelidze, Thomas V. & Vyacheslav V. Ivanov. 2010. *Indo-European and the Indo-Europeans: A reconstruction and historical analysis of a proto-language and proto-culture part i: The text. part ii: Bibliography, indexes*. Berlin: Mouton de Gruyter.
- Hackstein, Olav. 2012. The evolution of finite complementation in Tocharian. In Georges-Jean Pinault, Michael Peyrot, Jens Elmegård Rasmussen & Thomas Olander (eds.), *Tocharian and Indo-European studies*, vol. 13, 117–148. Copenhagen: Museum Tusculanum Press.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath. 2019. *Glottolog 4.1*. Available online at <http://glottolog.org>
- Haspelmath, Martin. 2001a. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 1492–1510. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110171549.2.14.1492>
- Haspelmath, Martin. 2001b. *Indefinite pronouns*. Oxford: Oxford University Press.
- Heine, Bernd & Tania Kuteva. 2006. From question to subordination. In *The changing languages of Europe*, 204–243. Oxford: Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780199297337.003.0006>
- Hendery, Rachel. 2012. *Relative clauses in time and space: A case study in the methods of diachronic typology*. Amsterdam: Benjamins. <https://doi.org/10.1075/tsl.101>
- Hoffner, Harold & H. Craig Melchert. 2008. *A grammar of the Hittite language*. Winona Lake, ID: Eisenbrauns.
- Igla, Birgit. 1996. *Das Romani von Ajia Varvara: Deskriptive und historisch-vergleichende Darstellung eines Zigeunerndialekts*. Wiesbaden: Harrassowitz.
- Jasanoff, Jay H. 1999. Some relative forms of the verb in Old Irish. In *Compositiones Indogermanicae in memoriam Jochem Schindler*, 205–222. Prague: Enigma.
- Kahle, David & Hadley Wickham. 2013. ggmap: Spatial visualization with ggplot2. *The R Journal* 5(1): 144–161. <https://doi.org/10.32614/RJ-2013-014>
- Kim, Ronald. 1999. Observations on the absolute and relative chronology of Tocharian loanwords and sound changes. *Tocharian and Indo-European Studies* 8: 111–138.
- Lehmann, Christian. 1984. *Der Relativsatz*. Tübingen: Narr.
- Leiwo, Martti. 2002. From contact to mixture: Bilingual inscriptions from Italy. In J. N. Adams, Mark Janse & Swain Simon (eds.), *Bilingualism in ancient society*, 168–194. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199245062.003.0007>
- Luján, Eugenio R. 2009. Relative clauses in Proto-Indo-European. In Vit Bubenik, John Hewson & Sarah Rose (eds.), *Grammatical change in Indo-European languages*, 221–234. Amsterdam: Benjamins. <https://doi.org/10.1075/cilt.305.22luj>
- Manzini, M. Rita & Leonardo M. Savoia. 2014. From Latin to Romance: Case loss and preservation in pronominal systems. *Probus* 26(2): 217–248.  
<https://doi.org/10.1515/probus-2014-0008>
- Matras, Yaron. 2004. *Romani – a linguistic introduction*. Cambridge: Cambridge University Press.
- Melchert, H. Craig. 2012. Genitive case and possessive adjective in Anatolian. In Vincenzo Orioles (ed.), *Per Roberto Gusmani. Studi in ricordo. Linguistica storica e teoretica*, vol. 2, 273–286. Udine: Forum.

- Meyer, Robin. 2017. Iranian-Armenian language contact in and before the 5th century CE. Oxford Wolfson College, University of Oxford dissertation.
- Morgenstierne, Georg, Josef Elfenbein, D. N. MacKenzie & Nicholas Sims-Williams. 2003. *A new etymological vocabulary of Pashto*. Wiesbaden: Reichert.
- Nadkarni, Mangesh V. 1975. Bilingualism and syntactic change in Konkani. *Language* 51(3): 672–683. <https://doi.org/10.2307/412892>
- Neukom, Lukas & Manideepa Patnaik. 2003. *A grammar of Oriya* (ASAS 17). Universität Zürich.
- Orel, Vladimir Ė. 2000. *A concise historical grammar of the Albanian language: Reconstruction of Proto-Albanian*. Leiden: Brill.
- Paudyal, Netra. 2009. The syntax of three-argument verbs in Nepali. Qualifying paper.
- Penney, John. 2011. Archaic and old Latin. In James Clackson (ed.), *A companion to the Latin language*, 220–235. Malden, MA: Wiley-Blackwell.
- Perder, Emil. 2013. A grammatical description of Dameli. Stockholm University dissertation.
- Ringe, Don. 2017. Indo-European dialectology. In Jared Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*, vol. 1, 62–75. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110261288-006>
- Widmer, Manuel, Sandra Auderset, Johanna Nichols, Paul Widmer & Balthasar Bickel. 2017. NP recursion over time: Evidence from Indo-European. *Language* 93(4): 799–826. <https://doi.org/10.1353/lan.2017.0058>
- Yakubovich, Ilya. 2008. The origin of Luwian possessive adjectives. In Karlene Jones-Bley, Martin E. Huld, Angela Della Volpe & Miriam Robbins (eds.), *Proceedings of the 19th Annual UCLA Indo-European Conference, November 3–4, 2007*, vol. 54 JIES Monograph, 193–217.

## Résumé

On présente souvent l'utilisation de pronoms interrogatifs en fonction de marqueurs de relativisation comme un phénomène caractéristique des langues d'Europe. La distribution de ce phénomène à travers les langues de la famille indo-européenne est ici étudiée de manière empirique. L'analyse d'un échantillon représentatif de langues indo-européennes anciennes et modernes montre qu'on trouve des faits de marquage de la relativisation au moyen de mots interrogatifs à toutes les périodes de la famille indo-européenne, tant en Europe qu'à l'extérieur de l'Europe. Un examen séparé des sous-groupes de la famille indiquerait que ce type de construction est le produit d'innovations parallèles qui, par contact langagier, se sont diffusées ultérieurement. La présente étude montre par ailleurs que le caractère fléchi ou invariable d'un mot interrogatif n'est pas déterminant quant à sa capacité à servir de marqueur de relativisation.

## Zusammenfassung

Die Markierung eines Relativsatzes durch Interrogativpronomina wird oft als typisches Merkmal europäischer Sprachen betrachtet. Diese Studie präsentiert eine empirische Herangehensweise zur zeitlichen und räumlichen Verteilung von Interrogativpronomina in der indogermanischen Sprachfamilie, die auch als Relativpronomina verwendet werden. Basierend

auf einer umfassenden Stichprobe von alten und modernen indogermanischen Sprachen wird gezeigt, dass Relativsätze, die mit einem Interrogativpronomen markiert werden, in allen Zeitstufen des Indogermanischen innerhalb sowie auch außerhalb Europas zu finden sind. Eine Analyse pro Sprachzweig deutet darauf hin, dass es sich um parallele Innovationen handelt, die sich anschließend durch Sprachkontakt verbreitet haben. Die Studie zeigt auch, dass Interrogativpronomina als Relativsatzmarker verwendet werden können unabhängig davon, ob sie flektiert sind oder nicht.

## Address for correspondence

Sandra Auderset  
Max-Planck-Institut für Menschheitsgeschichte  
Kahlaische Strasse 10  
07745 JENA  
Germany  
auderset@shh.mpg.de