

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

In the recent literature considerable attention has been paid to the diachronic pathways of wh-based subordination. The most widely accepted hypothesis since at least Haudry (1973, 150), elaborated in Heine and Kuteva (2006, 209), assumes a pathway from indirect questions through headless relative clauses:

- (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 relative clause: *Do you know the woman who came?*

But this explanation clearly does not work for the grammaticalization of wh-based *correlative clauses* such as (2) from Latin. Such clauses are widespread in many Indo-European languages, and are generally assumed to be diachronically prior to canonical, headed relative clauses.

- (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)

Correlatives in early IE languages are anaphoric in nature and admit an indirect (bridging) referential relationship between the wh-phrase and the correlate as in ordinary pronominal anaphora, cf. (3) from Hittite.

- (3) *PÍŠ ga-pár-ta=na=kán ku-in A-NA DÛ EME ši-pa-an-ta-aš nu UZUNÍG.GUG*
 animal:ACC=CONJ=PTC REL.ACC to made tongue sacrifice CONJ intestines.ACC
UZUZAG.UDU ḥa-ap-pí-ni-it za-nu-zi
 shoulder.ACC flame.INSTR burn.PRS.3S
 'He roasts the intestines and the shoulders of **the animal**¹ **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.')

The anaphoric nature of the relationship suggests that interrogative-based correlatives have the same origin as widely assumed for the (cross-linguistically more common) demonstrative-based correlatives. This is thought to involve a parataxis-like construction, cf. (4), from Bambara (Givón 2009, 98).

- (4) *ce min ye muru san, n ye o ye.*
 man REL PAST knife buy I PAST him see
 ('**That man** bought the knife, I saw **him**' →) 'The man **who** bought the knife, I saw **him**.'

But we cannot simply transfer this analysis to wh-based correlatives, because interrogatives do not introduce discourse referents that could act as antecedents to pronominals or definite descriptions. The key to solving this problem lies in the fact that those IE languages that have wh-based correlatives also allow the use of wh-interrogatives as *indefinites* in discourse subordinated contexts.

- (5) *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>)
- (6) *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.)

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

This use of interrogatives as indefinites without additional marking is particularly prone to occur in conditional contexts. Crucially the conditional need not be syntactically indicated but can arise from semantic subordination as in (5), cf. Culicover and Jackendoff (1997).

The interrogative-indefinite semantic connection is well-known (Kratzer and Shimoyama 2002), but most Indo-European languages (Russian and Hittite among them) do not allow using bare interrogatives as indefinites without additional marking in independent sentences or paratactic contexts. Hence, the most plausible assumption is that wh-correlatives have developed from constructions like (5). This is strengthened by the fact that the few non-Indo-European languages which have developed wh-correlatives also show traces of using interrogatives as indefinites in subordinate clauses. In particular, Udmurt (Uralic) allows sentences like (5) (Belyaev 2012), while Tibetan (Sino-Tibetan) correlatives obligatorily include the conditional marker *na* (Cable 2009). In both of these languages, interrogatives cannot be used as indefinites in independent sentences.

Therefore, a plausible scenario of the development of wh-based correlatives is the following:

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. Definite/specific correlative: ***Which** animal he slaughtered yesterday, he burnt **it**.*

We implement our semantic analysis of these stages in an extended version of PCDRT (Haug 2013). Anaphoric relations are captured in terms of a function \mathcal{A} taking anaphors to antecedents. Crucially, to capture bridging, we do not require that anaphora entails full coreference. Schematic meaning representations are given in (7).

- | | |
|--|--|
| (7) a. $[[x P(x)] \Rightarrow [y Q(y)]]$ | discourse subordination without anaphora |
| b. $[[x P(x)] \Rightarrow [y Q(y), \mathcal{A}(y) = x]]$ | generalizing correlative |
| c. $[P(x), Q(y), \mathcal{A}(y) = x]$ | specific correlative |

In other words, the generalizing correlative arises out of conditional discourse subordination via the grammaticalization of a requirement that the indefinite in the protasis is anaphorically referenced in apodosis. Later, this is generalized to specific contexts. Subsequent grammaticalization may yield total coreference (as opposed to a loose anaphoric link) and hence a canonical relative.

An interesting prediction of this analysis is that the generalizing reading of wh-correlatives must be primary since the use of interrogatives as indefinites requires discourse subordination as in (5) and therefore the conditional reading. This means that specific/definite readings must be later developments and we predict that there are no languages where wh-correlatives only have a definite reading. This prediction is borne out in our data.

The larger question, of course, is *why* the requirement arises that the indefinite in the protasis must be picked up in the apodosis. We conjecture that this is motivated by discourse 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.

Asher, N., and A. Lascarides. 2003. *Logics of conversation*. Belyaev, O. I. 2012. “Korreljativnaja konstrukcija i konstrukcija odnositel’nogo predloženiya s vnutrennej veršinoj v besermjanskom dialekte udmurtskogo jazyka.” In *Finno-ugorskie jazyki: fragmenty grammatičeskogo opisanija: Formal’nyj i funkcional’nyj podxody*. Moscow. Cable, S. 2009. “The syntax of the Tibetan correlative.” In *Correlatives cross-linguistically*. Amsterdam and Philadelphia. Culicover, P. W., and R. Jackendoff. 1997. “Semantic Subordination despite Syntactic Coordination.” *LI* 28 (2): 195–217. Givón, T. 2009. *The genesis of syntactic complexity: diachrony, ontogeny, neuro-cognition, evolution*. Amsterdam and Philadelphia. 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*. Heine, B., and T. Kuteva. 2006. *The changing languages of Europe*. Oxford. Kratzer, A., and J. Shimoyama. 2002. “Indeterminate pronouns: the view from Japanese.” In *The proceedings of the Third Tokyo Conference on Psycholinguistics*. Tokyo. Probert, P. 2006. “Clause boundaries in Old Hittite relative sentences.” *TPhS* 104 (1): 17–83.