Interrogative–indefinites and k^w-relatives in the history of English and French

Pathways in a locked room

Rob Truswell, University of Edinburgh
`rob.truswell@ed.ac.uk`

(Joint work with Nik Gisborne)

Romanistentag, Kassel, 2/10/19

Dramatis personae

- ▶ Interrogative forms: *Où est-il allé?*
- ▶ Use of the same forms in headed relatives: *du lit* **ou** *elle etoit* (morin,.3432)
- ▶ Use of the same forms as indefinites: *Se jo truis* **o***, mult grant bataille i ert.* (1100-roland-v,193.2648)
- ▶ Other uses (free relatives, exclamatives, etc.) not discussed in detail today.

- ▶ Among relativizers, we distinguish:
    - ▶ relative complementizers (heads, no piedpiping, no connectivity, often monomorphemic, but including *que*/*qui* alternation).
    - ▶ relative specifiers (phrasal, allow piedpiping, show connectivity).

In Present-day French

- ▶ French relativizers are all etymologically interrogative, but some are complementizers and some are specifiers.
- ▶ *Que* is a canonical relative complementizer: monomorphemic, invariant.
- ▶ *Où* and the *lequel*-series are relative specifiers (e.g. phrasal, allow piedpiping of P).
- ▶ *Dont* is complicated: it's invariant and synchronically monomorphemic, but (unusually) substitutes for PPs. Assume it's a complementizer.
- ▶ *Qui* is unreasonably complicated: it alternates with *que* (subject/nonsubject), so with simple NP gaps is a complementizer, but also piedpipes prepositions as an animate form (*de qui* etc.), so with other gaps is (part of) a specifier.

In Present-day English

- ▶ Relative complementizer *that* (monomorphemic, invariant).
- ▶ Zero-relatives.
- ▶ A series of *wh*-specifiers, most major *wh*-words except *what* (in most idiolects).

All of this is very unusual

	IE	Other
Spec	25 (62.5%)	8 (6.1%)
Int	*16 (40%)*	*3 (2.3%)*
Dem	*4 (10%)*	*0 (0%)*
Sp	*5 (12.5%)*	*0 (0%)*
No Spec	15 (37.5%)	124 (94%)

Table 1: Headed relative specifiers in 172 languages (based on De Vries 2002)

▶ Research question: how do IE languages keep ending up in this sparsely populated typological space?

Section 1

K^w-relatives: Diachronic typology

Candidate explanations

Not just inheritance from PIE

▶ Proto-Indo-European probably did not have headed k^w-relatives (it probably did not have embedded relatives at all, Kiparsky 1995, Clackson 2007, *pace* Probert 2014).

Not just contact

▶ Comrie (1998) identifies relative pronouns (including k^w-relatives) as a European areal type, whose distribution is explained by contact.

▶ But Indo-Aryan languages have 'European-type' relatives, and are not in Europe.

▶ Individual relative pronoun systems are different, so direct borrowing is unlikely.

A hybrid: Parallel evolution

1. A distinctive initial state;

2. Something to motivate a recurring path for emergence of interrogative relatives.

3. Borrowing possible in principle at any stage on this pathway (see e.g. Probert 2014).

▶ Parallel endogenous innovations seem unparsimonious, but they do happen.

(1)   de fout   wie hun eigenlijk maken
      the mistake who they actually make
      'the mistake which they actually make'

                                    (Johan Cruyff, via Boef 2012)

(2)   adnominal adjectives (those who are not modifying the noun
      predicatively)   (Belk 2016: 179)

▶ Several independent innovations of headed k^w-relatives attested or reconstructed in the literature.

Roadmap

Key claims in this talk:

- ▶ K^w-relatives emerge diachronically from uses of interrogative–indefinites in conditionals. (Section 2)
- ▶ The distinctive properties of the initial state involve left-adjoined structures ('diptyque normal', e.g. conditionals, correlatives) and interrogative–indefinite k^w-forms. (Section 2)
- ▶ But different languages follow different diachronic pathways within this broad template. (Section 3)
- ▶ So different IE languages are following different pathways within a narrowly confined 'locked room'. (Section 4)
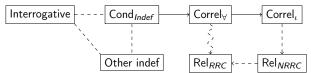
Section 2

The initial state

Conditional/correlative → relative

- Haudry (1973), etc.: headed k^w-relatives are diachronically descended from early IE k^w-correlatives.

(3) qui ..., is ... → [vir qui] ..., is ... → vir, [qui ...], is ...

- Belyaev & Haug (2014): all languages with k^w-correlatives allow generalizing reading, not all allow definite reading.

(4) cui   testimonium defuerit, is tertiis diebus ob portum
     REL.DAT witness   is.absent he third day   to door
     obuagulatum ito.
     to.summon   go
     'He whose witness is absent, he shall go to summon him every third day.'

(5) quam   earum in iis   locis optimam dicent esse, eam
     REL.ACC of.them in these places best   they.say be   that.ACC
     maxime   serito
     in.particular sow
     '(The one) which they say is best in these places, sow that one in
     particular.'   (Belyaev & Haug 2014)

Conditional/correlative → relative

▶ Andrews (1975): systematic ambiguity between generalizing correlative / conditional + indefinite.

(6) yasya   yat   paitṛkam   ritkam   sa
     who.GEN what.NOM paternal.NOM inheritance.NOM he.NOM
     tad   grhnīta,   netaraḥ
     that.ACC should.get not.another
     'Of whom what is the paternal inheritance, he should get it and
     not somebody else.'
     'If someone has something as a paternal inheritance, then he
     should get it and not someone else.'   (Sanskrit)

▶ This gives a way of linking indefinite interpretations of k^w- to relative structures:

▶ If this map is accurate, then the important properties of early IE revolve around conditionals, correlatives, and k^w-indefinites.

More on k^w-

- Early IE k^w- forms are interrogative–indefinites.
- But this is not that rare.
- Haspelmath (1997): 62 languages in 100-language sample have interrogative-based indefinites.
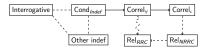- 28 (?31) allow bare interrogative–indefinites (same form with both functions).

K^w- and conditionals

- ▶ Early IE interrogative–indefinites seem to have been dependent indefinites.
- ▶ Particularly common in the antecedent of conditionals.
    - ▶ 50% of Old English bare *hw*-indefinites;
    - ▶ 'one of the favourite contexts of *quis*' (Pinkster 2015: 1104)

  (7) gif hwa hit bletsað . . .
       if who it blesses
       'If anyone blesses it, . . .'

  (8) Metum vero si qui sustulisset, omnem vitae diligentiam sublatam
       fore . . .
       'If indeed anyone succeeded in getting rid of fear, the careful
       conduct of life . . . would be got rid of entirely.'

       (Pinkster 2015: 1164)

- ▶ We don't know how rare this is, but note that almost all bare interrogative–indefinites in Haspelmath's (1997) 100-language survey are translated with 'some', not 'any'.

Frontability

▶ Luján (2009) develops a different map: $\boxed{\text{Indef}} - \boxed{\textbf{Int}} - \boxed{\text{Rel}}$

▶ Compare ours:

▶ Working hypothesis: Luján's map doesn't directly reflect a diachronic pathway, but rather the frontability of different k^w-forms in different contexts.

  ▶ Indefinites: typically (not always) unfronted;
  ▶ Interrogatives: sometimes fronted, sometimes not (marked by intonation, question particle, etc.)
  ▶ Relatives: always fronted (Downing 1978, but see De Vries 2002).

Frontability and typology

- *Wh*-fronting isn't rare; an interrogative–indefinite ambiguity isn't rare; but it's rare to find both in the same language.
- I cross-checked the data on the 'i=i' ambiguity in Gärtner (2009) against the data on *wh*-fronting in Dryer (2013). Results (from an unbalanced sample of 48 languages) below.

	Fronted *wh*	in situ *wh*
i=i	12	22
i≠i	10	4

Table 2: Interaction of *wh*-fronting and indefinite/interrogative ambiguity ($p = 0.03$, Fisher's exact test)

Interim summary
Licensing dependent indefinites

▶ Dependent indefinites need to be paired with an operator.
▶ Three main ways of identifying an operator:
  1. Affixation;
  2. Independent particle;
  3. Movement.
▶ All three widely attested with IE k^w-forms.
  ▶ Affixation gives the various series of indefinite pronouns; *quiconque* etc.
  ▶ Independent particles occur when bare indefinites are used in conditionals, etc.: *si quis* etc.
  ▶ Movement common in questions and required in relatives.

Interim summary
Dependent indefinites in the initial state

- ▶ Early Indo-European seems to have had:
  - ▶ Interrogative words which were also used as dependent indefinites.
  - ▶ Positional marking of certain functions of those words (always relatives, sometimes interrogatives).
  - ▶ The 'diptyque normal': use of left-adjunction structures to mark conditionals, concessives, correlatives, topic–comment structures, etc.
- ▶ Although interrogative–indefinites are common, bare dependent interrogative–indefinites are rarer, and the combination with k^w-fronting rarer still.
- ▶ This constellation of rare features, all pertinent to the B&H/Haudry diachrony, can help explain why more languages haven't developed k^w-relatives.

Section 3

Different paths to the same endpoint

Summary, key points

1. Latin and OE have nontrivial structural similarities, plausibly attributable to their common ancestor.
2. Latin and OE are not structurally identical in relevant respects. They have already diverged, while retaining the cluster of IE properties identified above.
3. K^w-forms can develop new uses as headed relativizers along parallel pathways — and divergent pathways! — in these two branches of IE (Romance here represented by French).

Latin and OE: Similarities

- ▶ Both have several series of interrogative-based indefinite pronouns and determiners.
  - ▶ *quilibet*, *quicumque*, *aliquis*, ...
  - ▶ *gehwa*, *locahwa*, *hwahwugu*, ...
- ▶ Both use bare k^w-indefinites in the scope of operators including conditionals. Both at least disprefer bare k^w-indefinites in matrix affirmative declaratives.
- ▶ Both already have headed relatives.
- ▶ Both use the 'diptyque normal' in constructions including conditionals and generalizing correlatives.
- ▶ Neither has multiple k^w-correlatives (which are the clearest evidence against a free relative analysis). A few examples in Late Latin (Philomen Probert, p.c.), none in English.

  (9) quis quantum credidit sibi dari tantum gratiae Dei miserantis accepit
  'Who believed how much to be given to himself, that much of merciful God's grace he received'
  (Tyconius, *Liber regularum* 3.15, 4th century CE)

Latin and OE: Differences

▶ Different sets of environments license bare k^w-indefinites.

(10)   An, opsecro hercle, habent quas gallinae manus?
Really, I ask you, have hens got hands of some sort?
<div align="right">(Pinkster 2015: 1104)</div>

(11)   þeah   hwa mæge ongitan hwæt oðer   do,   he ne   mæg
though who may   perceive what another does, he NEG may
witan hwæt he þencð
know what he thinks
'Although someone may perceive what someone else does, he
cannot know what they think.'   (coboeth,Bo:39.132.1.2620)

▶ Latin (sometimes) allows generalizing and definite correlatives.
OE only allows generalizing interpretations.

▶ The two languages use the left-adjoined position for different
functions (OE doesn't use it for topical individuals).

▶ Latin *already* has headed k^w-relatives. OE headed relatives
contain complementizer *þe* and/or inflected demonstratives.

Further divergence

- Latin may have developed multiple k^w-correlatives, unlike English (see above).
- French has mainly reanalysed the animacy distinction of *qui*/*que* as a subject/nonsubject distinction in relatives. English hasn't (but compare *whose*).
- English has innovated a (not quite categorical) distinction between *what* (free relatives) and *which* (headed). French hasn't.
- Many other examples with conditionals, unconditionals, etc.
- So there is nothing deterministic about change in this area, or about which structures will emerge.

But: *Quel* and *which*

- ▶ The k^w-series in the two branches include forms (*quel* and *which*) which do not 'fit' (are not restricted to one of the major ontological categories in the way that e.g. *who* or animate *qui* canonically are).

- ▶ When headed k^w-relatives first emerge, *quel* and *which* are not used.

- ▶ They begin to be used as relative specifiers at similar times.

    (12)   c' est lor cors senz les=quels il ne puyent estre perfait. (1190-SBERNAN-P-BFM,31.1010)

    (13)

þetindes	[...]	bi	hwicche	me	climbeð	to	þe	blisse	of
the.stakes		by	which	man	climbs	to	the	bliss	of

heouene
heaven
'the rungs by which one climbs to the bliss of heaven'
(cmancriw-2-m1,II.261.268, c.1225)

Not borrowing: Parallel change

▶ See Sakalauskaite (2016) for evidence that *(le)quel* and *(the) which* are used in different contexts, even in translations.

▶ The grammar of *lequel*-relatives is not the grammar of *the which*-relatives.

▶ *Lequel* is always arguably nonrestrictive. *The which* is always nonrestrictive when it has a complement NP, but can be restrictive otherwise.

> (14)   For if we luf   God in al oure hert, þar   es na thyng in us
> for if we love God in all our heart there is no thing in us
> thurgh   þe whilk we serve to syn.
> through the which we serve to sin
> 'For if we love God with all our heart, there is nothing in us
> through which we serve to sin.'   (cmrollep-m24,110.794)

▶ Subtly different, but broadly similar, near-simultaneous emergence of a typologically rare construction in two neighbouring and genetically related languages.

Section 4

Conclusion: The locked room

Why isn't there more divergence?

- ▶ We have seen that IE languages have followed a range of pathways from a common starting point.
- ▶ And yet, an IE relative 'type' has emerged.
- ▶ This seems like a paradox.
- ▶ We think that the resolution of the paradox comes from the fact that *wh*-phrases are confined within a very limited semantic space.
- ▶ Moreover, the interpretation of *wh*-phrases is mutable and sensitive to the local context.
- ▶ This increases the odds of different pathways emerging, but converging on the same space.

Choice points and pathways

- ▶ A canonical correlative is a way of saying two things about a single entity.

- ▶ A canonical conditional is a way of saying two (causally related) things about a single situation.

- ▶ But 'situations' can correspond to individuals (Elbourne 2001).

  (15) a. If a bishop meets a bishop, he blesses him.
        b. $\forall s.[\exists b_1.[\text{bishop}(b_1, s) \wedge \exists s' \supseteq s.[\text{bishop}(b_2, s') \wedge \text{meets}(b_1, b_2, s')]][\exists s'' \supseteq s'.[\text{bless}(b_1, b_2, s'')]]]$

- ▶ The communicative intentions of these structures are often clearer than the compositional routes through which those interpretations are indicated.

- ▶ This is fertile ground for reanalysis.

- ▶ The emergence of headed k^w-relatives is a likely outcome of that reanalysis, because they are yet another way of saying two things about a single individual.

The locked room

- ▶ The semantic space occupied by IE indefinite–interrogatives is a limited and largely encapsulated one.
- ▶ But still, the PIE initial state only occupies a subpart of this space.
- ▶ There are many ways to grow from this initial state, but the growth will tend to be in similar directions.
- ▶ The result is repeated independent emergence of similar, but not identical, constructions.

References I

Andrews, A. (1975). *Studies in the Syntax of Relative and Comparative Clauses*. PhD thesis, Massachusetts Institute of Technology.

Belk, Z. (2016). *Attributes of Attribution*. PhD thesis, University College London.

Belyaev, O. & Haug, D. (2014). The genesis of *wh*-based correlatives: From indefiniteness to relativization. Paper presented at Sinn und Bedeutung 19, Göttingen.

Boef, E. (2012). *Doubling in Relative Clauses: Aspects of Morphosyntactic Microvariation in Dutch*. PhD thesis, Universiteit Utrecht.

Clackson, J. (2007). *Indo-European Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Comrie, B. (1998). Rethinking the typology of relative clauses. *Language Design*, 1, 59–86.

De Vries, M. (2002). *The Syntax of Relativization*. PhD thesis, Universiteit van Amsterdam.

Downing, B. (1978). Some universals of relative clause structure. In J. Greenberg (Ed.), *Universals of Human Language, vol. 4: Syntax* (pp. 375–418). Stanford, CA: Stanford University Press.

Dryer, M. (2013). Position of interrogative phrases in content questions. In M. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/93.

Elbourne, P. (2001). E-type anaphora as NP-deletion. *Natural Language Semantics*, 9, 241–288.

Gärtner, H.-M. (2009). More on the indefinite–interrogative ambiguity: The view from embedded non-finite interrogatives. *Linguistic Typology*, 13, 1–37.

Haspelmath, M. (1997). *Indefinite Pronouns*. Oxford: Oxford University Press.

Haudry, J. (1973). Parataxe, hypotaxe et corrélation dans la phrase latine. *Bulletin de la Société Linguistique de Paris*, 68, 147–186.

Kiparsky, P. (1995). Indo-European origins of Germanic syntax. In A. Battye & I. Roberts (Eds.), *Clause Structure and Language Change* (pp. 140–169). New York: Oxford University Press.

References II

Luján, E. (2009). On the grammaticalization of *$k^w i$-/$k^w o$- relative clauses in Proto-Indo-European. In V. Bubenik, J. Hewson, & S. Rose (Eds.), *Grammatical Change in Indo-European Languages: Papers Presented at the Workshop on Indo-European Linguistics at the XVIIIth International Conference on Historical Linguistics, Montreal, 2007* (pp. 221–234). Amsterdam: John Benjamins.

Pinkster, H. (2015). *The Oxford Latin Syntax. Volume 1: The Simple Clause*. Oxford: Oxford University Press.

Probert, P. (2014). Relative clauses, Hittite, and Standard Average European. In S. Jamison, H. C. Melchert, & B. Vine (Eds.), *Proceedings of the 25th Annual UCLA Indo-European Conference* (pp. 137–164). Bremen: Hempen.

Sakalauskaite, J. (2016). Exogenous and endogenous influences on *the which*. Master's thesis, University of Edinburgh.