

The genesis of *wh*-based correlatives: From indefiniteness to relativization

Oleg Belyaev and Dag Haug
Russian Academy of Sciences and University of Oslo

Sinn und Bedeutung 2014
(Formal Theories of Meaning Change)
Göttingen, 16 September 2014

1 Introduction

Haudry (1973, 150), Heine and Kuteva (2006, 209): from indirect questions to headed RCs.

- (1) a. constituent question: *Who came?*
b. indirect question: *I don't know who came.*
c. indirect question reanalyzable as headless RC: *You also know who came.*
d. headed RC: *Do you know the woman who came?*

However, *wh*-based *correlative clauses* such as (2) from Latin cannot be said to follow this route.

- (2) *quōs ferrō trucidārī oportēbat, eōs nōndum vōce vulnerō*
which.PL.ACC by.iron be.slaughtered it.was.needed they.ACC not.yet by.voice I.injure
'I do not yet injure by the voice **them who** should have been slaughtered by the iron.'
(Haudry 1973, 156)

Cross-linguistically, *wh*-based correlatives are one strategy for forming correlative clauses. Other languages have correlatives based on demonstratives. These are widely assumed to go back to paratactic structures:

- (3) *ce min ye muru san, n ye o ye.*
man REL PST knife buy I PST him see
'(That man bought the knife, I saw him)' → 'The man who bought the knife, I saw him.'
(Bambara > Mande > Niger-Congo, Givón 2009, 98)

However, we cannot immediately transfer such a scenario to *wh*-correlatives, as shown in (4).

- (4) *What man bought the knife? I saw him.* → *I saw the man who bought the knife.*

How *wh*-correlatives develop is still an open question. In this paper, we will first look data from Indo-European, where both Hittite and Latin have correlative clauses based on the reflex of the root **k^wi-* / **k^wo-*. Then we will validate our analysis with typological data.

2 Earlier research

In many languages, including Latin and Hittite, *wh-interrogatives* double as *indefinites*. If we substitute an indefinite for the interrogative, the evolution from parataxis to a correlative seems much more straightforward.

- (5) **Some man** bought the knife. I saw **him**. → I saw **the man who** bought the knife.

Such a scenario was first proposed for Latin correlative structures by Kroll (1910),¹ with examples such as (6).

- (6) [**quae pecunia recepta erit**], **ea pecunia emere** ...
 REL.NOM.SG money.NOM.SG received.NOM.SG will.be, that.ABL.SG money.ABL.SG buy
 liceto
 shall.be.permitted
 ‘It shall be permitted to buy ... with the money that will be received.’
 ← ‘Some money will be received; with that money it shall be permitted to buy ...’

A similar explanation was offered by Hahn (1946, 1949), and from there it made it to much of the modern literature.

The paratactic scenario automatically explains the left-adjoined structure and the presence of a correlate in the matrix, both of which is characteristic for correlative constructions.

By contrast, the interrogative theory has no direct explanation for this. It must start from a headless relative clause as in (1c) and assume (implicitly or explicitly) that the headless relative clause was fronted and resumed in the main clause. This is a possible scenario that can give rise to correlative structures in languages that already have free *wh*-relatives, such as German (7).

- (7) [**Wer immer strebend sich bemüht**], **den** können wir erlösen.

However, in languages such as Latin and Hittite, all the evidence points to the priority of correlative structures.

On the other hand, Luján (2009) raises the following problems with the indefinite theory.

- There are no languages where the same pronoun is indefinite and relative but not interrogative, i.e. the semantic map is [INDEF]–[WH]–[REL] rather than [WH]–[INDEF]–[REL].
- Grammaticalization paths from (unambiguous) indefinites to relative clauses are rarely or never attested, while RC markers derived from unambiguous interrogatives are attested.

According to Luján, this suggests that it is the interrogative function of **k^wi-* / **k^wo-* that has given rise to Indo-European RCs.

3 Two key observations

3.1 The anaphoric connection

The relation between the interrogative DP and the main clause “resumptive” in old IE was anaphoric, as attested by the possibility of an indirect (bridging) referential relationship between the *wh*-phrase and the correlate as in ordinary anaphora, cf. (8) from Hittite.

1. Kroll mentions that the idea had already been foreshadowed (“gelegentlich angedeutet”) by other authors, which we have been unable to check.

- (8) [**PÍŠ** *ga-pár-ta=na=kán ku-in* A-NA DÛ EME *ši-pa-an-ta-aš*] nu ^{UZU}**NÍG.GUG**
 animal:ACC=CONJ=Ptcl REL.ACC to made tongue sacrifice CONJ intestines.ACC
^{UZU}**ZAG.UDU** *ḫa-ap-pí-ni-it za-nu-zi*
 shoulder.ACC flame.INS burn.PRS.3SG
 ‘He roasts the intestines and the shoulders of **the animal**² **which** he had sacrificed to the artificial tongue.’ (lit. ‘**What animal** he had sacrificed to the artificial tongue, he roasts **intestines** and **shoulder** with the flame.’)

(Probert 2006, 63)

This anaphoric connection does not sit well with a theory based on *wh*-interrogatives, as interrogatives do not introduce discourse referents that can be picked up in the further discourse. On the other hand, it follows naturally on the indefinite theory.

3.2 Generalizing semantics

Correlative clauses have maximalizing semantics (Grosu and Landman 1998), which means they are either generalizing (9) or definite (10).

- (9) [**cui** *testimonium defuerit*], **is** *tertiis diebus ob portum obuagulat* ito.
 REL.DAT witness is.absent, he third day to doot to.summon go
 ‘he whose witness is absent, he shall go to summon him every third day’
- (10) [**quam** *earum in iis locis optimam dicent esse*], **eam** *maxime serito*
 REL.ACC of them in these places best they.say be, that.ACC in particular sow
 ‘(the one) which they say is best in these places, sow that one in particular’

As pointed out by Sturtevant (1930), generalizing correlative clauses are much more frequent than definite ones in Hittite and early Latin. This sits well with an origin in free relatives, but does not follow from an origin in paratactic structures with indefinites.

In sum, we seem to be at an impasse, as the observations point to opposite conclusions. We will argue that we can solve this by refining the indefinite account and consider both the type of indefinite involved and the nature of the source construction. Instead of assuming a plain indefinite in a paratactic context, we will argue that correlatives arise from Hamblin indefinites in (discourse) hypotactic environments.

4 The interrogative-indefinite affinity

4.1 The Indo-European situation

The literature on correlatives has ignored the (otherwise well-known) fact that the indefinite use of *wh*-words in both Latin and Hittite is restricted to special contexts.

In Hittite, *kui-* can have an indefinite meaning “only in combination with *takku/mān* ‘if’” (Hoffner and Melchert 2008, 149).

In Latin the most common contexts are

- Conditional clauses: *si/nisi quis venit* ‘if/unless someone comes’
- Questions: *num quis venit?* ‘will someone come?’
- Negated purpose clause: *ne quis venit* ‘so that no one comes’

2. *PÍŠ ga-pár-t* is really the name of an unknown animal rather than a generic term for animals.

Restrictions on the use of the bare interrogative **k^wi-* / **k^wo-* as an indefinite are found in other early IE languages:

Vedic “Das bloss *ka* in indefinitem Sinn wird in der alten indischen Sprache in positiven Sätzen nicht gebraucht” (Delbrück 1893, 511)

Old High German Fobbe (2004, 115-116), though Latin influence is a possibility

Gothic Fobbe (2004, 115) – there is a single exception which could be due to Greek influence, and there is a separate word *sums* which is used to translate Greek *tis* in all but this single example.

By contrast, Ancient Greek allows the use of the bare interrogative as an indefinite in all contexts, including past unnegated indicatives.

- (11) *ēlthe tis*
 came who
 ‘Someone came.’

Most research in Indo-European studies assumes that Greek preserves the original situation, but this contradicts what we otherwise know about grammaticalization: items tend to bleach and be extended to more contexts rather than the opposite (see Haspelmath (1997, section 6.4) for a defense of this view in the particular domain of indefinite pronouns). Because of this it is more likely that Greek represents an innovation. Modern German has also extended the use of bare interrogatives to specific contexts, although the diachronic connection with Old High German is unclear, since bare interrogatives as indefinites are not attested in Middle High German (Fobbe 2004, 115).

4.2 Typology

The situation that we can reconstruct for old IE is not in fact restricted to these languages, but have many typological parallels. Haspelmath (1997, 173) notes that in Chinese, bare interrogatives may be used as indefinites in general questions, conditionals, negative sentences, and other contexts where they can be interpreted non-specifically, but not in ordinary declarative sentences.

Gärtner (2009), in a typological study of the indefinite-interrogative affinity, divides languages into three groups:

- + [i=i] robust indefinite/interrogative ambiguity
- % [i=i] non-robust indefinite/interrogative ambiguity
- [i=i] no indefinite/interrogative ambiguity

Are all % [i=i]-languages % [i=i]-languages in the same (or at least a similar) way? Gärtner (2009) thinks not:

Paradigmatic non-robustness In Dutch, only *wat* has an indefinite reading, not *waar* and the other indefinites. More generally, it seems that many languages do not allow indefinite readings of (counterparts to) *why* and *how*.

Distributional non-robustness According to Gärtner (2009, 6) (following Haspelmath 1997, 254), Latin *quis* is phonologically restricted. The same view is found in some Latin grammars too, e.g. Menge (2000, 135): “Sie lehnen sich v.a. enklitisch an folgende Wörter an (hinter denen sie aber nicht unmittelbar stehen müssen): ...”

Semantic non-robustness Some languages restrict the use of bare interrogatives as indefinites to a (subset of) non-specific readings. We include Latin here, as the concept of non-local enclisis is not convincing.

4.3 The semantics of restricted bare interrogatives

Is there a common core to the semantic restrictions on bare interrogative indefinites? At least, languages with semantic restrictions do not allow the bare interrogative indefinite in specific contexts, such as past non-negated indicatives. This means we are not dealing with an ordinary quantifier or choice functional indefinite.

Yanovich (2005) presents an ingenious analysis of Russian bare interrogatives where they denote Hamblin sets, just like Japanese bare indefinite roots in the analysis of Kratzer and Shimoyama (2002). For example, the clause in (12) gets the semantics in (13).

(12) (...) *kto* prišel
 who came

(13) {‘*a* came’, ‘*b* came’, ‘*c* came’, ...}

This analysis explains

- why bare interrogatives can be used as indefinites. The denotation in (13) is in fact a proper (Hamblin) denotation for a question ‘Who came’, so the semantics of the question operator is an identity function (Kratzer and Shimoyama 2002).
- why bare interrogatives are not used in unembedded contexts. The denotation in (13) is not an appropriate type for a declarative sentence, since it is a non-singleton set of propositions.

The upshot is that meanings such as in (13) must meet a licenser, i.e. some operator which is able to take a Hamblin set as an input and return a singleton set.

Such operators could be negation, conditionals, subjunctives, or question morphemes, but as Yanovich (2005) notes, the class of licensers of Hamblin pronouns is not cross-linguistically stable and we should have some account of this. But for our purposes, we have a neat explanation of the relevant facts.

5 The proposal

Our analysis crucially relies both on the anaphoric connection between the relative DP and the matrix correlate, and on Yanovich (2005) Hamblin semantics of bare interrogatives. To be able to account for the anaphoric dimension, we use a compositional version of DRT, namely partial compositional DRT (Haug 2013). There is currently no DRT analysis of Hamblin indefinites, so for simplicity we will assume standard DRT indefinite semantics for the *wh*-words. This is less problematic than it sounds: In (proto-)correlatives, the indefinite will in any case occur in a conditional, and the DRT semantics for conditionals will make sure that we unselectively quantify over all indefinites in the antecedent. This yields presumably the same effect as a Hamblin semantics would do.³ The underlying Hamblin semantics is nevertheless important to explain the limited distribution and the scope facts, and in future work we will integrate Hamblin semantics and DRT.

3. Although Yanovich (2005) notes that *if*-clauses would be among the operators licensing Hamblin definites, he does not explicitly spell out that semantics.

5.1 The source construction

The starting point for our analysis are constructions like (14)–(15).

- (14) Uvidiš' kogo — begi sjuda i bej v dver' dva raza.
 you.will.see whom run here and hit in door two times
 'If you see **someone**, run here and hit the door two times.' (lit. 'You see **whom**...')
 (Russian, <http://www.litmir.net/br/?b=172594&p=63>)
- (15) nu=kan mān AWAT NARARI kuwapi šarā išparza-zi #
 CONN=LOC.PTCL if summons assistance when up come-PRS.3SG
 'If at **some point** (lit. **If when**) a summons for assistance comes up...'
 (Hittite, NH/INS (CTH 106.A.1) Bo 86/299 rev. iii 44–46; Andrej Sideltsev, p.c.)

In both these constructions, the bare interrogative is interpreted as an indefinite in the scope of a conditional: In (14), the indefinite interpretation arises due to discourse subordination, represented in the SDRT notation (Asher and Lascarides 2003), while in (15), an explicit conditional subordinator is present, which indicates hypotaxis (though not, in the case of Hittite, embedding). Both of these possibilities may serve as a source for the wh-correlative construction. Simplified semantics are given in (16)–(17).

- (16)
- | | |
|-------------------------------|--|
| π_1, π_2 | |
| Consequence(π_1, π_2) | |
| $\pi_1 :$ | x_1
<hr/> see(you, x_1) |
| $\pi_2 :$ | <hr/> run. here. and. hit. the. door. twice(you) |
- (17)
- | | | |
|--|---------------|-----------------------------------|
| t
<hr/> a. summons. for. assistance. comes. up(t) | $\Rightarrow$ | t'
<hr/> $t < t'$
$P(t')$ |
|--|---------------|-----------------------------------|

As a first step (not further discussed here) we assume that discourse subordination may grammaticalize into syntactic subordination (hypotaxis). This is the first step in a process of syntactic grammaticalization:

- discourse subordination (parataxis) $\rightarrow$ syntactic subordination (hypotaxis) $\rightarrow$ embedding

In the rest of the paper, we focus on the semantic evolution.

5.2 The subsequent grammaticalization

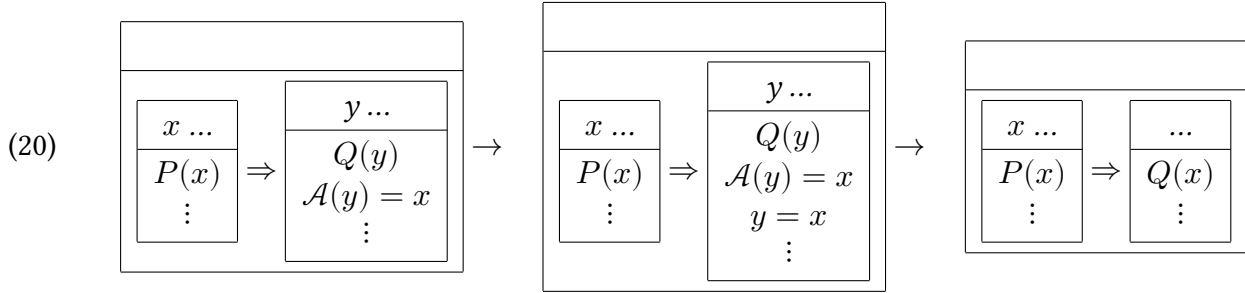
We assume the schematic source semantics in (18).

- (18)
- | | | |
|---------------------------------------|---------------|---------------------------------------|
| $x \dots$
<hr/> $P(x)$
$\vdots$ | $\Rightarrow$ | $y \dots$
<hr/> $Q(y)$
$\vdots$ |
|---------------------------------------|---------------|---------------------------------------|

Our proposal is that the structure in (18) grammaticalize into wh-correlatives by following two diachronic clines:

- (19)
- non-obligatory anaphora $\rightarrow$ obligatory anaphora $\rightarrow$ full coreference $\rightarrow$ non-anaphoric RC
 - generalizing (conditional) interpretation $\rightarrow$ definite interpretation

In semantic terms, the anaphoric cline can be represented as in (20).



$\mathcal{A}(y) = x$ means roughly “the antecedent of y is x ”, without implying full coreference (see Haug (2014) for details). Informally, the stages are

1. anaphora required, but bridging allowed
2. full coreference required,
3. variable binding instead of anaphora

As pointed out by Brasoveanu (2008), the anaphoric analysis of correlatives automatically predicts the maximalization effect, because anaphora must be exhaustive, as shown by the following examples from Kadmon (1990).

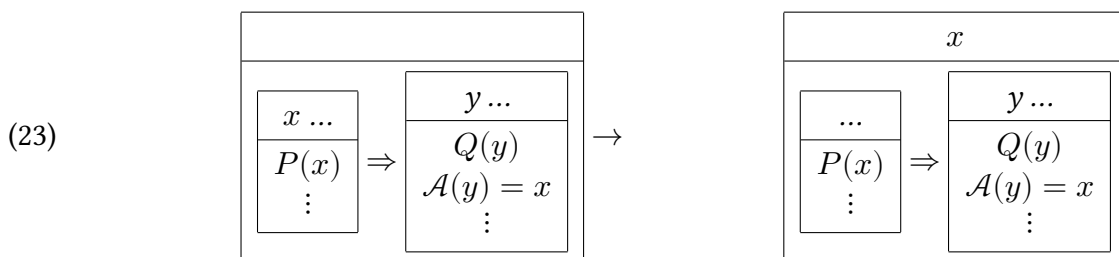
(21) Leif has a chair (Leif may have several chairs)

(22) Leif has a chair. It is in the kitchen. (Leif has only one chair)

The standard variable binding semantics of correlatives that we get in the last stage is appropriate for headed RCs: there is no anaphora, we do not predict maximalization.

The anaphoric cline is likely motivated by coherence (Asher and Lascarides 2003). The anaphoric link between the protasis and the apodosis also ensures that the two are mutually relevant. Thus an originally pragmatically motivated phenomenon becomes part of the core grammar and gives rise to true, syntactic subordination.

The second cline, from a conditional reading to a definite one, implies that the original Hamblin indefinite expands to specific contexts (in line with the trend for indefinites, see Haspelmath 1997, section 6.4), i.e. we get global discourse referent introduction as in (23).



In the source construction, the anaphora must be exhaustive in the local context (the antecedent of the conditional), which does not imply uniqueness in the global context. When the discourse referent is introduced globally, exhaustiveness entails uniqueness. The relevant transition context is probably examples such as (10), which can be construed as quantifying over a singleton set.

6 Interim conclusion

Putting all this together, we arrive at the following stages for IE correlatives.

1. Discourse subordinated conditional: *He slaughters **which** (i.e. 'some') animal, he burns a fire.*
2. Conditional correlative with obligatory anaphoric relation between the clauses: *He slaughters **which** animal, he burns the intestines.*
3. Conditional correlative with obligatory identity: *He slaughters **which** animal, he burns it.*
4. Definite/specific correlative: ***Which** animal he slaughtered yesterday, he burnt **it**.*

This theory solves the impasse we noted in section 3.2. Because the original structure is paratactic, we predict that there is a correlate in the matrix; and because the original structure is conditional, we predict that the semantics is generalizing. The synchronic similarity between conditionals and wh-correlatives is well-known, cf. Bittner (2001); Arsenijević (2009).

In an Indo-European context, our analysis sets the wh-correlatives of Hittite and Latin apart from the correlative structures found in Vedic and Homeric Greek. These are based on a stem *yo-* which is commonly thought to be demonstrative. Already Sturtevant (1930) claimed that the structures (and not just the etymology of the relative pronoun) were different and argued that **k^wi-* / **k^wo-* was used in generalizing relative clauses and *yo-* in definite ones (Sturtevant 1930, 148). We will see that Sturtevant's view – which is predicted on our analysis – is valid not just in IE, but by correlative structures more generally.⁴

7 Typological predictions

Our analysis predicts an interesting and non-trivial asymmetry:

- wh-correlatives:
 - historical pathway: univ → def
 - ⇒ must possess a universal interpretation; definite is optional (and possibly has special marking)
- dem-correlatives:
 - def → univ
 - ⇒ must possess a definite interpretation; universal is optional (and possibly has special marking)

These predictions appear to be borne out.

7.1 Languages with wh-correlatives

According to Lakshmi Bai (1985), this is exactly the situation in **Dravidian**, where participial RCs and (wh-based) correlatives are clearly distributed: the former are used in definite contexts while the latter are used in generalizing contexts.

4. In an unfortunate turn of events, Hahn (1949) changed the terminology and referred to Sturtevant's generalizing relative clauses as 'restrictive' and the definite relative clauses as 'non-restrictive', confusing future research.

- (24) a. [*nii kuṭutta*] **pen** *enkayoo veccuṭṭeen*
 you given pen somewhere displaced
 ‘I have misplaced **the pen** that you gave me.’
 b. [**enta** *payyan motalil varuvaanoo*] **avanukku** *itai kuṭuttu viṭu*
 which boy first will come to him this give
 ‘Give this **to whichever boy** comes first.’

(Tamil, Lakshmi Bai 1985, 181)

In particular, Hock (2008) notes that **Old Tamil** only allowed correlatives in clauses with conditional verbs.

According to Arsenijević (2009), correlatives in Serbo-Croatian are only generalizing and, in fact, are semantically conditional:

- (25) [**Ko** *ima magarca*], **taj** *ga i bije*.
 who has donkey that CL.ACC and beats
 ‘**Anyone who** has a donkey beats it.’

(Arsenijević 2009, 142)

The same used to be the case in the history of Basque, where, as Rebuschi (2009) argues, correlatives were in fact almost a subtype of conditionals:

- (26) Bainan [**nor-k ere** *irau-nen bai-du akhabantza-raino*], eta **hura** *iza-nen da*
 but who-ERG ever last-PROSP AUX end-until and Dem be-PROSP AUX
 salbatua.
 saved.sg
 ‘... but he that endureth to the end shall be saved.’
 (‘*Qui autem perseveraverit in finem, hic salvus erit.*’)

(Basque, Haraneder 1740, Rebuschi 2009, 94)

At the same time, there are no languages where wh-correlatives have only a definite interpretation.

In **Tibetan** (Tibeto-Burman > Sino-Tibetan), correlatives do have a definite interpretation (in addition to a generalizing one), but the language uses a morpheme otherwise meaning ‘if’ as a marker of correlatives (Cable 2009).

- (27) [**Deb cogtse gādi-i** *’og la yod na*] **sgam yang de-i** *’og la yod red*
 book table what-GEN under DAT be if box also that-GEN under DAT be AUX
 ‘The box is under **the table** the book is under.’

(Cable 2009, 205)

Another Tibeto-Burman language, **Newari**, as reported by Gärtner (2009, 13), is [%i=i], which suggests that the indefinite use of interrogatives in conditional adjunct clauses (as opposed to discourse subordination, as is the norm in Indo-European) lies at the origin of correlatives in Tibetan.

7.2 Languages with dem-correlatives

Ket (Yeniseian) has both wh-correlatives and dem-correlatives (Georg 2007; Nefedov 2012). The available data consistently show that wh-correlatives are used for generalizing RCs while dem-correlatives are used for definite RCs (Stefan Georg, p.c.):

- (28) a. [**qode** kuŋa qaj bat d{i}-asan-l-bed] **tu-de** da-o-n-a-dij
 REL thou.DAT just PtcI 1-speak-PST-make this-F 3F-PST-PST-3COREF-reach
 ‘That (woman) I was about to tell you about (just) showed up.’ (DEM-based)
 (Nefedov 2012, 218)
- b. [**ána** òn du-b-ded] **túde** bílda itelem
 who much 3-3n-read that all he.knows
 ‘Who reads a lot, knows everything.’ (WH-based)
 (Georg 2007, 174)

The same seems to obtain in **Udihe** (Tungusic > Altaic), judging from the available examples:

- (29) [**Ni** zomi: bi:-si-ni] **uti** ŋala-ni kokci-zaŋa-ni.
 who steal.PRP be-PST-3SG that hand-3SG get.crooked-FUT-3SG
 ‘He **who** steals something, **his** hand will become crooked.’ (WH-based)
 (Nikolaeva and Tolskaya 2001, 667)
- (30) Nua-ni **uti** **xoton-digi** eme-gi-e-ni [si anana **uta-du** bagdi-e-i].
 he-3SG that city-ABL come-REP-PST-3SG you long.ago that-DAT live-PST-2SG
 ‘He came from **the** same **city** you used to live in.’ (DEM-based)
 (Nikolaeva and Tolskaya 2001, 666)

In **Kita** (Mande > Niger-Congo), correlatives are DEM-based. The generalizing reading is possible, but only in the presence of the ‘if’ marker, otherwise absent:

- (31) [ídi worì **mín** tà] à dú n mà!
 2 ACP.POS money.DEF REL take 3 give 1SG POSTP
 ‘Return to me **the money that** you’ve taken!’
 (Creissels 2009, ex. 14d)
- (32) [**ní** **mín** bòyi-da] à b’ ó tà
 if REL fall-ACP.POS 3 INACP.POS DEM gather
 ‘He gathers **whatever** falls down.’
 (lit. ‘If **which** falls down, he gathers **that**.’)
 (Creissels 2009, ex. 22b)

Vedic has correlatives based on the stem *yo- commonly thought to be a demonstrative. These are often definite.

- (33) sóma [**yás** te mayobhúva ūtáyah sánti] **dāsúṣe** tábhir
 soma, REL.NOM.PL.F yours delightful means.of.help.NOM.PL.F are with.those us
 no avitá bhava
 helper be
 ‘Soma, which delightful means of help are yours, with those be our helper.’
 (RV 1.91.9)

The generalizing reading often has special marking in the form of a subjunctive.

- (34) [**yás** túbhyaṃ dāsān ná] **tám** áṃho aśnavat
 REL.NOM.SG you.DAT honour.3.SG.SBJV not that.ACC fear reach.3.SG.SBJV
 ‘Whoever honours you, fear shall not reach him.’
 (RV 2.23.4b)

Similar facts hold in Ancient Greek, although in this language the correlative clause has mostly been replaced by canonical relatives even in the earliest attested stages. There is a clear contrast with Latin, which has *wh*-correlatives and does *not* use the subjunctive in generalizing relative clauses.

Sanuma (Yanomam) is another language which consistently utilises DEM-based correlatives for relativization. Unfortunately, no clear statements are found in the literature concerning the available interpretation. Yet the examples of correlatives in this language only involve definite interpretations:

- (35) [**ipa ulu a** lo-le], **i niha** sa hapalo ke
 my son 3:SG sit-PRS REL to 1:SG speak IP
 ‘I just now spoke **to my son who** is sitting here.’

(Borgman 1990, 135)

7.3 A micro-typology of different stages of grammaticalization

The two clines shown in (19) are not logically related to each other, and languages may instantiate any combination of the stages shown.

7.3.1 [–an;–def]

This is the case in modern Standard **Russian**,⁵ which neither has definite correlatives nor obligatory anaphoric resumption.

- (36) a. [**Kto** ne rabotaet], **tot** est.
 who not works that eats
 ‘Whoever does not work eats.’
 b. # [**Kogo** ty videl], **tot / on** moj drug.
 whom you say that he my friend
 (‘The one you saw is my friend.’)

(movie “Operaciya Y”)

Another example of such a language is **Ashti** (Dargwa > North-East Caucasian). Here, correlatives are universal-only, and the language is [%i=i]:

- (37) a. [**si** **lu**rat-li-ži-w **ɤaj-ka:k’-ul=il**], **te:** **ummat-la**
 what language-OBL-SUPER-M[ESS] speech-DOWN:M:speak.PRS=IQ that nation-GEN
 w-i:χ^w-a: har-il insan
 M-be.IPFV-HAB.3 every-ATTR person
 ‘Every person speaks the language of whatever nation he comes from.’
 (lit. ‘Which language he speaks, from that nation is every person.’)
 b. * [u-dil juz **si** **admi-li-j** b-ik:-a-t:i=il], **te:t** di-la uci
 thou-ERG book what person-OBL-DAT N-give.PFV-PRET-2=IQ that me-GEN brother
 sa-w
 COP-M
 (‘The man to whom you gave the book is my brother.’)

5. Spoken Russian does seem to have definite correlative RCs, but these are poorly studied. It appears that they may arise due to performance issues rather than as a genuine grammatical construction — akin to resumptives in English.

- c. [ča saq'ʕ-an-ni=ja], unc:a ač-b-a:q'-a
 who come.PFV-3-COND=ADD door open-N-do.PFV-IMP[SG]
 'If someone comes, open the door.'
 (lit. 'If **who** comes, open the door.')

(own fieldwork)

Old Church Slavonic was [%i=i], but at the same time, did not yet possess wh-correlatives (these developed in Slavic later; OCS had the *yo- series of pronouns):

- (38) ašte kūto xošetü po mñe iti ...
 if who wants after me go:INF
 'If **anyone** wants to come after me ...'

(Haspelmath 1997, 136)

For other Indo-European examples and references, see Haspelmath (1997, 136–137). In general, any language which can be described as being [%i=i] according to Gärtner (2009) but lacks grammaticalized correlatives goes here. Gärtner himself provides the following list of [%i=i] languages (p. 13):

- Belorussian, Chinese, Dutch, Latin, Lithuanian, Newari, Eastern Pomo, Russian, Old Church Slavonic, Slovene, Ukrainian

It is not clear which of these have correlatives and whether these correlatives can be definite, but e.g. **Dutch** apparently lacks wh-correlatives, which makes it another candidate.

7.3.2 [+an;–def]

This includes languages which are not [%i=i] but do have correlatives, but the correlatives only have a universal interpretation.

Dravidian languages and **Basque**, illustrated above, are examples of this type.

7.3.3 [–an;+def]

Such languages have already developed a definite interpretation, but have not yet grammaticalized an obligatory anaphoric relation.

Hittite is also an example of this type, cf. (15). Hittite correlatives could also have a definite interpretation (REFs).

Kryz (Lezgi > North-East Caucasian) correlatives have a definite interpretation if used with an indicative verb form (Authier 2009, 379–380), but a generalizing one if used with a conditional verb form (Authier 2009, 376). However, in conditional sentences we find examples where there is no correlate in the main clause:

- (39) a. [hal-id q'ay-cu] a-d an q'ay-i!
 whose-notN die-PRF.F 3-notN also die-OPT
 'Let **the one because of whom** she has died also die!'
 b. [hal-ir riki ġe-t'a-na], nabada riki ačuğ yi-yu vun!
 who-ERG door PV-strike-COND God.forbid door open do-DEONT 2
 'Whoever knocks at the door, God forbid you to open it!'

(Authier 2009, 377)

Another example is **Udmurt** (Finno-Ugric > Uralic):

- (40) [**kin kitin** ule], mi kwin'ez — tatin, gurtin
 who where lives us two there in.village
 [There are four of us: three brothers and one sister] 'Who lives where, the two of us — here, in the village.'
 (Udmurt Corpus⁶: *Udmurt dunne*, 2009.11.11)
- (41) [**kin' mar-e** b̂gat-e], s-e=ik vuza.
 who what-ACC be.able-PRS.3SG that-ACC=Ptcl sell[PRS.3SG]
 'They sold, what they could.'
 (lit. "Who could what, that they sold.")
 (Besermen dialect, spoken text⁷: *Sverdlovsk* 1.56)

7.3.4 [+an;+def]

This type includes languages where wh-correlatives have a definite interpretation, but where the use of interrogatives as indefinites in conditionals is no longer allowed without a main-clause correlate.

An example of such a language is **Ossetic** (Iranian > Indo-European):

- (42) a. [**sə** lɜp:u-jə fed-t-aj], wəj u me= fšəmɜr
 what boy-GEN see.PFV-TR-PST.3SG DemDist[NOM] be.PRS.3SG 1SgPoss brother
 'The boy you saw is my brother.'
- b. [kʷəšt-mɜ čī ne rba-səd-i], wəm-ɜn ɜχsa nɜ
 work-ALL who.NOM NEG PV-go-PST.INTR.3SG DemDist-DAT money NEG
 rac:ən-ɜn
 give.PFV:FUT-1SG
 'Whoever didn't come to work, I won't give money to him/her.'
- c. * [kɜd kʷəšt-mɜ čī ne rba-səd-i], wəm-ɜn ɜχsa nɜ
 if work-ALL who.NOM NEG PV-go-PST.3SG DemDist-DAT money NEG
 rac:ən-ɜn
 give.PFV:FUT-1SG
 ('If someone (lit. "who") didn't come to work, I will not give money to him/her.')
- (own fieldwork, cf. also Belyaev (2014))

Surprisingly, clear examples of such languages are comparatively rare. Languages that have wh-correlatives seem to normally allow using interrogatives as indefinites in conditional contexts in one way or another, which provides strong support for our hypothesis.

Further progress along these lines may lead to the anaphoric semantics being abandoned altogether, giving way to the set-intersection semantics that is typical for canonical RCs. Such, apparently, is the situation in Hindi (Indo-Aryan > Indo-European)⁸, where only repetition of the same noun or expressives are allowed in "double-headed" correlatives:

- (43) [**jis aadmii-se** tum bahut pyaar-se baat kar rahe the], **us** ^{OK} **aadmii-ne** /
 REL man-with you much love-with talk do PROG.PL be.PST.PL that man-ERG
 * **Tiicar-ne** / ^{OK} **haraamii-ne** mujh-pe muqadma Thonk rakhaa hai
 teacher-ERG bastard-ERG me-on court.case 'apply'-PFV be.PRS.SG

6. http://web-corpora.net/UdmurtCorpus/search/?interface_language=en

7. The Besermen texts have been analyzed by Olga Biryuk. Some of them are available online: <http://corpora.iling-ran.ru/>

8. Hindi itself has historically demonstrative-based correlatives. There is not enough data available for languages with wh-correlatives in order to say anything about the possibility of bridging.

‘The man that you were talking with so nicely, that ^{OK} man / * teacher / ^{OK} bastard is suing me.’

(Potts et al. 2009)

8 Conclusion

- *wh*-correlatives grammaticalize from bare interrogatives used as Hamblin indefinites in conditional contexts
- The conditional context can either be discourse subordination or overtly marked with a conditional subordinator
- There are two clines of grammaticalization: anaphoric integration and “scope raising” of the indefinite
- The distinct origin of *wh*-correlatives predicts some interesting asymmetries with demonstrative-based correlatives, and these predictions are borne out typologically

List of glosses

ABL: ablative; ACC: accusative; ADD: additive particle; ALL: allative; ATTR: attributive; AUX: auxiliary; COND: conditional; CONJ: conjunction; COP: copula; DAT: dative; Dem: demonstrative; Dist: distal deixis; DOWN: marker indicating downwards movement; ERG: ergative case; ESS: essive orientation; F: feminine gender; FUT: future; GEN: genitive; HAB: habitual; IMP: imperative; INF: infinitive; INS: instrumental; INTR: intransitivity marker; IP: immediate past; IPFV: particle; IQ: indirect question; M: masculine; N: neuter gender; NEG: negation; NOM: nominative; OBL: oblique; OPT: optative; PFV: perfective; PL: plural; Poss: possessive pronoun; PRET: preterite; PRF: perfect; PROG: progressive; PROSP: prospective; PRS: present; PST: past tense; Ptcl: particle; PV: preverb; REL: relativization marker; SG: singular; SUPER: superessive-superlative; TR: transitivity

References

- Arsenijević, B. 2009. “[Relative {conditional {correlative clauses}}].” In *Correlatives cross-linguistically*, 131–156. Amsterdam and Philadelphia. Asher, N., and A. Lascarides. 2003. *Logics of conversation*. Authier, G. 2009. *Grammaire kryz, langue caucasique d’Azerbaïdjan, dialecte d’Alik*. Leuven and Paris. Belyaev, O. I. 2014. “Korreljativnaja konstrukcija v osetinskom jazyke v tipologičeskom osveščanii.” Cand. thesis, Lomonosov Moscow State University. Bittner, M. 2001. “Topical Referents for Individuals and Possibilities.” In *Proceedings of SALT 11*, 36–55. Ithaca. Borgman, D. M. 1990. “Sanuma.” In *Handbook of Amazonian languages*, 2:15–248. Brasoveanu, A. 2008. “Uniqueness effects in correlatives.” In *Proceedings of Sinn und Bedeutung 12*, 47–65. Oslo. Cable, S. 2009. “The syntax of the Tibetan correlative.” In *Correlatives cross-linguistically*. Amsterdam and Philadelphia. Creissels, D. 2009. “Les relatives corrélatives: le cas du malinké de Kita.” *Langages* 174:39–52. Delbrück, B. 1893. *Vergleichende Syntax der Indogermanischen Sprachen*. Vol. 3, bk. 1 of *Grundriss der vergleichenden Grammatik der Indogermanischen Sprachen*. Strassburg. Fobbe, E. 2004. *Die Indefinitpronomina des Deutschen. Aspekte ihrer Verwendung und ihrer historischen Entwicklung*. Heidelberg. Gärtner, H.-M. 2009. “More on the indefinite-interrogative affinity: the view from embedded non-finite interrogatives.” *Linguistic Typology* 13:1–37. Georg, S. 2007. *Introduction, phonology, morphology*. Vol. 1 of *A descriptive grammar of Ket*. Folkestone. Givón, T. 2009. *The genesis of syntactic complexity: diachrony, ontogeny, neuro-cognition, evolution*. Amsterdam and Philadelphia. Grosu, A., and F. Landman. 1998. “Strange Relatives of the Third Kind.” *Natural Language Semantics* 6 (2): 125–170. Hahn, E. A. 1946. “The origin of the relative *k^wi- k^wo-*.” *Language* 22:68–85. ———. 1949. “The non-restrictive relative clause in Hittite.” *Language* 25:346–374. Haspelmath, M. 1997. *Indefinite pronouns*. Oxford. Haudry, J. 1973. “Parataxe, hypotaxe et corrélation dans la phrase latine.” *BSL* 68 (1): 147–186. Haug, D. T. T. 2013. “Partial dynamic semantics for anaphora: compositionality without syntactic coindexation.” *JoS*. ———. 2014. “The anaphoric semantics of partial control.” In *Proceedings of*

SALT 24. Heine, B., and T. Kuteva. 2006. *The changing languages of Europe*. Oxford. Hock, H. H. 2008. "Dravidian syntactic typology: a reply to Steever." In *Annual Review of South Asian Languages and Linguistics 2008*, 163–200. Amsterdam and Philadelphia. Hoffner, H. A. J., and H. C. Melchert. 2008. *Reference Grammar. A Grammar of the Hittite Language*. Winona Lake, IN. Kadmon, N. 1990. "Uniqueness." *Linguistics and Philosophy* 13:273–324. Kratzer, A., and J. Shimoyama. 2002. "Indeterminate pronouns: the view from Japanese." In *The proceedings of the Third Tokyo Conference on Psycholinguistics*. Tokyo. Kroll, W. 1910. "Der lateinische Relativsatz." *Glotta* 3 (1): 1–18. Lakshmi Bai, B. 1985. "Some notes on correlatives constructions in Dravidian." In "For Gordon H. Fairbanks," *Oceanic Linguistics Special Publications*, no. 20:181–190. Luján, E. R. 2009. "On the grammaticalization of *k^wi/*k^wo in Proto-Indo-European." In *Grammatical change in Indo-European languages*, 221–236. Amsterdam and Philadelphia. Menge, H. 2000. *Lehrbuch der lateinischen Syntax und Semantik*. Völlig neu bearbeitet von Thorsten Burkard und Markus Schauer. Darmstadt. Nefedov, A. 2012. "Relativization in Ket." In *Clause linkage in cross-linguistic perspective*, 1–36. Berlin. Nikolaeva, I., and M. Tolskaya. 2001. *A grammar of Udihe*. Berlin and New York. Potts, C., A. Asudeh, S. Cable, Y. Hara, E. McCready, L. Alonso-Ovalle, R. Bhatt, et al. 2009. "Expressives and identity conditions." *Linguistic Inquiry* 40 (2): 356–366. Probert, P. 2006. "Clause boundaries in Old Hittite relative sentences." *TPhS* 104 (1): 17–83. Rebuschi, G. 2009. "Basque correlatives and their kin in the history of Northern Basque." In *Correlatives cross-linguistically*, 81–130. Amsterdam and Philadelphia. Sturtevant, E. H. 1930. "Relatives in Indo-European and Hittite." In *Curme Volume of Linguistic Studies*, 141–149. Baltimore. Yanovich, I. 2005. "Choice-functional series of indefinite pronouns and Hamblin semantics." In *Proceedings of SALT 15*, 309–326. Ithaca.