JHS Journal of Historical Syntax
Volume 9, Article 10: 1–39, 2025 · DOI: 10.18148/hs/2025.v9i2-10.185

SOURCES FOR QUESTION PARTICLES*

George Walkden
University of Konstanz

Tian Li
Hubei University of Automotive Technology

Laura Weisshaar
University of Konstanz

Abstract This paper deals with the diachrony of polar question particles (PQPs). We propose five source representations that can give rise to such particles: i) disjunction, ii) subordinators, iii) *wh*-words, iv) epistemic modality, and v) focus particles. Data is drawn from a typological survey, supported by in-depth case studies from the literature. The historical developments proposed are compatible with established formal approaches to grammaticalization.

1 Introduction

Among the languages of the world, the most common strategy for marking polar questions is a particle, as shown in Figure 1 (from Dryer 2013). In Dryer's sample, 585 of 955 languages mark polar questions using a question particle, and a further fifteen languages use both a question particle and interrogative verb morphology. Thus, 62.8% of the languages in the sample – nearly two thirds – deploy polar question particles (henceforth PQPs). This

* Versions of this material were presented at the Konstanz linguistics research colloquium in June 2021, at DiGS 23 in New York (remotely) in June 2022, at SyntaxLab in Cambridge in May 2023, at the Angus McIntosh Centre in Edinburgh in June 2023, and at the Questions at the Interfaces workshop in Konstanz in June 2023. Thanks to audiences at these venues, and in particular to Theresa Biberauer, Lisa Gotthard, Caroline Heycock, Georg Kaiser, Richie Kayne, Natasha Korotkova, Chris Reintges, Maribel Romero, Craig Sailor, Tamisha Tan, Elly van Gelderen, Ramona Wallner and Becky Woods, as well as to two anonymous reviewers for JHS. This research was funded by the Deutsche Forschungsgemeinschaft (DFG – German Research Foundation) as part of FOR 2111 'Questions at the Interfaces'.

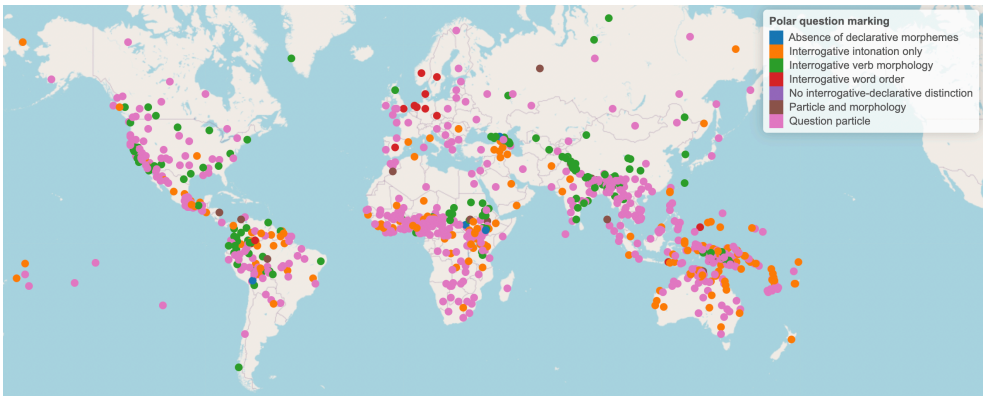


Figure 1 WALS feature 116A: marking of polar questions (Dryer 2013). Map generated using the R package lingtypology (Moroz 2017); base map is from OpenStreetMap.

is in stark contrast to the well-studied interrogative word order familiar from western European languages such as English, Spanish and Czech, which is found in only thirteen languages of the sample, nine of which are European (and seven of which are Germanic).

Given how prevalent question particles are, strikingly little attention has been paid to their diachronic origins.[1] The aim of this paper is to remedy this situation. Specifically, we aim to bridge the gap between typological surveys and diachronic treatments of individual particles in individual languages, by identifying common pathways of grammaticalization for PQPs. We identify five potential diachronic origins: i) disjunction, ii) subordinators, iii) *wh*-words, iv) epistemic modality, and v) focus particles. We also argue that these historical developments are compatible with formal approaches to grammaticalization such as those of Roberts & Roussou (1999, 2003) and van Gelderen (2004, 2011, 2021).

In section 2 we define our terms and lay out the background to our research. Section 3 details the empirical basis of our study, and in section 4 we present, exemplify and motivate the five sources for question particles. Section 5 contains discussion of our approach and how it compares to existing proposals, and section 6 concludes.

1 Synchronic typological surveys, by contrast, are not in short supply: these include Moravcsik (1971), Ultan (1978), Sadock & Zwicky (1985), König & Siemund (2007), and of course Dryer (2013).

2 Definitions and background

Adopting the assumptions of mainstream Minimalist syntax (following e.g. Chomsky 2000, 2001) and of Distributed Morphology (Halle & Marantz 1993) –in particular Late Insertion, the idea that morphological material is realizational and inserted postsyntactically– we can define PQPs as in (1).

(1) *Polar question particle* (*PQP*): A Vocabulary Item, associated with overt phonological material, that can serve as spellout of a syntactic head associated with polar question semantics in a matrix clause.

This definition leaves open the possibility that the same Vocabulary Item can be spelled out in other contexts too – for instance, in *wh*-questions – or that it might be restricted to a particular subtype of polar questions, e.g. biased (Romero & Han 2004, Sudo 2013), rhetorical (Han 2002, Rohde 2006), or otherwise non-canonical (Trotzke & Czypionka 2022). It thus contrasts with other definitions of PQPs, such as that of Bhatt & Dayal (2020: 1142), who classify as such only those elements that occur *solely* in polar questions. For a recent, more fine-grained investigation of types of particles in polar questions, see Gonzalez (2023).

By 'polar question semantics', we mean the standard semantics for polar questions adopted in the tradition following Hamblin (1958, 1973): a function that takes a proposition p and returns a set of propositions consisting of p and its negation, i.e. $\{p, \neg p\}$.

As for the underlying syntax of question particles, we assume it involves a syntactic head that is in or close to the periphery of the clause. Monforte (2018b) outlines two stances on this in the literature:

i Question particles are in the CP-domain (e.g. Rizzi 2001, Prieto & Rigau 2007, Kuwabara 2013, Paul & Pan 2017, Bhatt & Dayal 2020)

ii Question particles are in the TP-domain or lower (e.g. Hagstrom 1998, Duffield 2007, Bayer & Obenauer 2011, Scherf 2017)

Monforte proposes that both may be correct for different languages. Similarly, the proposal for polar questions in Holmberg (2016) involves two separate positions – a PolP below TP, and a Q-Force position high in the CP-domain – with polarity elements moving (overtly or covertly) from PolP to the Q-Force position for scopal reasons (see also Chen 2022).[2] Under Holmberg's approach, the Hamblin semantics of questions is encoded in Pol, in this

2 The idea that question particles are first Merged in a clause-internal position and undergo movement to a higher position goes back to Yanagida (1995) and Hagstrom (1998, 2000).

case a free variable with two possible values, positive and negative: a proposition with such a free variable is equivalent to a set consisting of a proposition and its negation (Holmberg 2016: 15–17). The key syntactic ingredient of PQPs is thus Pol, which may subsequently move into the C-domain (in matrix clauses) to Q-Force, signalling that the utterance is a request for information from the addressee. We adopt this stance here, noting that the typically peripheral position of PQPs in the clause cross-linguistically is an argument for a spellout position relatively high in the clause structure.[3]

2.1 *Formal approaches to grammaticalization*

A further aim of this paper is to assess the extent to which the sources of question particles are compatible with (and can be understood in terms of) formal approaches to grammaticalization. In this section the key approaches are introduced that will be of relevance to the discussion in section 5.

Roberts & Roussou (2003) propose that grammaticalization is the creation of new functional material involving '[s]uccessive upward reanalysis along the functional hierarchy' (2003: 202), and that it leads to 'structural simplification' (2003: 2). The driver of their approach is a markedness hierarchy as regards the realization of formal features, given in (2).

(2) **Markedness hierarchy**: $F^*_{Move/Merge} > F^*_{Move} > F^*_{Merge} > F$

Features may be strong (indicated by the * diacritic) or weak; strong features must be realized at PF. The least marked option is a weak feature. If the feature is strong, then a realization by means of movement (i.e. Internal Merge) is more marked than a realization by means of (External) Merge. The most marked option of all is one in which feature realization involves both External and Internal Merge.

All of the empirical case studies Roberts & Roussou investigate can be reduced to one of three basic patterns, given in (3).

(3) (a) $[_{XP}\ Y{+}X\ [_{YP}\ \ldots\ t_Y\ \ldots]] > [_{XP}\ Y{=}X\ [_{YP}\ \ldots\ Y\ \ldots]]$
(b) $[_{XP}\ X_F\ \ldots\ [_{YP}\ \ldots Y_F\ \ldots]] > [_{XP}\ X_F\ \ldots\ [_{YP}\ \ldots\ Y\ \ldots]]$
(c) $[_{XP}\ YP\ X\ \ldots\ [\ldots\ t_{YP}\ \ldots]] > [_{XP}\ Y{=}X\ \ldots\ [\ \ldots\ ]]$

3 In principle, of course, the syntax of question particles could be more interesting and complex than just a single head in the clausal spine. (We are grateful to Richard Kayne for raising this point in the question period of DiGS 2022.) Bailey (2012), for example, proposes that Chinese question particles are not heads of a right-headed CP, but rather disjunctors heading a null or elliptical disjunct; Pan & Paul (2016) and Paul & Pan (2017), in response, defend the traditional view. We will assume for simplicity's sake that all the question particles we discuss are heads in the CP-domain or TP-domain, while noting that more fine-grained investigation may show this assumption to be wrong in individual cases.

The case in (3a) involves reanalysis of a moved head as first Merged in the higher position, e.g. the grammaticalization of modals in the history of English. The case in (3b) involves the loss of agreement and innovation of a new first-Merged exponent, as for instance in the development of new modal particles such as *na* and *mu* in the history of Greek. The case in (3c) involves the reanalysis of a (moved) phrase in specifier position as instead being the head of that position, and is in fact a two-step process: i) the loss of movement and ii) spec-to-head reanalysis.

van Gelderen (2004) presents a formal account of grammaticalization that is similar in its essentials. The driving forces boil down to the principles in (4).

(4) (a) **Head Preference or Spec to Head Principle**:
Be a head, rather than a phrase. (van Gelderen 2004: 11)

(b) **Late Merge Principle**:
Merge as late as possible. (van Gelderen 2004: 12)

The Head Preference Principle in (4a) captures spec-to-head reanalysis as illustrated in (3c). The Late Merge Principle in (4b) drives reanalysis of internally Merged material as externally Merged in a higher position, as illustrated in (3a) and (3b).[4]

If these authors are correct, then what we should expect to see in the grammaticalization of question particles – as in other instances of grammaticalization – is i) upward reanalysis or ii) spec-to-head reanalysis. In section 5 we will return to see whether this is the case.

2.2 *Previous work on the diachrony of question particles*

Until recently, despite a number of works dealing with the diachrony of individual question particles or languages, there had been no systematic attempt to identify a set of sources for question particles comparable to e.g. Bybee, Perkins & Pagliuca (1994) for future markers or Diessel (1999: ch. 6) for demonstratives. The first edition of Heine & Kuteva's *World Lexicon of Grammaticalization* (2002) mentions only two pathways relating to the emergence of question particles, given in (5); S-QUESTION is their term for a polar question marker.

4 In later work, van Gelderen (e.g. 2009, 2011) formulates an explanation for these generalizations based on Feature Economy, i.e. reanalysis from semantic to interpretable to uninterpretable features. Her most recent work (van Gelderen 2021) explores the idea that third-factor principles in the sense of Chomsky (2005) may be the driving force. For the purposes of this paper, the precise causal factors behind the change are not crucial.

(5) (a) NEGATION > S-QUESTION (Heine & Kuteva 2002: 216–217)

(b) OR > S-QUESTION (Heine & Kuteva 2002: 226–227)

As regards (5a), Heine & Kuteva (2002) mention final question particles in Sino-Tibetan languages (Harris & Campbell 1995: 294–5), Turkish *-mi*, and more tentatively English tag questions. The correct analysis of such cases is discussed in section 4.1 of the present paper. As regards (5b), they adduce Moré, Hausa, Kxoe, Latvian and Basque, among others. In both cases the wording is cautious, e.g. for (5b) 'Further investigation is required to study the exact nature of this process and the interaction of conjunctions and negation markers' (Heine & Kuteva 2002: 227).

A short paper by Bencini (2003) also broaches the issue of the diachrony of question particles. Based on 22 particles from 23 languages, she identifies two main sources – negation and disjunction, as in (5) above – and three source constructions, given in (6) (from Bencini 2003: 613).

(6) (a) Source: S (or) *not* S. Reanalysis: S *not* > S QP

(b) Source: *or* S. Reanalysis: *or* S > QP S

(c) Source: S *or*. Reanalysis: S *or* > S QP

While (6a) and (6c) yield final question particles, (6b) yields initial question particles. Bencini does not claim that this is an exhaustive list of sources, and suggests that markers of doubt, conditionals and hypotheticals may be a further source (Bencini 2003: 609).

van Gelderen (2009: section 3.1) proposes a diachronic pathway not found in earlier literature: interrogative pronoun to question particle, drawing on examples from English, Hindi/Urdu, and Lombard (see also Hackstein 2004). Section 3.2 of van Gelderen (2009) also discusses negative verbs in Chinese varieties as a source of question particles.

The most detailed studies on sources for question particles to date are Metslang, Habicht & Pajusalu (2017) and Aigro (2020). These authors' starting point is Estonian, a language that is very rich in question particles, including in its recent history. These authors are also the only ones to date whose aim is more ambitious than simply stating a handful of possible pathways. The main sources proposed by Metslang et al. (2017: 494) are given in (7).

(7) (a) Means of linking sentences in discourse:

a. markers of coordination (disjunctive 'or', conjunctive 'and', 'also', adversative 'but'), typically on the left periphery,

b. markers of subordination (insubordination of complementizers 'if', 'whether', e.g. German *ob*), typically on the left periphery,

(b) Markers of alternation ('or not', 'or', 'not'), typically on the right periphery,

(c) Markers of epistemic modality ('perhaps', 'maybe'), which can appear in various places within a sentence.

The empirical basis of Metslang et al.'s study is the languages of the Circum-Baltic area, with comparisons drawn to various other particles mentioned in the literature; they also mention (in their section 6) interrogative pronouns. The study is qualitative and does not claim to be exhaustive as regards potential pathways (see also Metslang, Habicht & Pajusalu 2011). Later in the paper they provide a slightly different list of sources, given in (8), from Metslang et al. (2017: 515).

(8) (a) 'coordination markers: conjunctive, disjunctive and adversative coordination;

(b) subordination markers, the source of which is the subordinating conjunction in the case of insubordination of the subordinate clause;

(c) epistemic markers, which arise from particles expressing epistemic modality;

(d) pronominal markers, which arise from interrogative pronouns and proadverbs in tag questions.'

Metslang et al.'s findings – specifically the list of sources in (7), rather than that in (8) – have been incorporated into the second edition of the *World Lexicon of Grammaticalization* (Kuteva, Heine, Hong, Long, Narrog & Rhee 2019). They propose the pathways in (9).

(9) (a) VP-AND > S-QUESTION (Kuteva et al. 2019: 59–60)

(b) NEGATION > S-QUESTION (Kuteva et al. 2019: 293–294)

(c) OR > S-QUESTION (Kuteva et al. 2019: 306–308)

(d) PERHAPS > S-QUESTION (Kuteva et al. 2019: 321)

VP-AND stands for a coordinating conjunction in Kuteva et al.'s terminology (compare (7a-a)). Unlike Metslang et al. (2017), but following Heine & Kuteva

(2002) and Bencini (2003), they consider NEGATION > S-QUESTION to be a pathway, though in the second edition they clarify that they consider the context for this pathway to be markers of alternation (compare (7b)).

In contrast to previous studies which have approached the origins of particles through a mainly syntactic lens, Aigro (2020) is a study that proposes a uniform SEMANTIC characterization for the sources of question particles. Her argument is that all of the functional categories that plausibly serve as sources are markers of POLARITY. These sources are given in (10), from Aigro (2020: 39–40). She also provides numbers for some of these sources, based on her own reading of the literature.

(10)
(a)	Disjunctive coordination	(21 languages)
(b)	Negation	(15 languages)
(c)	Conjunctive coordination	(Lithuanian, Bulgarian, Estonian)
(d)	Embedded PQPs	(Swedish, Estonian)
(e)	Conditional markers	(Hua, Russian, ...)
(f)	Interrogative pronouns	(Polish, Bengali, Kannada, Tocharian)

Aigro problematizes four other categories mentioned as sources in the literature: inferential and adversative coordination, subordination markers, and epistemic modality markers. This is because it is not clear whether these are actually PQPs rather than simply discourse material that tends to occur in polar questions. A different concern can be raised about source (10e), conditional markers. Aigro rightly observes that conditional markers are often homophonous with PQPs (e.g. König & Siemund 2007).[5] However, while the pathway from PQP to conditional marker appears to be well-trodden (Heine & Kuteva 2002: 249; Hopper & Traugott 2003: 186 and literature cited there; Kuteva et al. 2019: 353–354; Iatridou & Zeijlstra to appear), we are not aware of any evidence that the reverse pathway is attested at all. Thus, this source will not play any role in what follows.

This concludes our overview of the literature. Specific studies on the genesis of individual particles are discussed in section 4, where we present our own proposals for sources and pathways. Further critical discussion and comparison of our proposal with previous ideas is found in section 5.

5 Indeed, for English it has even been proposed that this state of affairs is synchronically non-accidental (Starr 2014). See Iatridou & Zeijlstra (to appear) for arguments that the link is a purely diachronic one.

3 The survey

In this section we outline the empirical material that we have taken into account in our proposal (section 4). We consider it important to do so explicitly in order to avoid the misleading impression that prose around 'commonness' or 'rarity' of pathways can create, and also to create a basis for future work on question particles, since we are under no illusion that this paper will be the last word on the issue.

The full dataset can be found in the Appendix. At present, we have taken into account 83 question particles in 67 languages. A summary of the dataset by genealogical grouping can be found in Table 1, and a summary by linguistic area can be found in Table 2.[6] The dataset contains a wide variety of families and areas, though at the same time some major families and areas are absent: most notably, Australia is not represented in the dataset at all, and New Guinea, Oceania and the Americas are barely present. At the same time, Inner Asia and particularly Europe are overrepresented.

Importantly, the dataset should not be thought of, or used, as a representative sample of the world's languages, or of possible human languages. The notion of a sample of languages that is representative of the population in a statistical sense ('languages as independent trials') is in any case problematic because of founder effects: see Maslova (2000: 311–312). For this reason, the present paper – like most work in diachronic typology (Hendery 2012: 9–11) – eschews inferential statistics, aiming only to establish which pathways are attested. Numerical values are included because we consider it important to be maximally explicit about the nature and quantity of our evidence.

For each particle in the dataset, the following pieces of information are provided in the Appendix: i) the language in which the particle is found, ii) the family that language belongs to, iii) the form of the particle, iv) the position of the particle in the clause, v) the pathway we have analysed it as emerging from, and vi) a source. Only one source is provided for each particle; where multiple sources exist, we have provided the most in-depth discussion we have been able to find, giving preference to more recent sources (and those that themselves make reference to relevant sources) in case of doubt. The pathway of origin is our own interpretation, and in some cases differs from, or goes beyond, what is said in the source provided.

Two strategies were adopted to collect the data for this paper. First, case studies discussed in previous literature on the diachrony of question particles were collected and incorporated, particularly from Bencini (2003),

6 Information about area is taken from AUTOTYP (Nichols, Witzlack-Makarevich & Bickel 2013), with the exception of Turku (not in AUTOTYP), which is spoken in Chad and was hence assigned to the African Savannah area.

Family		Languages	Particles
Afro-Asiatic		7	9
Atlantic-Congo		2	2
Austro-Asiatic		1	1
Austronesian		3	5
Dravidian		3	3
Indo-European	(Celtic)	3	3
	(Germanic)	3	4
	(Indo-Iranian)	3	3
	(Italic)	5	7
	(Slavic)	8	9
	(other)	5	6
Japanese		1	1
Kartvelian		1	1
Khoe		1	1
Mande		1	1
Mayan		1	1
Na-Dene		1	1
Quechuan		1	1
Sino-Tibetan		4	4
Tai-Kadai		3	5
Trans-New Guinea		1	1
Turkic		1	1
Uralic		4	8
Uto-Aztecan		1	1
	(*Isolates*)	3	4
Total		**67**	**83**

Table 1 Language families in the dataset

Continent	Area	Languages	Particles
Africa	African Savannah	6	6
	Greater Abyssinia	2	2
	N Africa	1	3
	S Africa	1	1
C America	Mesoamerica	2	2
E N America	E North America	1	1
W N America	Basin & Plains	1	1
S America	Andean	1	1
N-C Asia	Inner Asia	10	14
	N Coast Asia	1	1
S/SE Asia	Indic	6	7
	Southeast Asia	7	9
W and SW Eurasia	Europe	21	26
	Greater Mesopotamia	4	4
NG and Oceania	Interior New Guinea	1	1
	Oceania	2	4
Total		**67**	**83**

Table 2 Linguistic areas in the dataset

Metslang et al. (2017), and Aigro (2020). Secondly, we took a subset of the languages described in WALS (Dryer 2013) as having PQPs, and searched the literature for information on their historical origins. In principle, such information can come either from historical (textual) evidence or from synchronic considerations (distribution across functions), though the latter does not always present a clear picture.

We have tried to err on the side of inclusivity. Since we are drawing case studies from a variety of sources and grammatical traditions, there are potentially problems of granularity of analysis. Our aim is to follow the 'Middle Way' laid out by Baker & McCloskey (2007) for typology – looking at an intermediate number of languages, at an intermediate level of detail – but we may not always have succeeded in achieving the level of detail we aim at. Issues are expected to be of two kinds.

First, as noted, our definition of polar questions in (1) includes those that are restricted to a particular subtype of polar questions, e.g. biased or rhetorical questions. For example, Kieviet (2017: 481) states that the Rapa Nui clause-initial particle *hoki* 'is used especially when the speaker expects a certain answer to the question, whether affirmative ... or negative', suggesting a characterization in terms of bias (see also Du Feu 1996: 18).[7] Similarly, the Old English clause-initial particle *hwæþer*, discussed in detail by Eckardt & Walkden (2022), is described there as used in pedagogical questions: questions in which the questioner already knows the answer, but is attempting to elicit it from the addressee for teaching purposes. We have not attempted to capture this level of detail in our dataset, as there is simply not enough available information on the semantic and pragmatic nuances of PQPs cross-linguistically, and what we have is not easily comparable. Nevertheless, we hope to return to these issues in future work. In the meantime, particles that introduce biased or rhetorical (etc.) polar questions are included in our dataset.

Secondly, and more seriously, some of the particles we have included may not be true PQPs at all. We have included the Japanese clause-final particle *no* as a PQP, following Hayashi (2010). However, Hinds (1986: §1.1.1) states that there is disagreement as to whether *no* is a true question particle. Since *no* is also a nominalizer (not only used in questions), one possible analysis is that *no*-questions are derived via ellipsis of a polar question matrix clause, e.g. involving the copula *desu*. However, *no* and *no desu* are not always interchangeable (McGloin 1980: 146–7, note 9), casting doubt on a naïve ellipsis analysis. Another particle whose status is unclear is Urdu/Hindi *kyā*. Bhatt &

7 Kieviet (2017) also notes that the particle is common in older texts, but less common in today's Rapa Nui.

Dayal (2020) analyse this as a PQP; Butt, Jabeen & Bögel (2020), on the other hand, treat it as a focus-sensitive operator that associates with the focused material.[8] It is not always easy to distinguish PQPs from discourse particles that are licensed *in* polar questions (on which see e.g. Czypionka, Romero & Bayer 2021), and we may have made the wrong decision in some cases.

Putting these considerations aside, we now turn to the pathways we propose for the origins of PQPs.

4 Pathways

Languages – in the sense of entities persisting over time, beyond the individual's I-language – are convenient fictions, and so a note on our use of the term 'pathway' is in order. The key notion is what Walkden (2021) terms reanalysis$_2$: a change event in which a perceiver/acquirer assigns to an input sentence a parse that is different from the one intended by the person who produced that sentence. If multiple such change events occur, in different varieties, in a way that affects the same type of lexical item with the same effects (e.g. a negative adverb becomes reanalysed as a head that functions as a negator; see Willis, Lucas & Breitbarth 2013), we can speak of a pathway. Pathways are thus generalizations over change events. They are not necessarily unidirectional, except in the rather trivial definitional sense: in any given case, whether the reverse pathway (e.g. negative heads being reanalysed as negative adverbs) commonly occurs, or occurs at all, is an empirical question. Pathway in this sense – like reanalysis$_2$ – is also a descriptive notion, devoid of any explanatory or causal force.

Pathways may also be identified in terms of sequences of change events, e.g. a demonstrative reanalysed as a definite article that is then in turn reanalysed as a marker of non-genericity (Greenberg 1978, van Gelderen 2007, 2021). In such cases we can talk about the PROXIMATE source of a lexical item – the source that is separated from it temporally only by the most recent change event – as opposed to its ULTIMATE source at the start of the pathway. A non-generic marker that has traversed the aforementioned pathway has two sources: its proximate source is a definite article, but its ultimate source (at least as regards this particular pathway) is a demonstrative. The distinction will be important in what follows.

With these preliminaries in mind, we now turn to the five pathways.

8 Although *kyā* is prototypically clause-initial in polar questions and is annotated as such in our dataset, Bhatt & Dayal (2020) and Butt et al. (2020) show that it may also occur in other positions.

4.1 *Disjunction*

As seen in section 2, the idea that PQPs may have their origin in disjunction is well established in the literature (Heine & Kuteva 2002, Bencini 2003, Bailey 2012, Jayaseelan 2012, Metslang et al. 2017, Kuteva et al. 2019). In our dataset, 44 of our 83 PQPs – just over half – plausibly originate via this pathway.

The core of the pathway is that a structure with two full clauses linked by a disjunctor – *X or Y*, or *X or not* – is reanalysed as a single clause with a PQP.[9] Typically, one of the two clauses in the proximate source structure will be heavily elided or reduced so that it consists of just one or two words. We illustrate with the case of Mandarin Chinese *ma*, following the account in Aldridge (2011).

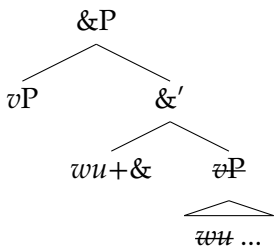
Chinese is known for its 'A-not-A' alternative questions in which two clauses of opposite polarity are juxtaposed without an overt disjunctor. Earlier stages of Chinese allowed this too, as in (11).

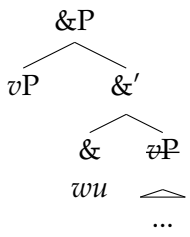
(11) 秋寒有酒無? (Bai Juyi, ninth century)
Qiu *han* *you* *jiu* ***wu****?*
autumn cold have liquor not.have
'In the autumn cold, is there any liquor?'

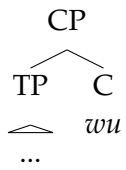
Aldridge proposes that the negative existential *wu* – pronounced with an initial /m/ – grammaticalized into present-day *ma* in three stages during the Middle Chinese of the fifth to tenth centuries. In the first, *wu* grammaticalized from a lexical verb to a verbal functional head. A-not-A questions as in (11) are derived, according to Aldridge, by coordination of *v*Ps under &P, as in (12). As part of this process, *wu* could move to the head of &P. As a second step, *wu* was reanalysed as first Merged in &, becoming a disjunctor, potentially with an elided complement, as in (13). The third and final step involves the reanalysis of this low disjunctor as a C head, much higher in the structure, as in (14).[10]

9 The source structure may itself in principle be a declarative, or – perhaps more likely – an alternative question.

10 For Aldridge (2011) this last step is in fact two steps: one in which *wu* becomes a disjunctor in C preceding an elided TP complement, and one in which the CP is reanalysed as head-final. For our purposes nothing rests on this.

(12)


(13)


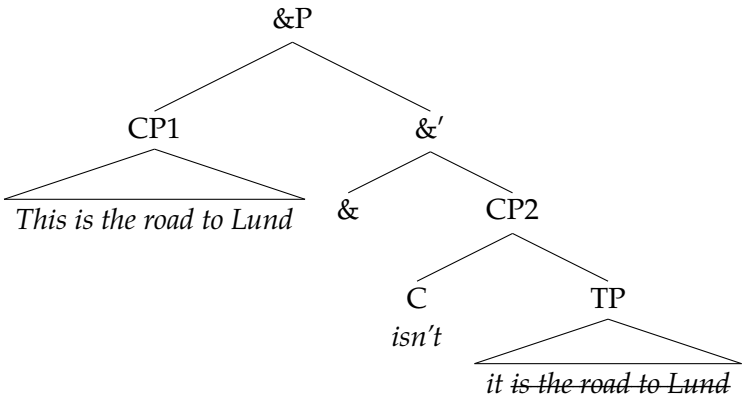
(14)


Other particles that plausibly arose via this route include Malayalam *o* (Jayaseelan 2012), Niuean *nakai* (Starks & Massam 2015), and Estonian *või* (Aigro 2020).[11] A variant of this pathway in which it is the first disjunct that is elided rather than the second can give rise to clause-initial particles, e.g. Cornish *na* (Buchholz & Fiedler 1987) and Mandinka *fó* (Creissels 2020).

Some varieties of English use invariant *innit* as a particle, as in (15), from Sailor (2009).

(15) Tom's the one who likes that Swedish death-metal shite, **innit**?

This *innit* clearly derives from a tag question diachronically. Sailor (2009) argues that tag questions are reduced interrogative clauses involving VP ellipsis. To us this suggests that the tag-question origin of particles like *innit* can be assimilated to the disjunction pathway. Building on McCawley (1988) and Holmberg (2016: §4.8.2), we propose that tag questions can be analysed as two CPs connected by a null disjunctor, as in (16).

(16)


From here, the tag has the potential to be reanalysed as an invariant disjunctor, as in the second step of the pathway above, (13), and then as a clause-final C head. Thus the tag question source follows the disjunction pathway. The

11 Though Aigro proposes that the mediating structure is disjunction of NPs, not of *v*Ps as in (13).

final particle *aśśi* in Tocharian A may well also have its origin in a tag question (Hackstein 2004: 174–5).

Another source which we believe can be assimilated to the disjunction pathway is negation. Heine & Kuteva (2002), Bencini (2003), Metslang et al. (2017), and Kuteva et al. (2019) all mention negation as a source of PQPs. In early work such as Heine & Kuteva (2002) and Bencini (2003) it is the only source mentioned other than disjunction. We agree with these authors insofar as negators or negative words clearly can and do develop into PQPs given enough time; the case study of Mandarin *ma* as laid out in Aldridge (2011) is one such case (see also van Gelderen 2009: §3.2). However, though negators may serve as an ULTIMATE source for PQPs, it is not clear that they can also be a PROXIMATE source.

We propose instead that negators develop into PQPs via the disjunction pathway, and that the proximate source is a disjunctor which itself has its origin as a negator, as in the history of Chinese. Nguyễn (1997: 237–8) suggests that the same is true of Vietnamese *không*, another of the examples in Bencini's (2003) survey. Likewise, Gyuris (2017: 49) cites Simoncsics (2003) as arguing that Hungarian final *e* is 'without doubt the remainder of the Uralic negating verb stem **e-*, which has been assumed to have died out'; she also adduces parallels with Chinese A-not-A constructions. If so, the diachrony of Hungarian *e* is precisely parallel to that of Mandarin *ma*.

We cannot, of course, prove that there are no languages in which negation is a proximate source for PQPs. We leave for future research the question of whether negators can ever develop directly into PQPs without an intermediate disjunction step of reanalysis.

4.2 *Subordinators*

Fifteen of the question particles in our dataset have their origin in subordinators: these include French *est-ce que*, Japanese *no*, Rapa Nui *hoki*, and Tzotzil *mi*.

This pathway relates to the phenomenon known as insubordination (e.g. Evans 2007, Traugott 2017, Corr 2018). We propose that it involves the reanalysis of a biclausal struture as monoclausal, where the original matrix clause is either elided or heavily reduced.

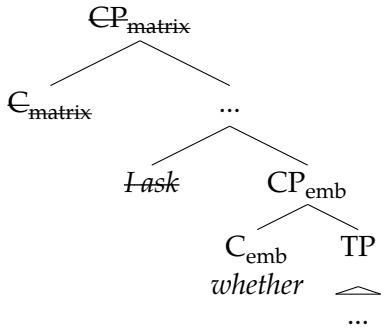
Ellipsis of everything but the embedded clause is likely the origin of particles such as Rapa Nui *hoki*, as in (17), and Old English *hwæþer*, as in (18).

(17) ***¿Hoki*** *ko* *tike'a* *'ā* *e* *koe* *te* *tātane* *ra'e?*
PQ PRF see CONT AG 2SG ART devil first
'Have you ever seen a devil?' (Kieviet 2017: 481)

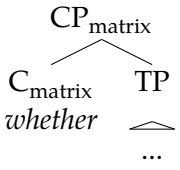
(18) ***Hwæðer*** *nu* *gimma* *wlite* *eowre* *eagan* *to* *him* *getio*
whether now jewels looks your eyes to them attract
'Does the beauty of jewels attract your eyes ... ?'
(Eckardt & Walkden 2022: 41)

Both particles are also used as complementizers to introduce embedded polar questions (Walkden 2014: 148–9 on Old English; Kieviet 2017: 190 on Rapa Nui). In the case of Old English *hwæþer*, it is demonstrably the complementizer use that is older (Walkden 2014: 154–5). We tentatively assume that, where the same particle is used to introduce both embedded and unembedded polar questions, the latter function is historically derived from the former via reanalysis, as schematized in (19), before, vs. (20), after.

(19)

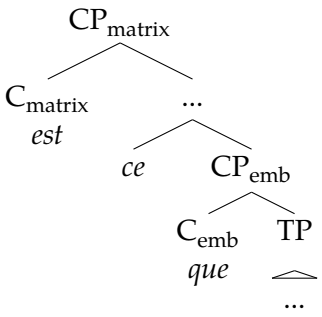


(20)

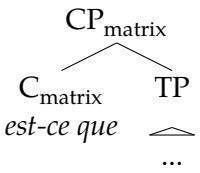


By contrast, reduced (but not elliptical) matrix clauses are the direct source of particles such as French *est-ce que* (originally 'is it (the case) that'; Druetta 2003, Elsig 2009, Tailleur 2013: ch. 3), Portuguese *é que*, and others.[12] Here we also see a reanalysis from biclausal to monoclausal, schematized in (21) vs. (22) (cf. Tailleur 2013: 62 on *est-ce que* in *wh*-questions).

(21)



(22)



12 The status of Portuguese *é que* as a PQP is dubious: most speakers of both European and Brazilian Portuguese we have consulted do not accept it. WALS (Dryer 2013) cites Parkinson (1988: 158) as a source for the claim, but in the second edition of this chapter (Parkinson 2009: 228) the claim is removed. It may be that *é que* is a question particle for some speakers/varieties but not for others.

A particularly interesting case is found in Egyptian Arabic, where there are three question particles apparently derived ultimately from personal pronouns: *huwwa* (masculine), *hijja* (feminine), *humma* (neuter) (Gary & Gamal-Eldin 1982: 4, Soltan 2011), as in (23).

(23) ***huwwa*** *ʔinti* *gaaja* *ʔinnaharda*
Q.3SG you.F.SG coming today
'Are you coming today?' (Gary & Gamal-Eldin 1982: 4)

These are the only examples we have found of gendered personal pronouns becoming question particles, and here again the distinction between ultimate source (here, pronouns) and proximate source becomes relevant. A crucial fact about these pronouns is that they can also serve as copulas in Arabic varieties (Eid 1991, 1992, Tan 2025: §2.1). The reanalysis of pronouns as copulas is a cross-linguistically very well attested process (Lohndal 2009, van Gelderen 2015, Tan 2025). Since Arabic is a null subject language, the structure of (23) before reanalysis is likely to have originally been biclausal ('is it (the case) that'), just as in (21) – and indeed this is the synchronic analysis proposed by Eid (1992). However, Soltan (2011) shows that Eid's analysis runs into problems, and argues that *huwwa* and its ilk are instead PQPs in the C-domain – which arose diachronically from reanalysis of a biclausal structure as monoclausal, just as in the other examples in this subsection, with the copula as proximate source.

4.3 Wh-*words*

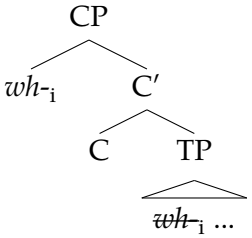
There are fourteen PQPs in our dataset that most plausibly derive from *wh*-words, including Belorussian *ci* (Mayo 1993: 926), Polish *czy* (Metslang et al. 2017: 513), Indonesian *apa* (Danusugondo 1975: 1–2, Sneddon 1996: 311, 320), and Urdu/Hindi *kyā* (Bhatt & Dayal 2020, Butt et al. 2020). As discussed in section 2, the pathway from interrogative pronoun to PQP is presented by van Gelderen (2009), and mentioned briefly in Metslang et al. (2017: 512–4) and Aigro (2020: 240). Some examples are given in (24)–(26).

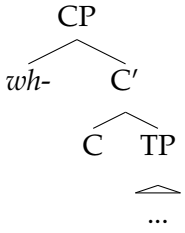
(24) ***Czy*** *to* *jest* *pan* *Krakowski?*
Q that is pan Krakowski
'Is that Mr. Krakowski?' (Polish; Hackstein 2004: 175)

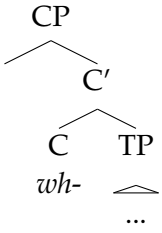
(25) ***Apa*** *ini* *hari* *Selasa?*
Q this day Tuesday
'Is this Tuesday?' (Indonesian; Sneddon 1996: 311)

(26) ***kya*** ***anu=ne*** *uma=ko* *kɪtab* *d-i?*
what Anu.F=ERG Uma.F=DAT book.F.SG.NOM give-PERF.F.SG
'Did Anu give a/the book to Uma?'
(Urdu/Hindi; Butt et al. 2020: 86)

These originate via progressive reanalysis and semantic bleaching of *wh*-words in *wh*-questions, as schematized in (27) vs. (28) and (29). As laid out by van Gelderen (2009) and Walkden (2013), there are likely two stages to the process. In the first, a *wh*-word originally moved to Spec,CP is reanalysed as first Merged there. In the second, the element in Spec,CP is reanalysed as a C head.

(27)


(28)


(29)


Evidence for the existence of the intermediate stage in (28) is provided by varieties in which a *wh*-word that originally had argumental function takes on a high adjunct role, as in (30)–(31), from Munaro & Obenauer (1999).

(30) ***Was*** *rennst* *du* *so* *schnell?*
what run you so fast
'Why are you running so fast?'
(German; Munaro & Obenauer 1999: 184)

(31) ***Cossa*** *zighè-tu?!*
what shout-2SG
'Why are you shouting?'
(Pagotto; Munaro & Obenauer 1999: 192)

In these examples the word originally meaning 'what' plays a 'why'-like role, indicating that it is first Merged very high in the clause (Rizzi 1990, 2001). Here the *wh*-word is still clearly in specifier position, since the C head is occupied by the finite verb; this structural signposting presumably prevents the acquirer from reanalysing the *wh*-word from specifier to head. In languages where C is not regularly filled overtly in this way, such as Polish and Urdu/Hindi, there is nothing to prevent spec-to-head reanalysis.

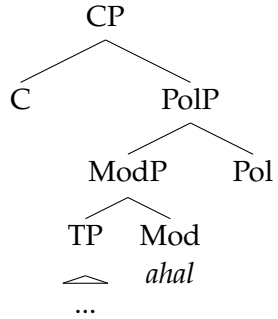
4.4 *Epistemic modality*

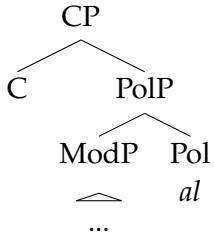
The remaining sources we have identified are rarer in our dataset. Four particles most plausibly have their origin in adverbial elements with epistemic modal function: Basque *al*/*a* (Monforte 2018b), Greek *mípos* (Mackridge 1985: 301, Roussou 2015: 148), and Niuean *ka* and *kia* (Starks & Massam 2015). We illustrate following Monforte (2018a,b). An example of *a* is given in (32 a), and an example of *al* in (32 b).

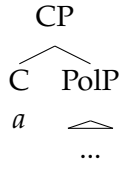
(32) (a) *Nehor* *ikusi* *duzu-**a**?*
anybody see AUX-A

(b) *Inor* *ikusi* ***al*** *dezu?*
anybody see AL AUX
'Did you see anybody?'

Monforte (2018a) argues that the question particle *al* derive historically from the epistemic particle *ahal*, which he characterizes as conveying the degree of probability that the speaker gives to the proposition.[13] In broad strokes, we can characterize the grammaticalization process as involving a reanalysis of a modal head high in the T-domain (Cinque 1999) as a higher polarity head in the T-domain, then finally as a question particle in the C-domain, as illustrated in (33) vs. (34) and (35).

(33)


(34)


(35)


Monforte (2018b) argues that *a* occupies a higher position than *al*.[14] Since *a* is also the more phonologically reduced particle, we can posit that *a* is a further grammaticalized version of *al*, in C. Intermediate stages may have involved head-movement, for instance of the particle in Pol to C before its reanalysis

13 There are dialectal differences within Basque as regards the use of *a* and *al*, as well as distributional syntactic differences: for present purposes these details are not crucial. See Monforte (2018a,b) for more discussion and data.

14 Monforte (2018b) analyses this as Bayer & Obenauer's (2011) Particle Phrase (PartP). We have assumed that this is more or less equivalent to Holmberg's (2016)'s PolP.

as a particle first merged in C (as shown in (35)). An earlier stage may have involved reanalysis of a modal adverb in specifier position as the modal particle *ahal* in head position as in (33). See Coniglio (2022) for the argument that modal particles consistently originate as adverbs.

Beyond Basque, another very similar case is Niuean *kia*. Starks & Massam (2015: 199–202) argue that this originates in the Proto-Polynesian subjunctive/irrealis particle *kia. Similarly, Niuean *ka* appears to originate as a confirmation-seeking particle *kaha* (Starks & Massam 2015: 206–9).[15] Since we have only few examples of this trajectory, however, more research is needed to establish how widespread the pathway from epistemic modal to question particle is.

4.5 Focus particles

The final source for PQPs is as a focus particle. Five particles in our dataset plausibly originate via this pathway: Estonian *ju*/*jo* (Aigro 2020: 248–50) and *kas* (Aigro 2020: 242–4), present-day Lithuanian *ar* (Aigro 2020: 240), Old Lithuanian *biau*, *be* (Ostrowski 2012), and Slave *nį*/*élį* (Rice 1989: 1124, 1130). We illustrate using Estonian *jo*/*ju*, following Aigro (2020). Aigro shows that this particle served as a focus marker in the early 1600s – as in (36 a) – and became reanalysed as an (affirmative-biased) PQP in the eighteenth century, as in (36 b), though this latter use does not survive into the present-day language.

(36) (a) *Toddest, se on **io** vx röhmu weerdt.*
Really, DEM COP.3SG FOC.AFF one joy.PRT worth
'Really, this is indeed worth joy.' (COWE, 1601; Aigro 2020: 248)

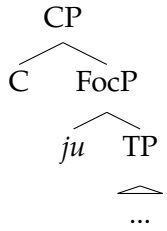
(b) ***Jo** neil siis ka murret on wihmasest aastast?*
PQP.AFF 2PL.AD then also worry.PRT COP.3SG rainy.ELA year.ELA
'They have troubles because of the rainy year, right?'
(COWE, 1732; Aigro 2020: 249)

Focus particles like *only*, *also* and *even* tend to be syntactically flexible, occurring in a position where they can take scope over the focused constituent. With that in mind, reanalysis from a wide-scope focus particle, as in (37), to

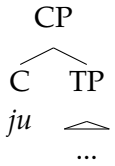
15 Though the final position of the latter particle might suggest a proximate source as a tag question instead, as in the pathway in section 4.1.

a polar question particle in C, as in (38), is very straightforward.[16]

(37)



(38)



With the important exception of Aigro (2020), focus particles are not discussed as a source in previous literature on PQPs. Metslang et al. (2017) make the case that conjunctive markers can be a source for PQPs; however, all of their examples are problematic. What they call markers of inferential coordination, e.g. Swedish *alltså* 'thus' and German *also* 'so, therefore' (Metslang et al. 2017: 505), and markers of adversative coordination, e.g. Estonian *aga*, *kuid*, *ent* 'but' (Metslang et al. 2017: 506–7), are not clearly PQPs at all in the present-day languages, as Aigro (2020: 240) observes. Other particles that they discuss as connectives are better analysed as focus markers, like Estonian *kas* and Lithuanian *biau*, *be*. With this in mind, we have found no persuasive evidence for a pathway from coordinating conjunctions to PQPs.[17]

5 Discussion

Table 5 summarizes the five pathways we have proposed in section 4. The list of possible ultimate sources provided here is not intended to be exhaustive.

Our set of pathways is different from the lists of sources that have been presented in previous literature. Some sources that have been prominent in previous work, such as negation, have been subsumed under other pathways (in this case disjunction). Others, such as coordinating conjunctions, are not present at all. And our final pathway, from focus particles, is not discussed at all in previous literature, except by Aigro (2020).

Ideally a set of pathways such as this one would aim for exhaustiveness. Yet there is one particle in our dataset that resists categorization: this is the French particle *ti* (also written *tu*, *ty*, *t'y*), found in several varieties, as in (39).

16 Semantically, too, the two are close: both the standard Hamblin (1973) semantics for questions and the dominant approach to focus interpretation involve sets of propositions, as noted in Rooth (1992: 84–85).

17 Metslang et al. (2017: 501) and, following them, Aigro (2020: 240) state that the PQP *a* in Bulgarian and Macedonian derives from the coordinating conjunction *ja*, citing Englund (1977). We have been unable to trace this claim to that book, or any other source, so we have left these particles out of our dataset.

Pathway	Proximate source structure	Ultimate sources	Sec.	No.
reduced disjunct > disjunctor > PQP	alternative question	negation, verbs, tags	§4.1	44
reduced matrix clause/ subordinator > PQP	biclausal polar question	complementizer, reduced clause, pronoun/copula	§4.2	15
wh-word > higher *wh*-word > PQP	*wh*-question	what, how, why	§4.3	14
epistemic modality > polarity > PQP	polarity-marked declarative	modal adverbs, modal particles	§4.4	4
focus particle > PQP	clause with wide-scope particle	focus particle	§4.5	5
				82

Table 3 Pathways for question particles

(39) *Il vient* ***-tu****?*
he comes Q
'Is he coming?' (Quebec French; Morin 2017: 3)

Morin (2017) discusses Quebec French *tu* in detail, arguing that it is indeed a PQP that is enclitic on the finite verb. It originates in a 'complex inversion' construction, shown in (40): see Roberts (1993: ch. 2), Morin (2009), and the literature cited there.

(40) *Jean vient -il*
Jean comes -he
'Is Jean coming?' (Morin 2017: 82)

Two phonological changes laid the groundwork for the reanalysis of *-t-il* as a PQP. The first is final obstruent deletion, whereby the /t/ in verb forms such as *vient* is not pronounced except in contexts with a following vowel. This motivated the emergence of 'linking /t/' in cases of inversion where no etymological /t/ would be expected, such as *aime-t-il* 'does he love?'. The second

relevant change is the loss of word-final /l/ following high vowels. These two changes conspired to create a sequence /ti/ that was not transparently related to a verb-pronoun sequence as in (40), but restricted to historical inversion contexts such as polar questions, and hence ripe for reanalysis, informally put. Rather than grammaticalization, this change is an instance of EXAPTATION, in Lass's (1990) terms: a functionless ('junk') phonological sequence is repurposed as a PQP.[18]

This example serves to illustrate that the question of 'where does functional category X come from' is unlikely to be answered exhaustively in general: language learners and users are just too creative for that. The best we can do is to catalogue likely pathways, and that is what we have done here.[19] In the remainder of this section we consider whether the five pathways we have identified can be given a uniform syntactic (§5.1) or semantic (§5.2) characterization.

5.1 Formal syntactic approaches to the grammaticalization of PQPs

In section 2 we noted that formal approaches to grammaticalization (Roberts & Roussou 2003, van Gelderen 2004) characterize grammaticalization as involving either i) upward reanalysis or ii) spec-to-head reanalysis. These reanalyses usually give rise to 'lexical split': pairs of homophonous or near-homophonous items instantiating the old and the new analyses, as with Urdu/Hindi *kyā*, where the old *wh*-word exists alongside the PQP/focus marker. Here we revisit the pathways we have presented to see how well this characterization fits them.

The first pathway – disjunction to PQP (§4.1) – is a case of functional material being reanalysed as functional material higher in the structure: a disjunctor embedded under a null matrix interrogative C becomes reanalysed as the exponent of that C. As such it involves upward reanalysis, and straightforwardly exemplifies the effects of van Gelderen's (2004) Late Merge Principle.

18 Contrary to appearances, in all likelihood *tu* does not originate via grammaticalization of the French second-person pronoun *tu*, though the latter may have exerted a secondary analogical influence on the form of the particle in Quebec French (Morin 2017: 83–85). For this reason, and because of the crucial role of the liaison consonant, we do not consider *ti*/*tu* as having grammaticalized from a pronoun.

19 As a reviewer notes, assuming that *all* PQPs arise via pathways such as these raises the question of what the earliest stages of human language looked like: did they mark polar question status exclusively by means of intonation? Lass (2000: 216–218) observes that a corollary of a strong theory of unidirectionality in grammaticalization is that a stage must have existed in which languages were isolating and consisted exclusively of lexical (not grammatical) material (see Walkden 2019: 5–7 for discussion). As the discussion above should make clear, we do not aim for exhaustiveness or assume strict unidirectionality in this paper.

The second pathway – subordinator to PQP (§4.2) – involves either the reanalysis of an embedded C as a matrix C, or the reanalysis of a whole embedding clause as a matrix C. In either case the reanalysis is upward. A direct parallel is provided by the development of the Greek future/modal particle *tha* from the verb *thelo* 'want' plus a complement introduced by (*h*)*ina* (Roberts & Roussou 2003: 58–71), in which a biclausal structure becomes monoclausal.

The third pathway – from *wh*-word to PQP (§4.3) – involves both upward reanalysis and spec-to-head reanalysis. The first step of the pathway, in which a moved *wh*-word becomes reanalysed as first Merged in the left periphery, is another case of the Merge-over-Move preference in action. The second step involves an element in the specifier of CP becoming reanalysed as a C head.[20]

The fourth pathway – from marker of epistemic modality to PQP (§4.4) – also involves both upward reanalysis and spec-to-head reanalysis, but this time the other way round. In a first step, an epistemic adverb in a specifier position in the T-domain is reanalysed as a head. Subsequent steps involve that element being reanalysed as a higher head.

The fifth pathway – from focus particle to PQP (§4.5) – is not so obviously a case of upward reanalysis or spec-to-head reanalysis. Here, one kind of head is analysed as a different kind of head. One question that can be asked here is whether this is grammaticalization in the strict sense at all. One of the hallmarks of grammaticalization is phonological weakening, and there is no real evidence for that in any of the five examples in our dataset. Likewise, it is not clear in what sense the change can be characterized as involving semantic bleaching. An alternative is to view the change from focus particle to PQP as an instance of 'lateral' grammaticalization in the sense of Simpson & Wu (2002: 170), 'a process in which a functional head from one type of syntactic domain may under appropriate circumstances undergo re-interpretation as an equivalent functional head in a second domain', in this case the clausal domain. Given the small number of examples, the nature of this change deserves further study.[21]

A reviewer asks why we use the term grammaticalization for *any* of the five pathways, since none of the proximate sources involves a truly lexical (rather than grammatical) item, and since semantic bleaching and phonological reduction – the classic hallmarks of grammaticalization – are for the most part not in evidence. We have chosen to use the term because four of the five changes we have investigated fit the formal characterization of grammatical-

20 Or *mutatis mutandis* for some projection in an articulated left periphery.

21 As a reviewer notes, focus particles may also arise diachronically through pathways similar to those discussed in section 4: for instance, the focus marker in Tima, a Niger-Congo language, grammaticalizes from a copula in a cleft structure (Becker & Schneider-Blum 2020), an origin similar to the Arabic PQP discussed in section 4.2.

ization found in Roberts & Roussou (2003), but it is worth noting that, for us, the term is a label of convenience rather than anything deeper; the more important fact is that they are instances of upward reanalysis.

A final issue worth addressing in this subsection is how the diachronic origin of PQPs relates to their synchronic analysis. Up to now we have been treating question particles as largely identical. However, as noted in section 2, there are at least two possible positions for question particles to surface overtly in the clause (Holmberg 2016, Monforte 2018b): in the C-domain (Q-ForceP, in Holmberg's 2016 terms) and in the T-domain (PolP, in in Holmberg's 2016 terms). Metslang et al. (2017: 494) also observe that different sources seem to give rise to different linear positions for PQPs. It is not possible for us to investigate this issue systematically in our dataset in the scope of this paper, but impressionistically there does seem to be a link. The examples we have discussed of PQPs derived via the first three pathways – disjunction, subordination and *wh*-words – all seem to involve clause-peripheral particles. But this is not true of the fourth pathway, epistemic modality: Basque *al* and Niuean *kia* and *ka* can all surface clause-internally, suggesting that they may be in a lower position. This may relate to the fact that the pathway from epistemic modality to PQP involves reanalysis of an element that is already in the T-domain. We leave an in-depth study of the positional question to future research.

5.2 Semantic approaches to the grammaticalization of PQPs

One may also wonder whether the pathways we discuss here have a unified characterization at a semantic, as opposed to syntactic, level. A unified semantic approach to PQPs is pursued by Aigro (2020), who argues that polarity is central to the emergence of PQPs. Her claim is that the six sources she identifies 'can all be regarded as markers of polarity, marking the negation, affirmation or open polarity of a clause' (Aigro 2020: 240). For the proximate sources discussed in this paper, this is less obviously the case. Aigro does discuss *wh*-words, suggesting that these are markers of open polarity. But this does not seem to be an apt characterization of all *wh*-words: *why p*, for instance, presupposes the truth of *p* (Lawler 1971: 67). It is also not clear whether epistemic modality markers or focus particles bear any relation to polarity.

A very loose semantic characterization of the proximate sources for PQPs in our five pathways is as follows: they are all elements that can take propositional arguments. Zimmermann (2000) makes the case that both disjunctions and epistemic modality involve the generation of sets of alternative propositions. The same is true of focus at the clausal level in Rooth's (1992) theory.

By contrast, subordinators and high (*why*-like) *wh*-words do not obviously generate sets of propositions.[22] All five sources involve a propositional argument as input, however.

6 Conclusion

In this paper we have presented a survey of 83 polar question particles from 67 languages. Of these, more than half most plausibly have their origins in disjunction. The remaining particles can all be traced back via four other pathways: subordinators, *wh*-words, markers of epistemic modality, and focus particles (with one exception: French *ti*).

Our study has proposed some sources/pathways for PQPs that have received little or no attention in the literature to date, such as *wh*-words and focus particles. At the same time, it has called others – such as non-disjunctive coordination – into question. Four of our five pathways fit the predictions of formal approaches to grammaticalization, in that they involve either upward reanalysis or spec-to-head reanalysis, and the fifth – focus particles – is unlikely to be grammaticalization at all. Semantically, we have tentatively proposed that the proximate diachronic sources for PQPs all take propositions as arguments. There remain many avenues for future work on the diachrony of PQPs, some of which we have flagged up in the article.

References

Aigro, Mari. 2020. Polar question particles and their sources: a semantic approach to grammaticalization. *Transactions of the Philological Society* 118(2). 237–253.

Aldridge, Edith. 2011. Neg-to-Q: The historical origin and development of question particles in Chinese. *The Linguistic Review* 28. 411–447.

Bailey, Laura. 2012. *The syntax of question particles*: Newcastle University dissertation.

Baker, Mark & James McCloskey. 2007. On the relationship of typology to theoretical syntax. *Linguistic Inquiry* 11. 285–296.

Bayer, Josef & Hans-Georg Obenauer. 2011. Discourse particles, clause structure, and question types. *The Linguistic Review* 28. 449–491.

Becker, Laura & Gertrud Schneider-Blum. 2020. Morphological marking of contrast in Tima. *Glossa* 5. 125.

Bencini, Giulia. 2003. Toward a diachronic typology of yes/no question con-

22 With the exception of question-subordinators like present-day English *whether*, of course, which do generate sets of propositions.

structions with particles. In *Proceedings of Chicago Linguistics Society 39*, 604–621.

Bhatia, Tej K. 1993. *Punjabi: A cognitive-descriptive grammar*. London: Routledge.

Bhatt, Rajesh & Veneeta Dayal. 2020. Polar question particles: Hindi-Urdu *kya:*. *Natural Language and Linguistic Theory* 38. 1115–1144.

Buchholz, Oda & Wilfried Fiedler. 1987. *Albanische Grammatik*. Leipzig: Verlag Enzyklopädie.

Butt, Miriam, Farhat Jabeen & Tina Bögel. 2020. Ambiguity resolution via the syntax-prosody interface: The case of *kya* 'what' in Urdu/Hindi. In Gerrit Kentner & Joost Kremers (eds.), *Prosody in syntactic encoding*, 85–117. Berlin: De Gruyter.

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago, IL: University of Chicago Press.

Chen, Zhuo. 2022. *No matter whether you ask: Yes-no questions and their kin in Mandarin*. Los Angeles: UCLA dissertation.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: essays on Minimalist syntax in honor of Howard Lasnik*, 89–156. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: a life in language*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36. 1–22.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. Oxford: Oxford University Press.

Coniglio, Marco. 2022. On the adverbial origin of German modal particles. In Xabier Artiagoitia, Arantzazu Elordieta & Sergio Monforte (eds.), *Discourse particles*, 13–40. Amsterdam: John Benjamins.

Corr, Alice. 2018. Matrix complementisers and 'speech act' syntax: Formalising insubordination in Catalan and Spanish. In Janine Berns, Haike Jacobs & Dominique Nouveau (eds.), *Romance Languages and Linguistic Theory 13: Selected papers from 'Going Romance' 29, Nijmegen*, 75–94. Amsterdam: John Benjamins.

Creissels, Denis. 2020. Grammaticalization in Manding languages. In Walter Bisang & Andrej Malchukov (eds.), *Grammaticalization scenarios: cross-linguistic variation and universal tendencies*, vol. 2, 695–728. Berlin: De Gruyter.

Czypionka, Anna, Maribel Romero & Josef Bayer. 2021. Question-sensitive discourse particles at the interfaces of syntax, semantics and pragmatics

– an experimental approach. *Glossa* 6. 24.
Danusugondo, Purwanto. 1975. *Bahasa Indonesian: Indonesian for beginners. volumes 1 and 2*. Sydney: Sydney University Press.
Dayley, Jon P. 1989. *Tümpisa (Panamint) Shoshone grammar*. Berkeley: University of California Press.
Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization*. Amsterdam: Benjamins.
Druetta, Ruggero. 2003. *Qu'est-ce que tu fais?* État d'avancement de la grammaticalisation de *est-ce que*. Deuxième partie. *Linguæ &* 2003. 21–35.
Dryer, Matthew S. 2013. Position of polar question particles. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/92.
Du Feu, Veronica. 1996. *Rapanui*. London: Routledge.
Duffield, Nigel. 2007. Aspects of Vietnamese clausal structure: separating tense from assertion. *Linguistics* 45. 765–814.
Eckardt, Regine & George Walkden. 2022. A particle-like use of *hwæþer*: Wisdom's questions in Boethius. In Xabier Artiagoitia, Arantzazu Elordieta & Sergio Monforte (eds.), *Discourse particles*, 41–64. Amsterdam: John Benjamins.
Eid, Mushira. 1991. Verbless sentences in Arabic and Hebrew. In Janine Berns, Haike Jacobs & Dominique Nouveau (eds.), *Perspectives on Arabic linguistics III*, 31–61. Amsterdam: John Benjamins.
Eid, Mushira. 1992. Pronouns, questions, and agreement. In Ellen Broselow, Mushira Eid & John McCarthy (eds.), *Perspectives on Arabic linguistics IV*, 107–141. Amsterdam: John Benjamins.
Elsig, Martin. 2009. *Grammatical variation across space and time: The French interrogative system*. Amsterdam: Benjamins.
Englund, Birgitta. 1977. *Yes/no-questions in Bulgarian and Macedonian*. Stockholm: Almqvist & Wiksell.
Evans, Nicholas. 2007. Insubordination and its uses. In Irina Nikolaeva (ed.), *Finiteness: theoretical and empirical foundations*, 366–431. Oxford: Oxford University Press.
Gary, Judith Olmsted & Saad Gamal-Eldin. 1982. *Cairene Egyptian Colloquial Arabic*. Amsterdam: North-Holland.
van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam: John Benjamins.
van Gelderen, Elly. 2007. The definiteness cycle in germanic. *Journal of Germanic Linguistics* 19(4). 275–308.
van Gelderen, Elly. 2009. Renewal in the left periphery: Economy and the

complementiser layer. *Transactions of the Philological Society* 107. 131–195.
van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. Amsterdam: Benjamins.
van Gelderen, Elly. 2015. The copula cycle. *Lingue e linguaggio* 14. 287–301.
van Gelderen, Elly. 2021. *Third factors in language variation and change*. Cambridge: Cambridge University Press.
Glinert, Lewis. 1989. *The grammar of modern Hebrew*. Cambridge: Cambridge University Press.
Gonzalez, Aurore. 2023. Interrogative particles in polar questions: the view from Finnish and Turkish. *Glossa* 8(1). 1–47.
Greenberg, Joseph H. 1978. How does a language acquire gender markers? In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Word structure*, vol. 3 Universals of human language, 47–82. Stanford, CA: Stanford University Press.
Gyuris, Beáta. 2017. New perspectives on bias in polar questions: A study of Hungarian *-e*. *International Review of Pragmatics* 9(1). 1–50.
Hackstein, Olav. 2004. Rhetorical questions and the grammaticalization of interrogative pronouns as conjunctions in Indo-European. In Adam Hyllested, Anders Richard Jørgensen, Jenny Helena Larsson & Thomas Olander (eds.), *Per aspera ad asteriscos. Studia Indogermanica in honorem Jens Elmegård Rasmussen sexagenarii Idibus Martiis anno MMIV*, 167–186. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
Hagstrom, Paul. 1998. *Decomposing questions*: Massachusetts Institute of Technology dissertation.
Hagstrom, Paul. 2000. The movement of question particles. In *Proceedings of NELS 30*, vol. 1, 275–286.
Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20: Essays in honor of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT Press.
Hamblin, Charles L. 1958. Questions. *The Australasian Journal of Philosophy* 36. 159–168.
Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10. 41–53.
Hammond, Lila. 2005. *Serbian: an essential grammar*. London: Routledge.
Han, Chung-Hye. 2002. Interpreting interrogatives as rhetorical questions. *Lingua* 112. 201–229.
Harris, Alice & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
Harris, Martin. 1988. French. In Martin Harris & Nigel Vincent (eds.), *The*

Romance languages, 209–245. London: Croom Helm.
Hayashi, Makoto. 2010. An overview of the question‐response system in Japanese. *Journal of Pragmatics* 42(10). 2685–2702. Question-Response Sequences in Conversation across Ten Languages.
Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
Hendery, Rachel. 2012. *Relative clauses in time and space: a case study in the methods of diachronic typology*. Amsterdam: John Benjamins.
Hinds, John. 1986. *Japanese*. London: Croom Helm.
Holmberg, Anders. 2016. *The syntax of yes and no*. Oxford: Oxford University Press.
Hopper, Paul & Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press 2nd edn.
Iatridou, Sabine & Hedde Zeijlstra. To appear. On the formal identity between interrogative and conditional complementizers: synchrony or diachrony? Ms., MIT/Göttingen. https://www.heddezeijlstra.org/.
Iwasaki, Shoichi & Preeya Ingkaphirom. 2005. *A reference grammar of Thai*. Cambridge: Cambridge University Press.
Jayaseelan, K. A. 2012. Question particles and disjunction. *Linguistic Analysis* 38. 35–51.
Jenner, Henry. 1904. *A handbook of the Cornish language*. London: D. Nutt.
Kieviet, Paulus. 2017. *A grammar of Rapa Nui*. Berlin: Language Science Press.
Kilian-Hatz, Christa. 2008. *A grammar of modern Khwe (Central Khoisan)*. Cologne: Rüdiger Köppe Verlag.
König, Ekkehard & Peter Siemund. 2007. Speech act distinctions in grammar. In Timothy Shopen (ed.), *Language typology and syntactic description*, 276–324. Cambridge: Cambridge University Press.
Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press 2nd edn.
Kuwabara, Kazuki. 2013. Peripheral effects in Japanese questions and the fine structure of CP. *Lingua* 126. 92–119.
Lass, Roger. 1990. How to do things with junk: exaptation in language evolution. *Journal of Linguistics* 26. 79–102.
Lass, Roger. 2000. Remarks on unidirectionality. In Olga Fischer, Anette Rosenbach & Dieter Stein (eds.), *Pathways of change: Grammaticalization in english*, 207–227. Amsterdam: John Benjamins.
Launer, Michael K. 1974. *Elementary Russian syntax*. Cambridge, MA: Slavica Publishers.
Lawler, John. 1971. *Any* questions? In *Papers from the seventh regional meeting*

of the Chicago Linguistic Society, 163–173.
Lohndal, Terje. 2009. The copula cycle. In Elly van Gelderen (ed.), *Cyclical change*, 209–242. Amsterdam: John Benjamins.
MacAulay, Donald. 1992. *The Celtic languages*. Cambridge: Cambridge University Press.
Mackridge, Peter. 1985. *The modern Greek language: a descriptive analysis of standard modern Greek*. Oxford: Oxford University Press.
Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4. 307–333.
Matthews, Stephen & Virginia Yip. 1994. *Cantonese: A comprehensive grammar*. London: Routledge.
Mayo, Peter. 1993. Belorussian. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 887–946. London: Routledge.
McCawley, James D. 1988. *The syntactic phenomena of English*. Chicago, IL: University of Chicago Press.
McGloin, Naomi Hanaoka. 1980. Some observations concerning no desu expressions. *The Journal of the Association of Teachers of Japanese* 15(2). 117–149.
Metslang, Helle, Külli Habicht & Karl Pajusalu. 2011. Developmental paths of interrogative particles: the case of Estonian. *Folia Linguistica Historica* 32. 149–188.
Metslang, Helle, Külli Habicht & Karl Pajusalu. 2017. Where do polar question markers come from? *Sprachtypologie und Universalienforschung* 70. 489–521.
Mitchell, Bruce. 1985. *Old English syntax*. Oxford: Clarendon.
Monforte, Sergio. 2018a. About the particle *al* in Basque. In Marta Díaz Ferro, Jorge Diz Ferreira, Ania Pérez Pérez & Ana Varela Suárez (eds.), *Novas perspectivas na lingüística aplicada*, 19–31. Lugo: Axac.
Monforte, Sergio. 2018b. Question particles in Basque. *Isogloss* 4. 29–53.
Moravcsik, Edith. 1971. *Some cross-linguistic generalizations about yes-no questions and their answers*: Stanford University dissertation.
Morin, Annick. 2009. On the Quebec French interrogative particle *tu*. In Enoch O. Aboh, Elisabeth van der Linden, Josep Quer & Petra Sleeman (eds.), *Romance languages and linguistic theory: Selected papers from 'Going Romance', Amsterdam 2007*, 201–222. Amsterdam: Benjamins.
Morin, Annick. 2017. *Questioning particles: A cross-linguistic approach to Quebec French polar interrogatives*: University of Toronto dissertation.
Moroz, George. 2017. *lingtypology: easy mapping for linguistic typology*. https://CRAN.R-project.org/package=lingtypology.
Munaro, Nicola & Hans-Georg Obenauer. 1999. On underspecified *wh-*

elements in pseudo-interrogatives. *University of Venice Working Papers in Linguistics* 9. 181–253.
Nguyễn, Đình-Hoà. 1997. *Vietnamese*. Amsterdam: John Benjamins.
Nichols, Johanna, Alena Witzlack-Makarevich & Balthasar Bickel. 2013. The AUTOTYP genealogy and geography database. 2013 release.
Ostrowski, Norbert. 2012. Old Lithuanian interrogative particle *biau*. *Baltistica* 47. 261–265.
Pan, Victor Junnan & Waltraud Paul. 2016. Why Chinese SFPs are neither optional nor disjunctors. *Lingua* 170. 23–34.
Parkinson, Stephen. 1988. Portuguese. In Martin Harris & Nigel Vincent (eds.), *The Romance languages*, 131–169. London: Croom Helm.
Parkinson, Stephen. 2009. Portuguese. In Martin Harris & Nigel Vincent (eds.), *The Romance languages*, 217–232. London: Routledge 2nd edn.
Paul, Waltraud & Victor Junnan Pan. 2017. What you see is what you get: Chinese sentence-final particles as head-final complementizers. In Josef Bayer & Volker Struckmeier (eds.), *Discourse particles: Formal approaches to their syntax and semantics*, 49–77. Berlin: Mouton de Gruyter.
Press, Ian. 2004. *Standard Breton*. Munich: Lincom Europa.
Priestly, T. M. S. 1993. Slovene. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 388–451. London: Routledge.
Prieto, Pilar & Gemma Rigau. 2007. The syntax-prosody interface: Catalan interrogative sentences headed by *que*. *Journal of Portuguese Linguistics* 6. 29–59.
Rice, Keren. 1989. *A grammar of Slave*. Berlin: Mouton de Gruyter.
Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.
Rizzi, Luigi. 2001. On the position Int(errogative) in the left periphery of the clause. In Guglielmo Cinque & Giampaolo Salvi (eds.), *Current studies in Italian syntax: Studies offered to Lorenzo Renzi*, 287–296. Amsterdam: Elsevier.
Roberts, Ian. 1993. *Verbs and diachronic syntax: a comparative history of English and French*. Dordrecht: Kluwer.
Roberts, Ian & Anna Roussou. 1999. A formal approach to "grammaticalization". *Linguistics* 37. 1011–1041.
Roberts, Ian & Anna Roussou. 2003. *A Minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.
Rohde, Hannah. 2006. Rhetorical questions as redundant interrogatives. *San Diego Linguistics Papers* 2. 134–168.
Romero, Maribel & Chung-Hye Han. 2004. On negative *yes*/*no* questions. *Linguistics and Philosophy* 27. 609–658.
Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Seman-*

tics 1. 75–116.

Roussou, Anna. 2015. Is particle a (unified) category? In Josef Bayer, Roland Hinterhölzl & Andreas Trotzke (eds.), *Discourse-oriented syntax*, 121–158. Amsterdam: John Benjamins.

Sadock, Jerrold M. & Arnold M. Zwicky. 1985. Speech act distinctions in syntax. In Timothy Shopen (ed.), *Language typology and syntactic description*, 155–196. Cambridge: Cambridge University Press.

Saeed, John I. 1999. *Somali*. Amsterdam: John Benjamins.

Sailor, Craig. 2009. *Tagged for deletion: A typological approach to VP ellipsis in tag questions*. Los Angeles UCLA MA thesis. Version incorporating organizational changes of 2011.

Scherf, Nathalie. 2017. The syntax of Swedish modal particles. In Josef Bayer & Volker Struckmeier (eds.), *Discourse particles: Formal approaches to their syntax and semantics*, 78–99. Berlin: Mouton de Gruyter.

Shevelov, George Y. 1993. Ukrainian. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, 947–998. London: Routledge.

Simoncsics, Péter. 2003. Egy uráli párhuzamosság: a nyenyec költői dikció ngey (ngæy) töltőeleme és a magyar -é szintaktomorféma. *Nyelvtudományi Közlemények* 100. 238–244.

Simpson, Andrew & Zoe Wu. 2002. IP-raising, tone sandhi and the creation of particles: Evidence for cyclic spell-out. *Journal of East Asian Linguistics* 11. 67–99.

Sneddon, James N. 1996. *Indonesian reference grammar*. St. Leonards, Australia: Allen and Unwin.

Soltan, Usama. 2011. On strategies of question-formation and the grammatical status of the Q-particle *huwwa* in Egyptian Arabic *Wh*-questions. *UPenn Working Papers in Linguistics* 17(1). 215–224.

Starks, Donna & Diane Massam. 2015. The origin of yes–no question particles in the Niuean language. *Journal of Linguistics* 51(1). 185–212.

Starr, William B. 2014. What 'if'? *Philosophers' Imprint* 14. 1–27.

Sudo, Yasutada. 2013. Biased polar questions in English and Japanese. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 275–295. Leiden: Brill.

Tailleur, Sandrine. 2013. *The French* wh*-interrogative system:* Est-ce que*, clefting?*: University of Toronto dissertation.

Tan, Tamisha L. 2025. Inclined to agree: From pronominal copula to predicative agreement. *Journal of Historical Syntax* 9. To appear. Special issue: Proceedings of the 23rd Diachronic Generative Syntax (DiGS) Conference.

Trail, Ronald L. 1970. *The grammar of Lamani*. Norman: University of Oklahoma.

Traugott, Elizabeth Closs. 2017. 'insubordination' in the light of the Uniformitarian Principle. *English Language and Linguistics* 21(2). 289–310.

Trotzke, Andreas & Anna Czypionka. 2022. Non-canonical questions from a comparative perspective: Introduction to the special collection. *Linguistics Vanguard* 8(s2). 205–207.

Trotzke, Andreas & Sergio Monforte. 2019. Basque question particles: Implications for a syntax of discourse particles. *Linguistic Variation* 19(2). 352–385.

Ultan, Russell. 1978. Some general characteristics of interrogative systems. In Joseph H. Greenberg (ed.), *Universals of human language*, vol. 4, 211–248. Stanford, CA: Stanford University Press.

Walkden, George. 2013. The status of *hwæt* in Old English. *English Language and Linguistics* 17. 465–488.

Walkden, George. 2014. *Syntactic reconstruction and Proto-Germanic*. Oxford: Oxford University Press.

Walkden, George. 2019. The many faces of uniformitarianism in linguistics. *Glossa* 4. 1–17.

Walkden, George. 2021. Against mechanisms: towards a minimal theory of change. *Journal of Historical Syntax* 5(33). 1–27. Special issue: Whither Reanalysis?

Willis, David, Christopher Lucas & Anne Breitbarth. 2013. Comparing diachronies of negation. In David Willis, Christopher Lucas & Anne Breitbarth (eds.), *The history of negation in the languages of Europe and the Mediterranean*, vol. 1, 1–50. Oxford: Oxford University Press.

Yanagida, Yuko. 1995. *Focus projection and wh-head movement*. Ithaca, NY: Cornell University dissertation.

Zimmermann, Thomas Ede. 2000. Free choice disjunction and epistemic possibility. *Natural Language Semantics* 8. 255–290.

Appendix: Languages and particles included in the survey

Language	Family	Particle	Position	Path	Source
Albanian	I-E	*mos*	Initial	Disj.	Buchholz & Fiedler (1987)
Albanian	I-E	*a*	Initial	Disj.	Bencini (2003: 610)
Amharic	Afro-As.	*way*	Final	Disj.	Bencini (2003: 610)
Arabic (Egyptian)	Afro-As.	*huwwa*	Initial	Sub.	Gary & Gamal-Eldin (1982: 4)
Arabic (Egyptian)	Afro-As.	*hijja*	Initial	Sub.	Gary & Gamal-Eldin (1982: 4)
Arabic (Egyptian)	Afro-As.	*humma*	Initial	Sub.	Gary & Gamal-Eldin (1982: 4)
Arabic (Syrian)	Afro-As.	*hal*	Final	Disj.	Bencini (2003: 610)
Arabic (Turku)	Afro-As.	*wála*	Final	Disj.	Kuteva et al. (2019: 308)
Basque	Isolate	*ote*	Other	Disj.	Trotzke & Monforte (2019)
Basque	Isolate	*al*, *a*	Other	Epis.	Monforte (2018a)
Belorussian	I-E	*ci*	Initial	Pron.	Mayo (1993: 926)
Breton	I-E	*ha*(*g*)	Initial	Sub.	Press (2004: 57)
Bulgarian	I-E	*da*	Initial	Sub.	Metslang et al. (2017)
Cantonese	Sino-Tibetan	*me*	Final	Disj.	Matthews & Yip (1994: 311)
Cornish	I-E	*na*	Initial	Disj.	Jenner (1904: 161)
English (colloq.)	I-E	*innit*	Final	Disj.	Sailor (2009)
English (Old)	I-E	*hwæþer*	Initial	Sub.	Eckardt & Walkden (2022)
English (Old)	I-E	*hu* (*ne*)	Initial	Pron.	Mitchell (1985: 680)
Estonian	Uralic	*või*/*vä*	Final	Disj.	Aigro (2020: 244–5)
Estonian	Uralic	*ega*	Initial	Disj.	Aigro (2020: 247–8)
Estonian	Uralic	*kas*	Initial	Foc.	Aigro (2020: 242–4)
Estonian	Uralic	*ja*/*ju*/*jo*	Initial	Foc.	Aigro (2020: 248–50)
Estonian	Uralic	*es*, *eks*	Initial	Disj.	Aigro (2020: 245–7)
Finnish	Uralic	*ko*	Second	Pron.	Metslang et al. (2017)
Florentine	I-E	*che*	Initial	Pron.	van Gelderen (2009: 148)
French	I-E	*est-ce que*	Initial	Sub.	Harris (1988: 237)
French	I-E	*tu*, *ti*, *ty*	Other	None	Morin (2017)
Georgian	Kartvelian	*tu ara*	Other	Disj.	Harris & Campbell (1995: 295)

Sources for question particles

Language	Family	Particle	Position	Path	Source
Greek	I-E	*mípos*	Other	Epis.	Mackridge (1985: 301)
Guadalupe Creole	Isolate	*anh/onh*	Final	Disj.	Bencini (2003: 610)
Hausa	Afro-As.	*kō*	Initial	Disj.	Bencini (2003: 610)
Hebrew	Afro-As.	*ha'im*	Initial	Sub.	Glinert (1989: 271)
Hungarian	Uralic	*e*	Final	Disj.	Bencini (2003: 610)
Indonesian	Austron.	*apa*	Initial	Pron.	
Irish	I-E	*an*	Initial	Sub.	MacAulay (1992)
Japanese	Japanese	*no*	Final	Sub.	Hayashi (2010: 2646)
Kannada	Dravidian	*e:nu*	Final	Pron.	König & Siemund (2007)
Kannada	Dravidian	*a*	Final	Disj.	Bencini (2003: 610)
Khwe	Khoe	*rè*	Final	Disj.	Kilian-Hatz (2008: 297)
Kobon	Trans-NG	*aka*	Final	Disj.	Bencini (2003: 610)
Lamani	I-E	*kããi/ka*	Final	Pron.	Trail (1970: 63)
Lao	Tai-Kadai	*boh*	Final	Disj.	Bencini (2003: 610)
Latin	I-E	*ne*	Final	Disj.	Bencini (2003: 610)
Latin	I-E	*an*	Initial	Disj.	Bencini (2003: 610)
Latvian	Uralic	*vai*	Initial	Disj.	Metslang et al. (2017)
Limbu	Sino-Tibetan	*i*	Final	Disj.	Bencini (2003: 610)
Lithuanian	I-E	*ar*	Initial	Foc.	Aigro (2020)
Lithuanian (Old)	I-E	*biau, be*	Initial	Foc.	Ostrowski (2012)
Macedonian	I-E	*da*	Initial	Sub.	Metslang et al. (2017)
Malayalam	Dravidian	*o*	Final	Disj.	Jayaseelan (2012: 36)
Mandarin	Sino-Tibetan	*ma*	Final	Disj.	Aldridge (2011)
Mandinka	Mande	*fó*	Initial	Disj.	Creissels (2020: 724)
Moré	Atlantic-Congo	*bi*	Final	Disj.	Kuteva et al. (2019: 307)
Niuean	Austron.	*ka*	Other	Epis.	Starks & Massam (2015)
Niuean	Austron.	*kia*	Other	Epis.	Starks & Massam (2015)
Niuean	Austron.	*nakai*	Other	Disj.	Starks & Massam (2015)
Nung	Tai-Kadai	*mi*	Final	Disj.	Bencini (2003: 610)
Panjabi	I-E	*kii*	Initial	Pron.	Bhatia (1993: 4)
Polish	I-E	*czy*	Initial	Pron.	Hackstein (2004: 175)
Polish	I-E	*li*	Second	Disj.	Metslang et al. (2017)
Portuguese	I-E	*é que*	Initial	Sub.	Parkinson (1988: 158)

Language	Family	Particle	Position	Path	Source
Quechua (Imbabura)	Quechuan	*-chu*	Final	Disj.	Bencini (2003: 610)
Rapa Nui	Austron.	*hoki*	Initial	Sub.	Kieviet (2017: 480–482, 490)
Romanian	I-E	*au*	Initial	Disj.	Bencini (2003: 610)
Russian	I-E	*li*	Second	Disj.	Launer (1974: 24)
Sango	Isolate	*wala*	Final	Disj.	Bencini (2003: 610)
Serbian	I-E	*li*	Second	Disj.	Hammond (2005: 63)
Slave	Na-Dene	*nį/élį*	Final	Foc.	Rice (1989: 22, 1124, 1130)
Slovene	I-E	*kaj*	Initial	Pron.	Priestly (1993: 430)
Somali	Afro-As.	*ma*	Other	Disj.	Saeed (1999: 197)
Swedish (Solf)	I-E	*åm*	Initial	Sub.	Aigro (2020: 240)
Thai	Tai-Kadai	*rú*	Final	Disj.	Iwasaki & Ingkaphirom (2005: 279–286)
Thai	Tai-Kadai	*rú-plàaw*	Final	Disj.	Iwasaki & Ingkaphirom (2005: 279–286)
Thai	Tai-Kadai	*máy*	Final	Disj.	Iwasaki & Ingkaphirom (2005: 279–286)
Tibetan	Sino-Tibetan	*am*	Final	Disj.	Bencini (2003: 610)
Tocharian A	I-E	*aśśi*	Final	Disj.	Hackstein (2004: 174–5)
Tümpisa Shoshone	Uto-Aztecan	*ha*	Second	Pron.	Dayley (1989: 44–45, 324–325)
Turkish	Turkic	*mi*	Final	Pron.	Harris & Campbell (1995: 295)
Tzotzil	Mayan	*mi*	Initial	Sub.	König & Siemund (2007)
Ukrainian	I-E	*čy*	Initial	Pron.	Shevelov (1993: 978)
Urdu/Hindi	I-E	*kyā*	Initial	Pron.	Bhatt & Dayal (2020)
Vietnamese	Austro-Asiatic	*không*	Final	Disj.	Bencini (2003: 610)
Yoruba	Atlantic-Congo	*bi*	Final	Disj.	Bencini (2003: 610)

Sources for question particles

George Walkden, Tian Li & Laura Weisshaar
Department of Linguistics
University of Konstanz
PO Box D 175
78457 Konstanz
Germany
george.walkden@uni-konstanz.de
www.ling.uni-konstanz.de/walkden/