Whys and Wherefores: the Aetiology of the Left Periphery (with reference to Vietnamese)

Nigel Duffield
Konan University, Japan[1]

- Minor edits 1/24/25

Pour l'enfant, amoureux de cartes et d'estampes,
L'univers est égal à son vaste appétit.
Ah ! que le monde est grand à la clarté des lampes !
Aux yeux du souvenir que le monde est petit !...

Charles Baudelaire, *Le Voyage (Les Fleurs du Mal)*

Abstract: This paper offers a detailed description of the left periphery of embedded clauses in Vietnamese. Five kinds of pre-subject constituent are considered in isolation, and in interaction with one another: subordinating conjunctions; embedded topics; fronted quantifier expressions; fronted adverbials, and the Vietnamese equivalent of English *why* (Italian *perché*). A systematic comparison is made with the functional sequence of Italian, proposed by Rizzi & Bocci (2017). Whilst largely consistent with the Italian pattern, our findings diverge in certain respects, esp. suggesting a modification of previous treatments of the *'*why*-to' constraint (Shlonsky & Soare 2011).

Keywords: Vietnamese Syntax; 'Left Periphery'; Complementation; Embedded Topicalization; '*Wh*-questions'; '*Why*'-questions; Scope and Constituency; Cartographic approaches.

Introduction: Why Peripheral, Why?

In this paper I direct attention to two *whys*: the theoretical question "*Why* CP*?*", and a set of questions concerning the left periphery in Vietnamese, ultimately focussing on the contrastive behaviour of the lexical item *tại sao (why, perchè*), relative to all other *wh*-expressions. Whilst the former question is paramount, I will argue that an answer to it is best arrived at by understanding the latter contrasts: hence, the empirical focus of this paper is *why~tại sao,* and on its interaction with other elements of the left periphery (in English and Vietnamese, respectively).

[1] Acknowledgement. Were it not for the timely and constructive responses to my countless requests for judgments, this paper would never have made it past the preliminary remarks stage: I am therefore exceptionally grateful to Ly Pham and Trang Phan for their generous help.

Let's begin, though, with the conceptual questions facing any cartographic approach to word order, namely, *why* is CP where it is? Up there? On the left (or sometimes, right)[2] periphery? Indeed, *Why* is CP considered 'peripheral' at all—part of another territory in the satirical map of *The Derivation* in Fig 1., posted more than a decade ago?

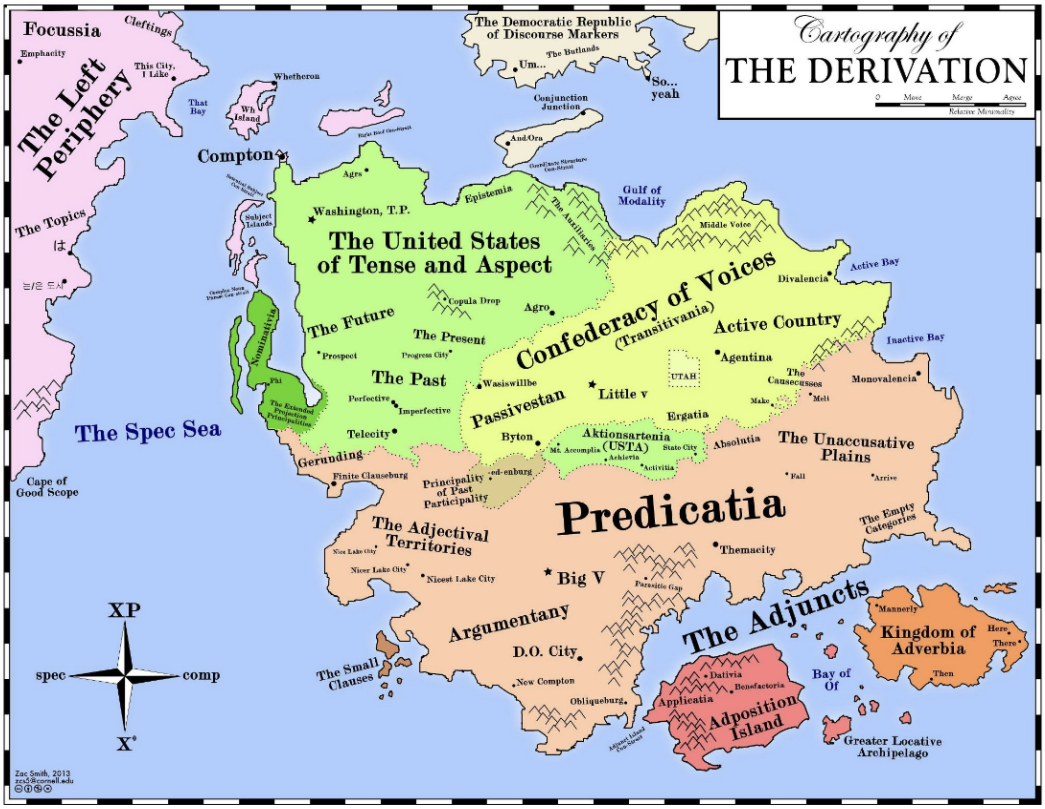


Fig. 1. Cartography of the Derivation.[3]

Designations matter greatly, both for analysis and theory construction. Although strictly speaking the term periphery denotes simply the edge or circumference of an object or region (compare the Parisian *Périphérique*), and is neutral between an inclusive ('integral') and an exclusive ('adjacent but separate') denotation, it is clear that the latter—non-integral, interpretation—is the more salient in normal discourse; in the case of computer hardware, for example, peripherals are additional, non-essential components. In other domains of inquiry, essential components are never treated as 'peripheral', no matter their distance from the centre. Aeronautical engineers don't consider the nose cone or empennage to be 'peripheral' structures on an aircraft, even though the tailplane is not an integral part of the fuselage. No one—other than the thoracic surgeon,

[2] Beyond this footnote, I will largely ignore apparently 'head-final'/ left-branching languages, such as Japanese or Korean. For convenience, I will assume, following Kayne (1994), Sheehan, Biberauer *et al.* (2017), Roberts (2019), and others, that that these are not simply the mirror-image of their right-branching counterparts (though *cf.* Saito (2015): rather, that SOV-I-C languages are underlyingly head-initial varieties, where 'generalized roll-up' has taken place (i.e., iterated phrasal movement to higher specifier positions). On that (probably erroneous) view, the left periphery is a universal domain, appearances to the contrary notwithstanding.

[3] © Zac Smith: used with permission.

perhaps—conceives of the head as peripheral to human anatomy, despite its being non-integral to the torso; this is in obvious contrast to how we think about the limbs (which are more intricately attached). In syntactic theorizing, on the other hand, generativists have tended to treat CP as external to the root: indeed, in the literal chart reproduced as Fig. 1., the left periphery is drawn as another country, something best left alone. (Which it often was, until Rizzi (1997), except whenever subordinate clauses—and extractions therefrom—were at issue (Chomsky 1977; 2000, 2005, 2008).[4]

The original impulse to exclude CP no doubt stems from the earliest conceptions of re-write grammars, in which S was taken as the initial symbol (Chomsky 1957, 1965):[5] if your top-down theory of phrase structure ("S→NP VP") has you start with what is now construed as a lower node (S = IP~TP), then everything to the left of this node inevitably becomes problematic (in a right-branching language; see fn 1). Subsequent Minimalist theorizing—by definition, proceeding from the bottom up—generally has had little to say about the derivation of root clauses beyond TP, except where features of a monolithic C are taken to trigger subject raising (Duffield, 2024), or in cases of A'-movement.[6]

It is worth noting, by contrast, that lower functional categories such as TP and NegP have always been treated as integral to S, even though they are also semantically peripheral to the core proposition: just like CP, TP is excluded from the core thematic domain (vP^{max}, *cf*. "Complete Functional Complex" (Chomsky 1986: 169). The inclusion of TP presumably has to do with the fact that bare propositions *'John love Mary'* and its permutations {SOV, VSO, VOS…} are grammatically unacceptable after the age of two—*Tarzan* movies, aphasic patients, and special learner populations aside; see Deprez & Pierce (1993), Guilfoyle & Noonan (1992); though *cf*. Chung & McCloskey (1987). In short, there is consensus that TP is required in root clauses, for grammatical well-formedness. And, since TP invariably dominates Aspect phrase, the latter projection is perforce also considered integral: *within*, not *without*. Compare the location of 'The United States of Tense and Aspect' in Fig.1 above.[7]

Developing recent trends in cartography, the conclusion here will be that some CP components are required for syntactic well-formedness, even in root clauses (and even if these elements often go unpronounced). In other words, CP may be top-most,

---

[4] Clearly, this phrase-structural notion of *peripheral* is at least partially separate from the "core" *vs*. "periphery" distinction, in discussions of the shape of internally represented grammars ('I-language", "narrow syntax" etc.); see Chomsky (1981, 1985, 1998); *cf*. Hyams (1986), Deprez & Pierce (1993); Culicover (1999), Moltmann (2020). And CP is obviously crucial to Phase theory: see Citko (2014). However, the CP of Phase Theory is usually treated as an undifferentiated single layer of structure, an escape hatch for movement: almost all the theoretical action takes place further down, in TP or below.

[5] Lasnik & Uriagareka (2022: 66) observe that Chomsky (1973) reverses the standard order of S and S' (S' → Comp S, Bresnan 1970) such that the initial symbol S dominates Comp (S —> Comp S'); in other words, root clauses start at (what would now be labelled) CP. However, this revision was quickly abandoned in Chomsky (1977): from Chomsky (1981) onwards, CP re-assumes its 'no-man's land' status.

[6] Emonds (1969: 6) offers a more qualified definition of root that permits nodes above S: "A root will mean either the highest S in a tree, an S immediately dominated by the highest S or the reported S in indirect discourse…" This definition is further elaborated in Emonds (2004).

[7] The non-peripheral treatment of NegP is likely due to its clause-medial position in the surface syntax of English. If the original object language of generative theory had been Modern Irish, or Chamorro, rather than English (though *cf*. McCawley 1970, Emonds 1980), Negation might well have been treated as peripheral.

but some sections of it are non-peripheral. At the same time, other elements may in fact lie beyond the periphery. In both directions, then, Vietnamese data suggests that the border needs to be redrawn.

The Cartographic Turn

Of course, syntactic cartography has hardly ignored CP: the clausal periphery has been grist to the cartographic mill for nearly 25 years (see Rizzi 1997, Cinque 1999, not to mention multiple articles in the sub-series "*The Cartography of Syntactic Structures* (Oxford University Press)." This prodigious output notwithstanding, researchers have focussed almost exclusively on *What (~Which)?, Where?* and *How?* questions; specifically, "*Which* grammatical elements comprise the left periphery?"; "*Where* is a particular element located *vis-à-vis* other CP constituents?"; "*How* did that element come to occupy its position, by movement or base-generation (internal or external *Merge*, in Minimalist terms)?"

Rizzi & Bocci (2017) offer a masterful exposition of this approach, for Italian: the structure in (1a) summarizes their answers to the first two questions (*What?* and *Where?*), the *how?* question being addressed by postulating specific kinds of feature-driven movements, all constrained by the formal principle of *Criterial Freezing* (1b)—itself a variant of other previously entertained "economy" conditions on movement. Just as the greater part of Rizzi & Bocci's (2017) paper is a descriptive commentary on these claims, using Italian data, so most of this paper is an examination of the corresponding Vietnamese facts.[8]

(1) a. [Force [Top* [Int [Top* [Foc [Top*[Mod [Top* [Qemb [Fin [IP …]]]]]]]]]]] (From Rizzi & Bocci 2017)

b. (42) Criterial freezing: A phrase meeting a criterion is frozen in place

Since I will also consider how adverb placement interacts with the functional sequence in (1), it is useful here to introduce Cinque's Adverb Hierarchy, reproduced in (2), which is simultaneously a hierarchy of functional heads and adverbial positions:

(2) $Mood_{speech\ act}$ > $Mood_{evaluative}$ > $Mood_{evidential}$ > $Mod_{epistemic}$ >T (Past) > T (Future) > $Mood_{(ir)realis}$ > Mod_{root} / $Aspect_{habitual}$ /T (Anterior) > $Aspect_{perfect}$ > $Aspect_{progressive}$ / $Aspect_{completive}$ > Voice > V.

Now, this concern with *What? Where?* and *How?* is all to the good, since, at least from an inductive perspective, *What?, Where?* and *How?* questions are logically prior to *Why?*[9] The problem is that cartographers—with the notable exception of Rizzi (2017, also

[8] An additional analytic question is whether external Merger is 'integrated'–extending the clausal spine–or appended, creating an adjunction structure (*cf*. 'Chomsky-adjunction'), within analyses where this distinction is a theoretical option. With respect to child language acquisition (see de Villiers 1991, below), for example, the evidence suggests that *why* is initially adjoined to bare propositions, and only later integrated into the clausal spine.

[9] For *deductive* theories, by contrast, *why* questions are often prior: Chomsky's (1995) discussion of 'virtual conceptual necessity' properties is a partial answer to a *why* question about the shape of grammar, in advance of any consideration of linguistic data. See fn. 8.

2018), to which I'll turn directly—are insufficiently concerned with what might be termed 'mid-level' *Why* questions, even once the map is drawn.[10]

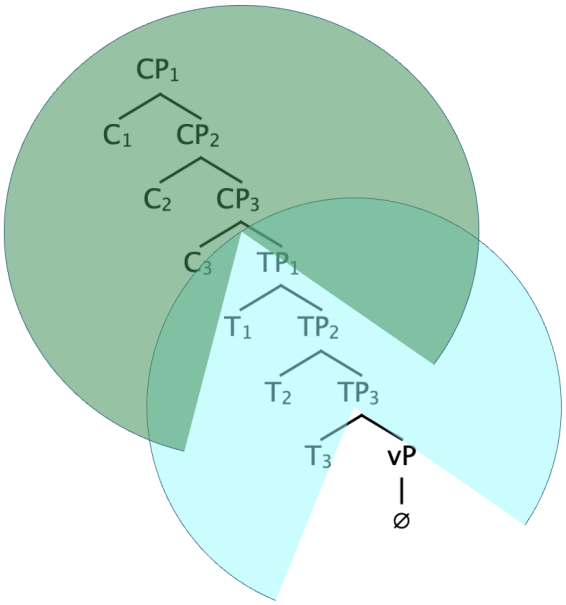
In the case of syntactic cartography, mid-level questions fall into two general categories: *broad-span* questions concerning general clausal architecture, and *narrow-span* questions that should explain cross-linguistic variation within a particular sub-domain. Broad-span *Why*-questions relate to the universal layering of syntactic domains diagrammed in (3): *why*, for example, CP elements are always projected higher than TP {CP_1 CP_2, CP_3…TP_1, TP_2, TP_3, }, rather than in inverse, or intercalated, orders {TP_1 TP_2, TP_3…CP_1, CP_2, CP_3, }, {CP_1 TP_2, CP_3…TP_1, CP_2, TP_3 }; *why* thematic subjects always raise out of the predicate phrase, even in VSO languages (McCloskey 1994, 2017); why T is supervenient on ASP; conversely, why T and Neg vary in their relative order cross linguistically (Ouhalla 1990); *why* almost all lexical categories appear below almost all functional categories (an exception being "Inner Aspect" projections, see above; see also Phan 2013, 2023, for discussion of Inner Aspect in Vietnamese.).[11,12]

[10] On one hand, *high* level, meta-questions ('Beyond explanatory adequacy') are answered by "because Merge" or "because Strong Uniformity" or even "just because innateness…", answers that are unsatisfactory when it comes to capturing micro-parametric, or typological variation. At the other extreme, *low level,* mini questions as to why a particular element is moved to a particular node in a given context are typically answered with reference to some attracting, proprietary formal feature (though *cf*. Roberts' (2019) "^" feature), Rizzi's (2014) *Subject Criterion:* see below.

[11] Until recently, no 'mainstream', Merge-based approach attempted to explain why a language could not start the derivation with T or C, rather than merging these functional categories later, as invariably happens. To be sure, there are consistently left-branching languages where the CP-domain is realized on the right periphery, but there are, to my knowledge, none where lexical categories dominate the corresponding non-propositional functional nodes. In principle, this should be possible: it merely involves reversal of items bearing probe and goal features.

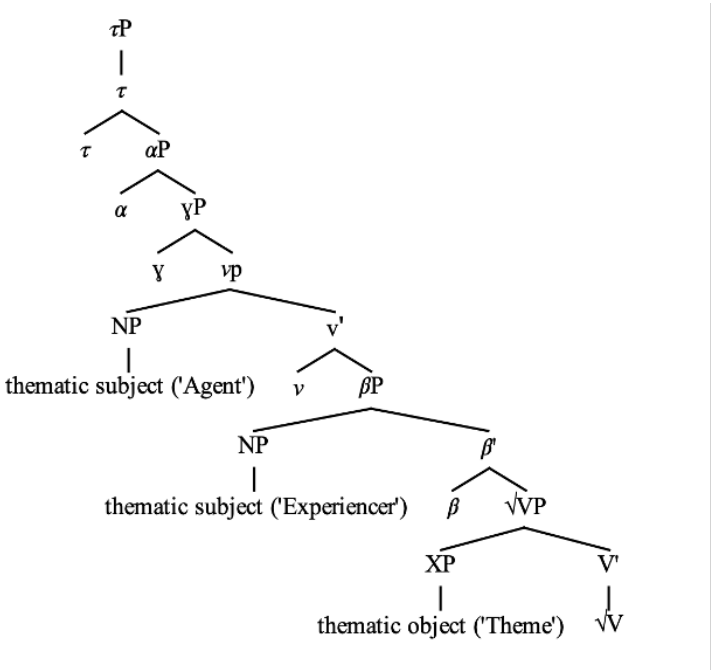
[12] Additional questions arise concerning the ordering of adjunct modifiers both relative to lexical heads—where they typically display similar cross-categorial harmony effects to those found with arguments, but where head-complement parameter explanations cannot apply (Travis 1984, *cf*. Hawkins 2001)—also relative to one another, where semantic factors evidently play a role (Cinque 1994, 1999; Sproat & Shih 1991; Teodorescu 20016; Endo, Shinuya et al. 2016).

(3)



More specific, 'narrow span' questions include the following: given a set of morphemes in a particular language appearing at the left edge of the clause, *why* are certain elements supervenient on other elements? In the figure in (4), for example, *why* must **τ** c-command **α**, and **α**, **γ** within a given subdomain?

(4)



On such questions cartographers are largely silent.[13] To my knowledge, the first discussion of the aetiology of CP at this level of explanation is found in the preliminary speculations in Rizzi (2017), who writes:[14]

> "Why is it that we typically find certain properties of ordering and cooccurrence restrictions, rather than others?...[...] Two broadly defined candidates come to mind:[15]
>
> 1. Certain properties could derive from *requirements of the interface systems*. For instance, it could be that functional head B may necessarily occur under functional head A (thus giving the linear order AB in head-initial languages and BA in head-final languages) because the opposite hierarchical order would yield a structure not properly interpretable...[...]...A *special* [my emphasis: NGD] case of the impact of interface requirements may be the ordering properties that follow from selectional requirements, e.g., the fact that the Force head in embedded clauses must be high enough to be accessible to higher selectors, which want to know if their complement is a declarative or a question, for instance (Rizzi 1997);
> 2. When the functional heads occurring in specific orders trigger movement, *the ordering may be a consequence of locality requirements*. For instance, Abels (2012) has argued that almost all the ordering effects observed in the Italian left periphery may follow from the theory of locality based on a version of featural Relativized Minimality, along the lines developed in Starke (2001), Rizzi (2004): if A is a stronger island-creating element than B, then B will not be extractable from the domain of A, neither long-distance, nor locally..."

Setting aside the issue of selection—which by definition applies only to a subset of subordinate clauses, and so is exactly a "special" (i.e., non-typical) case—neither of these responses is satisfactory. In the absence of a deeper explanation as to why "functional head B *necessarily* occurs under functional head A, alternatively, "why A is a *stronger*

---

[13] It is telling, I think, that there are no departments of cartography in research universities: geographers, historians, archaeologists, economists, social scientists of all types rely on cartography furnished by government workers (UK Ordnance Survey, US Geological Survey, for instance). Cartography is only ever the beginning of the story: no matter how accurate or detailed it may be, a good map offers no deeper explanation in the absence of a prior physical, social, or economic theory.

[14] Increasingly, Rizzi has become concerned with *why* questions: indeed, he devotes a full section of his 2017 paper to the issue. However, the proposal given there is explicitly circular: "A functional sequence may be taken as an *explanans*;...[...] Reciprocally, a functional sequences should be looked at as an *explanandum*, a complex set of properties which in and of itself is in need of a further and deeper explanation." In his discussion of functional sequences as *explananda*, Rizzi then reiterates the two kinds of principles listed above. Such argumentation may be satisfactory to some readers. This is not to say that cartography has no explanatory value. On the contrary, it has tremendous value as a heuristic, within a deductive theory of grammar. However, this does not mean that "the functional sequence" is an *explanans* since, absent a deeper explanation, we are no closer to removal of "puzzlement" (the criterion of explanation proposed by Emmon Bach, 1974:154, cited in Greenberg 1979: 279).

[15] Rizzi (2017) implies that Cinque & Rizzi (2010) is the original source of these speculations ("As pointed out in Cinque & Rizzi (2010)..."): however, from my reading, that position paper is exclusively concerned with *where*, *what* and *how* questions: the question *why* does not feature in that article at any point.

island-creating element than B [my emphases: NGD]", purely formal explanations do little more than to restate the distributions they are supposed to explain (*cf*. appeal to 'strong' *vs*. 'weak' features in earlier treatments of verb movement, see e.g., Chomsky (1989, 1992)).[16]

The alternative, functionalist proposal I have been promoting recently[17] is cast in terms of four principles, two of which—*Supervenience* and *'1-Arg'*—offer a partial answer to the mid-level questions introduced above. I will return to this proposal in the concluding sections, once we have addressed the core data question: "*Wherefore is* why *different?".* And before that, we have some mapping to do.[18]

Vietnamese clause structure

As may be familiar, Vietnamese is distinguished from most other MSEA languages in having a rich inventory of pre-subject constituents, whose relative distributions provide evidence of an articulated left periphery. While this super-structure is similar in many ways to that proposed by Rizzi & Bocci (2017), based on Italian, the Vietnamese facts presented below suggest significant points of divergence.

> Notionally, we may distinguish five classes of pre-subject constituents, as follows:
>
> (i) clause initial conjunctions—including various kinds of subordinating conjunction {complementizer, relativizers}, topic markers, and linking elements. These elements have a fixed distribution relative to one another, and are never found lower in the clausal hierarchy (to the right of the subject): **{rằng, liệu, (mà),[19] (là)}, {thì, (là)}**
> (ii), left peripheral phrasal constituents that are amenable either to a movement or to a co-indexation analysis (Move or Merge): **{topicalized XPs, NPs heading relative clauses** (see fn.)}
> (iii) speaker-oriented adverbial phrases (e.g., ***quả thật***, 'indeed')
> (iv) special phrasal constituents, which in all probability have been moved to a pre-subject position, given the distribution of the same constituents in the general case; {**universally quantified XPs**}
> (v) the *wh*-phrase *why* (***tại sao***), which—uniquely in this *wh*-in situ language—only ever appears to the left of the subject.

I consider each these classes in turn, subsuming the second category within the first: that is to say, I will assume the overt phrasal constituents in (ii) occupy the 'specifier'

---

[16] There is, of course, a way to make selection 'non-special', namely, by assuming that even root clauses contain illocutionary features within an expanded left periphery.. This move, which could be correct, does no more to solve the broader *why* question; since there is no reason in principle why Force features could not be merged early (in an autonomous syntactic theory).

[17] See Duffield (2021, 2022), and as developed further below.

[18] Related to these questions, it should be asked whether conventional labels used to categorize elements SAE languages {conjunction, complementizer, relative pronoun, topicalizer…} are useful or distracting when it comes to describing pre-subject functional categories (see, for example, {Clark, Marybeth, 1992}?

[19] Time and space constraints preclude consideration of *mà* in this draft.

position of their corresponding licensing heads (XP-**thì**, XP-**mà**, in the case of topicalization and relativization, respectively.)[20]

Clause-Initial (Pre-Peripheral?) Conjunctions

Although our core focus is with the last of these classes, it is best to start at the top, with the head elements that provide most direct evidence of a functional sequence on the left periphery. The examples in (5) illustrate the canonical distributions of all the Class (i) heads under consideration here:

(5) a. Tôi nói **rằng** [tôi **là** [cán bộ ngoại giao [**mà** cần liên hệ với sứ quán.]]]
PRN say ?? I COP staff foreign.affairs REL need contact with embassy
'I said that I was a diplomatic staff member who needed to contact the embassy.'

b. Cô gái hỏi **liệu** [cô có thể đi đến bữa tiệc được không.]
PRN girl ask ?? PRN Q possibility go arrive party CAN NEG
'The girl asked if she could go to the party.'

c. [Quyển sách [**mà** [anh thích nhất]] **thì** bán chạy.
CLF book REL PRN like best TOP sell run
'The book that you like most, (it) is selling well.'

Conjunctions I: {rằng, liệu, là}

Let us take {**rằng, liệu**} first. As these occur as the most peripheral elements in any embedded sequence and signal the illocutionary force of the following clause, a reasonable assumption is that they are *Force* heads, contrasting only in selectional features (+IND, +INT): that is, it might be thought that *rằng* introduces declarative clauses, whereas *liệu* introduces interrogatives. However, though it is initially plausible to associate both elements with *Force*, the examples in (6) and (7) demonstrate that *liệu* and *rằng* co-occur. More surprising perhaps, both word orders are observed— whereas *liệu* precedes *rằng* in (6), the opposite order is found in (7)—even while the unacceptability of the (b) and (d) examples in each set shows that this is not a matter of free variation.

(6) a. Đi tìm lời đáp: **Liệu rằng** tia UV có làm filler biến dạng?
go find answer: ray UV ASR make filler deform
'Finding the answer: do UV rays cause filler deformation?'[21]

b. Đi tìm lời đáp: *__Rằng liệu__ tia UV có làm filler biến dạng?
go find answer: ray UV ASR make filler deform
(as a)

[20] In both cases, I will adopt a base-generation (with coindexation) analysis here, though nothing hangs on this: in this paper, I am simply concerned to establish surface positions within the functional sequence.

[21] *Elle* magazine (VN edition, 17/02/2024).

c. **Liệu rằng** khi mất đi em có còn hối tiếc? [song lyric]
Q COMP when lose PRN ASR still regret?
'Will I regret it when I lose you?'

d. *__Rằng liệu__ khi mất đi em có còn hối tiếc?
COMP Q when lose PRN ASR still regret?
(as c)

(7) a. John hỏi **rằng liệu** [tôi có muốn hẹn hò với anh ấy không]
John ask I Q want go.out with PRN.DEM NEG
'John asked if I wanted to go on a date with him.'

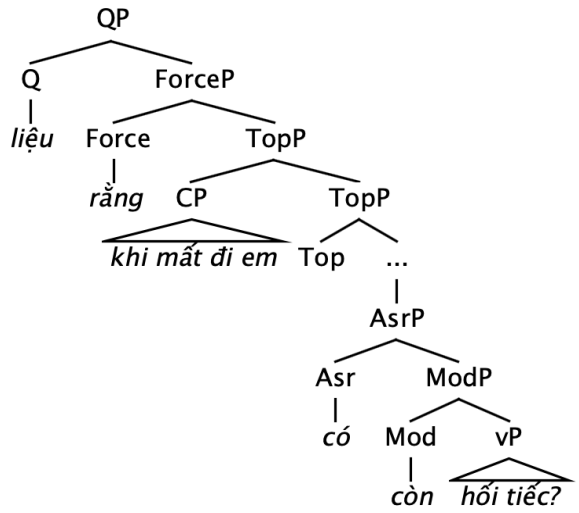
b. *John hỏi **liệu rằng** [tôi có muốn hẹn hò với anh ấy không]
John ask I Q want go.out with PRN.DEM NEG
(as a)

c. John đã hỏi mẹ **rằng liệu** [mẹ có thể đón cậu bé không]
John PAST ask mother mother can pick up child NEG-Q
'John asked his mother if she could pick him up.'

d. *John đã hỏi mẹ **liệu rằng** mẹ có thể đón cậu bé không.
John PAST ask mother mother can pick up child NEG-Q
(as c)

The clue to this puzzle lies in the fact that both *liệu* and *rằng* also serve as matrix predicates in certain contexts. In the examples in (6), for example, *liệu* works in this fashion: in fact, there are no cases of {*liệu…rằng*} sequences where *liệu* itself is preceded by another matrix predicate. Considering this, it is plausible to analyse *liệu* in (6c) as extra-clausal: this is diagrammed in (8) (*cf.* French *est-ce (que)*):

(8)



As for *rằng*, etymologically this element is derived from a matrix verb, synonymous with *nói* ('speak, say') The diachronic source of *rằng*, as a matrix predicate, is illustrated by the proverb and folktale examples in (9) below:

(9) a. 'Chẳng nói, chẳng **rằng**!'
NEG say, NEG speak
'Say nothing!

b.. '…Phú ông xin đổi ba bò, chín trâu…
rich man ask exchange three cows nine buffalo
'The rich man asked to exchange three cows and nine buffaloes…'

…Bờm **rằng**: Bờm chẳng lấy trâu…
Bom say: Bom NEG take buffalo
'..(and) Bom said: "I (Bom) will not take the buffaloes."'

Given its verbal origins, it is reasonable to analyse contemporary examples where *nói rằng* introduces direct speech, and is immediately followed by a colon or comma—examples such as those in (10)—in terms of a similar matrix analysis.

(10) a. Murakami Haruki **nói rằng**: "Tôi cho bạn biết tôi thích bạn …."[22]
Murakami Haruku say : "I let friend know I like friend…
'Haruki Murakami said: "I'm telling you that I like you…"'

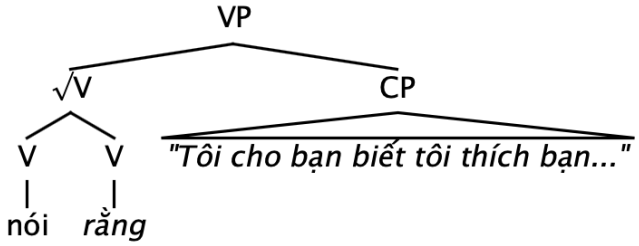
b. Tôi thấy lỗi **nói rằng**: "Một số quy tắc không được áp dụng nhất"
I see error say 1 number rule NEG CAN apply correct most
'I see an error saying "A number of rules are not most applicable"'[23]

Two variant possibilities present themselves: nói *rằng* might be analysed as a verbal doublet, diagrammed in (11a); alternatively, *nói* might function intransitively, with *rằng* heading a VP-adjunct, as in (11b). This would then be akin to the most likely analysis of the Early Modern English {…*spoke*, *saying*…} constructions in (12):

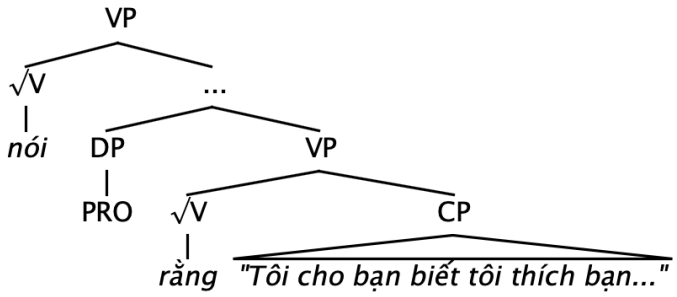
[22] The full citation from merrygoround.divongquanh 09/30/2024 reads:
"Murakami Haruki nói rằng: "Tôi cho bạn biết tôi thích bạn không hẳn là muốn ở cạnh bạn, chỉ là mong rằng sau này khi bạn gặp thăng trầm của cuộc đời thì đừng nản chí, bởi ít nhất đã có một người từng bị sức hút của bạn hấp dẫn. Trước đây như thế, về sau cũng vậy." ("Murakami Haruki said: "I tell you that I like you, not necessarily because I want to be with you, but because I hope that when you encounter ups and downs in life, you will not be discouraged, because at least one person has been attracted to you. It was like that before, and it will be like that in the future.")

[23] https://support.microsoft.com/…

(11) a.



b.



(12) a. "…and he **spoke, saying**: These are the horns which have scattered Juda every man apart, and none of them lifted up his head …(Zechariah 1:21 KJV)"

b. But the king **spoke, saying** to Daniel, "Your God, whom you serve continually, He will deliver you…(Daniel 6:16)"

Yet whilst these extra-clausal analyses may deal with {*liệu…rằng*} order, as well as with some instances of 'reporting' *rằng*, it is nevertheless clear that in the general case this element lies *within* the periphery: that is to say, its etymology notwithstanding, *rằng* has been grammaticized. This is strongly implied by the fact that it appears following 'non-reporting' predicates—i.e., those that cannot take quoted speech as a complement. This set includes the predicates *tin* ('believe'), *nhận ra* ('realize'), *bực tức* ('resent'), and *tiếc* ('regret'), discussed in the section on embedded topicalization below.[24]

In addition, *rằng* appears in concessive clauses as part of the complex conjunction *tuy rằng* (= 'although', *cf*. French *bien que*), where, once again, no verbal report is involved.

(13) a. Nó không nói gì, [**tuy rằng** [nó biết rất rõ]].
PRN NEG say what, [although PRN know very clearly
'He didn't say anything, even though he clearly knows (the answer).'

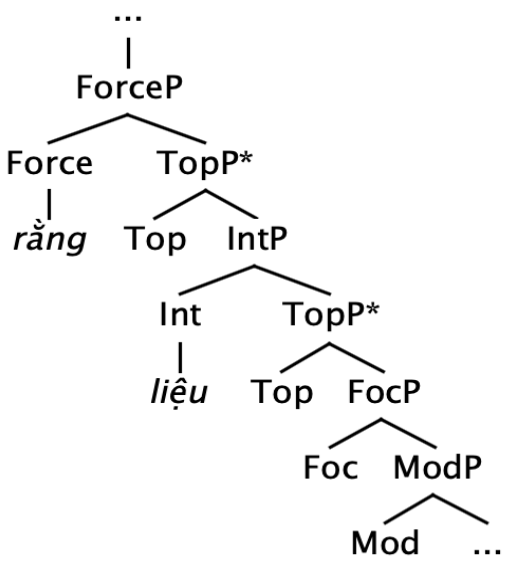
b. [**Tuy rằng** [nhớ tên]] [nhưng lại không nhớ ra gương mặt]
Although remember name but again NEG recall out face
'I remember the name, but not the face.'

So, if *rằng* is located within the complement clause in default contexts, lexicalising Force, then in examples such as those in (7) above, where *rằng* precedes *liệu* in non-reporting contexts, it follows that *liệu* must occupy a lower functional position within CP.

[24] See the examples in (35)-(36) below, discussed in relation to embedded topicalization (ET).

Within Rizzi & Bocci's cartography, **Int**(errogative) presents itself as a prime candidate: *liệu* would then be equivalent to Italian *se* (*cf*. Rizzi & Bocci, *ibid*.). This analytic possibility is diagrammed in (14):

(14) Positioning *rằng* and *liệu* (first pass)[25]



It is here that matters become interesting. For just as *rằng* and *liệu* are distinguished, so too are *rằng* and *là*,whenever the latter element behaves as a conjunction. In (5a) above we saw that *là* functions canonically as a copula, linking subjects and nominal predicates; see also the examples (15). Crucially, in its copular function *là* may not appear in construction with other kinds of non-verbal predicate, neither with adjectival (individual-level or stage-level) predicates (16), nor with prepositional predicates (17); examples taken from Phan & Duffield (2023).[26]

(15) a. John **là** người đàn ông đứng ở đằng kia.
John COP person gentleman stand LOC.COP location there
'John is the person standing over there.'

b. Người đàn ông đứng ở đằng kia **là** John.
person gentleman stand LOC location COP there.
'The person standing over there is John.'

(16) a. Anh ấy **(*là)** cao/thông minh/tốt. *[AP]
PRN.DEM COP tall/intelligent/nice
'He is tall/intelligent/nice.'

[25] We shall see directly that this is untenable, at least because Vietnamese does not allow the recursive higher topic position between Force and Int.

[26] As we shall see directly, *là* can appear adjacent to adjectival predicates, just not in its copular function.

b. Cô ấy **(*là)** bận/rất vội vàng/rất vui.
PRN.DEM COP busy/very hurry/very happy
'She is busy/in a great hurry/very happy.'

(17) a. Anh ấy ở/**(*là)** trên tàu. *[PP]
PRN.DEM LOC.COP/COP on train
'He is on the train.'

b. Tiền hoàn thuế của tôi ở/(*là) đâu?
money complete tax POSS 1.SG LOC.COP/COP where
'Where is my tax refund?'

c. Nhìn xem! Anh ấy ở/(*là) kia kìa.
look see PRN.DEM LOC.COP/COP DEM_3.DEM_3
'Look! He's over there.'

Rằng can never replace *là* in its copular function. However, the opposite substitution does seem to be possible, inasmuch as *là* alternates with *rằng* as a subordinating conjunction in a variety of contexts. The examples in (18) show that *là* readily introduces complement clauses, especially in spoken registers.

(18) a. Yui đã nói **là** đi đến nhà của bạn.[27]
Yui PAST say CONJ go to house POSS friend
'Yui said (she was) going to her friend's house.'

b. Tôi biết **là** chị đang yêu mọt ngừoi.
I know CONJ she PROG love 1 person
'I know that she is in love with someone.'

c. Lan nói **là** (chị ấy) thích học tiếng Anh.
Lan say CONJ PRN.DEM like study lge. English
'Lan said she liked learning English.'

However, just as was observed with *liệu*, no true substitution is involved here: *là* once again occupies a functional position lower than *rằng*. This is evidenced by the examples in (19) below, which show that *rằng* and *là* may co-occur, abeit only in one order, {*rằng…là*}. (Example (19c) shows that {*rằng..là*} collocations contrast d'irectly with {**mà..là*)} sequences: *là* cannot occur within relative clauses, nor may it substitute for *mà* as a relativizing head.)

[27] https://vn.wa-tera.com/to-iimasu/. In non-reporting contexts, *nói là* typically has a less quotative meaning than *nói rằng*: that is to say, *nói là* is 'say' in the sense of "mean, imply, suggest" compare English: "What does it say when he turns up especially early?" By contrast, *rằng* is almost always associated with mental states and embedded propositions.

(19) a. Phải nói [(***là**) [rằng [(**là**) thế hệ trẻ của chúng.ta rất tài năng.]]][28]
modal say ?? COMP ?? generation young of PRN very talented
'(I) have to say that our young generation is very talented.'

b. Bạn có thể nói [(***là**) [rằng [**là** mình ổn, nhưng…]]][29]
friend possible say ?? COMP ?? self fine, but…
'You can say that you're fine, but…'

c. Quyển sách [(***là**) [mà [**(*là**) anh thích nhất]]] **thì** bán chạy.
CLF book ?? REL ?? PRN like best TOP sell run
'The book that you like most, (it's) selling well.'

A further point to observe is whilst *là* may introduce interrogative clauses, it differs from *liệu* in having a "quasi-quotative", rather than subordinating, function: that is to say, the question that follows *là* is interpreted as 'direct-reported', not indirect.[30] In this respect, *là* functions similarly to colloquial English *like* in (20a), where inversion (SAI) signals the interrogative force of the following clause, whilst at the same time sequence of tense rules and deictic shift indicate that this is not direct speech (*cf*. *20b *vs*. 20c):

(20) a. 'She asked me, like, What was I doing there?' [direct reported]
b. %She asked me what was I doing. [*in standard varieties] [indirect]
c. She asked me: "What are you doing here?" [direct]

In short, *là* has a semantic status intermediate between *rằng* and *liệu*: whereas *rằng* is typically associated with propositional assertions, and *liệu* with alternative questions, *là* is apparently neutral between the two types of complement. This is suggested by the minimal contrasts in (21), involving *không biết* ('not know'), which, like English *know*, can take either a declarative or interrogative complement: the corresponding examples in (22) suggest that *là* is equipotent:

(21) a. Người mẹ không biết ***rằng**/liệu bọn trẻ đang ngủ hay thức.
mother NEG know COMP/INT children PROG sleep or awake
'The mother doesn't know whether the kids are sleeping or awake.'
(= the mother is unsure about whether her kids are asleep.)

b. Người mẹ không biết [rằng/***liệu** [bọn trẻ có đang ngủ.]]
mother NEG know COMP/INT children ASR PROG sleep
'The mother doesn't know that the kids are in fact sleeping.'
(= the kids are in fact asleep: the mother doesn't know that fact)

[28] (20a) from Duffield (2013).

[29] "Bạn có thể nói rằng là mình ổn, nhưng ánh mắt bạn lại không biết nói dối.." ('You can say that you are fine, but your eyes don't know how to lie').

[30] Another clear difference between *liệu* and conjunction *là* is that *liệu* can appear in utterance initial position, as in (7) above: this is impossible for *là*, which only appears intra-clausally.

c. Cô ấy không biết [***rằng**/liệu [tôi **có** đang vui vẻ tại bữa tiệc **không**]]
PRN.DEM NEG know COMP/INT I Q PROG happy at party NEG^{Q}
'She doesn't know whether I am having fun at the party'
(= it is unclear to her whether or not I am having fun.)

d. Cô ấy không biết [rằng/***liệu** [tôi đang vui vẻ tại bữa tiệc]]
PRN.DEM NEG know COMP/INT I PROG happy at party
'She doesn't know that I am having fun at the party'
(= I am having fun, indeed, but she doesn't know that fact)

(22) a. Người mẹ không biết **là** bọn trẻ đang ngủ hay thức.
mother NEG know ?? children PROG sleep or awake
'The mother doesn't know whether the kids are sleeping or awake.'
(= the mother is unsure about whether her kids are asleep.)

b. Người mẹ không biết **là** bọn trẻ có đang ngủ.]
mother NEG know ?? children ASR PROG sleep
'The mother doesn't know that the kids are sleeping.'
(i.e., the kids are in fact asleep: the mother doesn't know that fact)

…

c. Cô ấy không biết **là** tôi đang vui vẻ tại bữa tiệc hay không.[31]
PRN NEG know ?? I PROG happy at party or NEG
'She doesn't know whether I am having fun at the party or not'
(= it is unclear to her whether or not I am having fun.)

d. Cô ấy không biết là tôi đang vui vẻ tại bữa tiệc.
PRN.DEM NEG know ?? I PROG happy at party
'She doesn't know that I am having fun at the party'
(I am having fun, indeed, but she doesn't know that fact.)

Moreover, it turns out that conjunction *là* has a larger set of distributional privileges than either *rằng* or *liệu*. Another significant difference is that *là* introduces the complement of certain epistemic predicates such as those in (23), contexts where both *rằng* and *liệu* are excluded. It is likely no coincidence that many of these are raising predicates: the raised counterparts are illustrated in (24). This is exactly what one would expect if *là* occupies a lower node than *Force*, which seems to be the projection that blocks A-movement in English.

[31] Notice that the contrast between (21c) and (22c) is not quite minimal: whereas *liệu* introduces a standard Y-N question having the same form as a matrix question *[tôi **có** đang vui vẻ tại bữa tiệc **không**]* 'Am I having fun at the party?', with interrogative embraciation, *là* prefers to express the same idea using alternative statements, i.e. without interrogative ***có***, followed by ***hay không***. This is consistent with the idea that *là*, like English *like*, lies intermediate between direct and indirect subordination—"direct-reported" see (20a) above. This clearly requires closer investigation.

(23) a. ?Có **nên** *rằng/là đàn ông **là** trụ cột gia đình?[32]
ASR MOD men COP pillar family
'Should the man be the family breadwinner?'

b. Có **phải** *rằng/là cái chết **là** chấm hết?
ASR right CLF deathq COP end final
'Is Death really the end?'

c. **Chắc** *rằng/là anh ấy bị tắc đường nên đến muộn.
Likely PRN.DEM AUX delay road MOD come late
'Maybe he was stuck in traffic, so he was late.'

d. Nhà hàng kia đông quá! **Chắc** *rằng/là đồ ăn ngon lắm!
restaurant DEM_3 full very likely food l delicious very
'That restaurant is so crowded! The food must be very good.'

(24) a. Đàn ông có **nên** là ~~đàn ông là~~ trụ cột gia đình?[33]
man ASR MOD COP pillar family
'Should men be the family breadwinners?'

b. Cái chết có **phải** là ~~cái chết là~~ chấm hết?
CLF die ASR MOD COP end final
'Is death the end?'

c. Anh ấy **chắc** là bị tắc đường nên đến muộn.
PRN.DEM likely CONJ AUX delay road MOD come late
'Maybe he was stuck in traffic, so he was late.'

d. Nhà hàng kia đông quá! Đồ ăn **chắc** là ngon lắm!
restaurant DEM_3 full very food likely CONJ delicious very
'That restaurant is so crowded! The food must be very good.'

The VN conjunction *là* differs in subtle but crucial respects from its putative Italian counterpart *di*. Comparing the Italian examples in (25), involving *sembrare* ('seem'), it can be seen that whereas raising is allowed out of bare non-finite complements (25a), *di* blocks A-extraction (25b) just as extraction is blocked from the corresponding finite

[32] Notice that in the unraised cases, the copular form of *là* is required before nominal predicates (23a), (23b). This copula must be deleted in the raised alternants (24a), (24b). Though this might be explained as a case of haplology, there seems to be something more going on, since other multifunctional items—aspectual *vs*. preterite *đã*, for example—are also only able to appear once per clause. Also, even when a constituent intervenes, as in (23a)/(23b), there is a preference to drop one of the two *làs*; in such cases, though, it is the conjunction, rather than the copula, that must be dropped: *Có phải (?là) cái chết là chấm hết?*

[33] (19c) https://tatoeba-org.translate.goog/en/sentences/show/9013946?_x_tr_sl=vi&_x_tr_tl=en&_x_tr_hl=en&_x_tr_pto=sc (20a) https://spiderum.com/bai-dang /Dan-ong-co-nen-la-tru-cot-gia-dinh-hteaGrzfEuOL; (20d) https://hanoilanguage.vn/how-to-use-hinh-nhu-chac-la-seem-in-vietnamese/

complements introduced by *che* (compare 25c vs. 25d). A further difference is that raising is obligatory in Italian (and English) non-finite structures, as shown by the unacceptability of (25e), presumably for Case reasons. Notice that *di* only appears with the Control use (of sembrare (25f).[34]

(25) a. ma (loro) sembrano [usare la violenza per altre cose].
but they seem-3.PL use-INF DET violence for other things
'...but they seem to use violence for other things.'

b. *ma (loro) sembrano [**di** usare la violenza per altre cose].
but they seem-3.PL di use-INF DET violence for other things
(as a)

c. ma sembra [**che** (loro) usino la violenza per altre cose].
but seem-3.SG that they use-subjnc-3.PL DET violence for other things
'...but it seems that they use violence for other things.'

d. *ma (loro) sembrano [**che** usino la violenza per altre cose].
but seem-3.PL that use.subjn.3.PL det violence for other things
'...but they seem that (they) use violence for other things.'

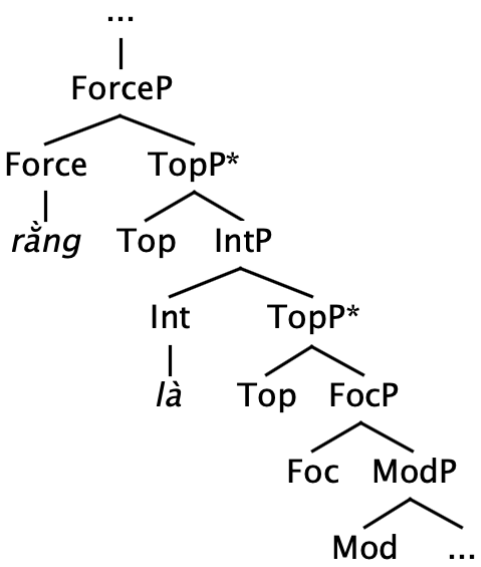
e. *ma sembra [(*loro) **di** (*loro) usare la violenza per altre cose].
but seem-3.sg them di them use-INF DET violence for other things
'[intended]...but they seem to use violence for other things.'

f. Mi sembra di impazzire.
me.DAT. seem-3SG di go.crazy
'It seems like I am going crazy.'

By contrast, as we have already seen, conjunction *là* is compatible with either finite or non-finite complements. This implies that, unlike *di*, which lexicalizes **Fin** (in Rizzi & Bocci's cartography, *là* occupies an intermediate position between Force and **Fin**. Once again, **Int** would seem to be the most promising candidate:

[34] I am extremely grateful to Guglielmo Cinque for correcting my errors concerning *sembrare* in an earlier draft. Ironically, the Vietnamese translation equivalent of Italian *sembrare*~seem—*hình như*—does not allow raising:

(26)



An additional reason not to identify *là* with di/**Fin** is that—as we shall see in the next section—*là* invariably precedes Topicalized constituents in Vietnamese, unlike Italian *di*: which, as Rizzi & Bocci point out, follows the topic constituent; this is illustrated in (27) (Rizzi & Bocci's examples (1) and (2)).

(27) a. Ho deciso **che**, la macchina, la comprerò quest'anno.
'I decided that, the car, I will buy it this year.'

b. Ho deciso, la macchina, **di** comprarla quest'anno.
'I decided, the car, of to buy it this year.'

Since the order in (27b) FORCE-TOP-*là* in Vietnamese is unacceptable, and since—as will be shown—the Vietnamese left periphery only seems to admit the lower Top position, it follows that *là* must occupy a relatively high position, higher than **Fin.** Once again, then, **Int** presents itself.

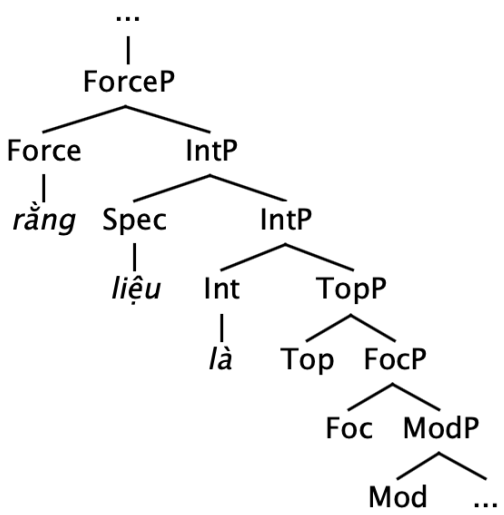
However, it should be clear that this yields a cartographic clash: the tree in (14) that derives {*rằng…liệu*} sequences is the same as the one that describes {*rằng…là*} order in (26). This leads to the prediction that *liệu* and *là* should be in complementary distribution, whenever they both function as conjunctions (i.e., following *rằng*). Unfortunately, this prediction is immediately disconfirmed by the examples in (28) below: these demonstrate that the sequence {*rằng…liệu..là*} is perfectly acceptable (in any context where {*rằng…liệu*} is also good).

(28) a. John hỏi rằng liệu (**là**) [tôi có muốn hẹn hò với anh ấy không]
John ask COMP INT ?? I Q want go.out with PRN.DEM NEG^{Q}
'John asked if I wanted to go on a date with him.'

b. John đã hỏi mẹ rằng liệu (**là**) mẹ có thể đón cậu bé không]
John PAST ask mother COMP INT ?? mother Q can pick up child NEG[Q]
'John asked his mother if she could pick him up.'[35]

Dismissing the *ad hoc* solution of an iterative **Int** node, which would in any case fail to explain the impossibility of *[là..liệu] order, some reanalysis is necessary. The simplest solution is to treat *liệu* as a specifier element in **Int**, with **là** as its head, as in (29), so preserving the Rizzi/Bocci hierarchy of C-related heads.[36]

(29)



Naturally, other analyses are possible, given the unspecified semantics of the criterial head **Int**, and the lexically underspecified nature of Vietnamese *là*. Considering topicalization more closely allows greater precision, though no certainty: the proliferation of Top positions in Rizzi & Bocci's structure is unhelpful in this endeavour, as is the surprising claim that in certain dialects Italian *che* may lexicalise any one of several heads in a given structure:

> 'In fact,...[...]...elements like *che* are typically versatile, and can occur in different positions in the clausal spine: in the dialects under consideration, they occur in the highest position as declarative force markers, and also in a lower position, lower than the *wh*-element in indirect questions...(Rizzi & Bocci, fn 32).

At all events, the data allow us to distinguish "high" from "mid-high" in the left periphery of Vietnamese clauses.

[35] Matters are further complicated by the fact that {*rằng…là..liệu*} sequences are also acceptable if there is a prosodic break between {*rằng…là*} and {*liệu*}: I will assume that this anomaly can be explained in terms of performance: *cf*. English "that..that" restarts.

[36] The tree in (29) anticipates the discussion of embedded topiclization in the next section, where it will be shown that Vietnamese differs from Italian in not allowing topicalization above **Int.**

Having now devoted more space than the average reader may regard as tolerable to the highest levels of the left periphery, we can inch our way further down to consider the "mid-high" region of CP.

Conjunctions II: {thì, (mà), là}

The "topic" marker *thì*, along with the distinguished phrase that typically precedes it, has been an object of theoretical interest for several decades—likely much longer within traditional scholarship. Notable studies in English include Thompson (1965/1987), Cao (1991/2004), Clark (1992), and Tran (2016, 2024). Nguyen M. C. (2021) provides an excellent summary; she also contributes some useful new data, particularly in respect of clause-initial *thì*.[37] The main reason that *thì* has attracted such interest is its indeterminate nature—construed more positively, its multifunctionality: as the examples below indicate, *thì* serves many different functions, including as an "aboutness topic" marker (30), a contrastive topic marker (31), and a clausal conjunction (32); in addition, as both Tran (2016) and Nguyen (2021) point out, *thì* frequently appears utterance-initially as a discourse linker (33)—for Nguyen, this is its primary function.

(30) a. Cô ấy **thì** gặp anh ấy. [subject topic]
PRN.DEM TOP meet PRN.DEM
'Speaking of her, she meets him.'

b. Anh ấy **thì** cô ấy gặp. [object topic]
PRN.DEM TOP PRN.DEM meet
'Him, she will meet.'

c. Hồi trước **thì** cô ấy gặp anh ấy. [adjunct topic]
time before TOP PRN.DEM meet PRN.DEM
'In the old days, she would meet him.'

(31) a. Ở miền Bắc có bốn mùa nhưng … [contrast topic]
LOC north EXIST four season but…
…ở miền Nam **thì** chỉ có hai mùa.
LOC South TOP only have two season
'In the North, there are four seasons, but in the South, only two.'

[37] I am grateful to Trang Phan for drawing my attention to Nguyen's dissertation (2021). Using the pragmatic-functional framework proposed in Fraser (2009), Nguyen treats all these instances as discourse-markers of various kinds (Elaborative, Logical, Contrastive DMs). Her conclusions are broadly in line with Clark (1992):

> "The claim is made here that *thì* has a general discourse-related topicalizing function, explicitly marking background for the main proposition which the speaker wishes to communicate, and that it is this explicit marking allowing for 'immediate' communication that makes this conjunction so popular. A subsidiary claim is that such a function is an inchoative one, that conjunctions such as *thì* introduce inchoative predications: Given *X*, then predicate *Y* as coming about (Clark 1992: abstract)."

It is not clear how either of these approaches should capture the constraints on clause-internal interactions detailed below.

b. Mày *thì* gầy, nó **thì** béo. [Clark 1992: [5]]
you TOP slim, PRN TOP fat
'You are slim, and he is fat.'

(32) a. [Anh mà đến] **thì** [chị ấy rất vui lòng.] [clausal linker]
PRN TOP come CONJ PRN.DEM very happy
'She'll be very happy, (if) you come.'

b. [Anh không thấy ai] **thì** [không cần ở lại một mình.]
PRN NEG see wh CONJ NEG need stay oneself
'(If) you don't see anyone, you don't have to stay there by yourself.'

c. (Nếu) trời sập **thì** cô ấy gặp anh ấy.
if sky fall CONJ PRN.DEM meet PRN.DEM
'If the sky falls, she will meet him.'

(33) A: [Gia đình ấy có nhà to] [discourse linker]
Family DEM2 have house big
'That family has a big house.'

B: **Thì** [cô ghen tị à?].
PTL 2 jealous PTL
'Are you jealous of them?'

For our purposes, the most pressing question about *thì*, regardless of label, is its distribution and interaction with respect to other left-peripheral elements, particularly in embedded contexts.

The first point to establish here is whether Vietnamese allows embedded topicalization (ET) in the first place; after that, we can determine to which topicalization class it belongs, within the typology of Miyagawa (2017), to be discussed directly.

It has long been known, since Hooper & Thompson (1973)'s response to Emonds (1969)'s original proposal, that English ET is restricted by the semantic class of the matrix predicate. In Hooper & Thompson's taxonomy, reproduced in (34), ET is excluded from C & D class predicates. This illustrated by the contrast between the examples in (35) and those in (36):

(34)

Hooper and Thompson (1973: 473–474)

Non-factive			Factive	
A	B	C	D	E
say	suppose	be (un)likely	resent	realize
report	believe	be (im)possible	regret	learn
exclaim	think	deny	be surprised	know
etc.	etc.	etc.	etc.	etc.

(35) a. I exclaimed that this book, I will never read. (Class A)
b. I think that this book, he read thoroughly. (Class B)
c. I found out that this book, no one is willing to read for the test (Class E)

(36) a. *It's likely that this book, everyone will read for the test. (Class C)
b. *He was surprised that this book, I had not read. (Class D)

In the same paper, Hooper & Thompson (1973: 485) observe that ET is disallowed in non-finite clauses, irrespective of semantic class; compare the examples in (37). They propose that this is because infinitival complements are reduced clauses, lacking a Topic position, something that will become crucial shortly.

(37) a. My friends said [$_{TOP}$ the more liberal candidates], they had always supported.]
b. My friends tend to support [the more liberal candidates]. [Miyagawa 2017: [24a]
c. *My friends tend [$_{TOP}$ the more liberal candidates] to support. [Miyagawa 2017: [24b]

From a cartographic standpoint (Rizzi 1997, 2001; Haegeman 2006, 2010) 'reduction' means truncation: by hypothesis, non-finite clauses are truncated structures, where the upper regions of the CP domain are deleted—or "spliced" in the terminology of Shlonsky & Soare (2011):

(38) ~~ForceP > IntP > TopP~~ … > Fin(ite)P

Miyagawa (2017) relates these two restrictions on ET, by proposing the following universal structural constraint:

(39) Topic Projection: The topic projection TopP is allowed for the complement of A, B, and E, but not for complement of classes C and D.

The constraint in (39) means that all unacceptable instances of ET can be ruled out cartographically: either the TopP is deleted, in the case of infinitivals, or it is unavailable in the first place (a function of semantically driven c-selection).[38]

However, the significance of Miyagawa's work lies less in his formalization of ET than in his proposal that this universal constraint on ET is parameterized in ways that are immediately relevant to our cartographic project.[39] Developing his (2010) work,

[38] As presented here, this is a pure stipulation. However, Miyagawa offers a semantic explanation of the restriction: "the complement of C and D predicates is "subjunctive" and contains a Focus Operator that induces a semantics of alternatives, the predicates…must select the focus operator directly to function properly. If TopP is projected […] this blocks selection…(Miyagawa 2017: 25)". In this way, the restriction is explained by selection, as Rizzi suggests (see earlier discussion)

[39] In fact, Miyagawa proposes *four* language types, predicted by Strong Uniformity, according to which one of two sets of features {**δ** = topic, focus; **ϕ** = agreement} located in C may

Miyagawa (2017) distinguishes two language types:[40] "agreement-based" languages such as English, and "discourse-configurational" languages, such as Japanese. Miyagawa shows that Japanese-type languages deviate from English-type languages by exhibiting a clearer morphosyntactic and syntactic division between different kinds of topic (Aboutness, Contrastive, and Familiar Topics).

Regarding the universality of Hooper & Thompson's taxonomy, Miyagawa points out that, with respect to 'Aboutness Topics', the English pattern {[A, B, E]; [*C,*D]} is exactly replicated in Japanese: he also shows that the permitted *vs*. disallowed complements are further distinguished by combining with different (head-final) complementizers (*to* vs. *koto*), indicative of different functional projections.[41]

Compare first the examples in (35) and (36) above with those in (40) and (41), from Miyagawa (2017: 12-13). Notice that these examples involve simple "aboutness topics":

(40) a. Class A:
Hanako wa [**sono hon wa** kodomo ga yonda *to*] itta
Hanako TOP that book TOP child NOM read C said
'Hanako said that, as for that book, her child read it.'

b. Class B:
Hanako wa [**sono hon wa** kodomo ga yonda *to*] sinziteiru.
Hanako TOP that book TOP child NOM read C believe
'Hanako believes that, as for that book, her child read (it).'

c. Class E:
Hanako wa [**Taroo wa** kanozyo ga suki da *to*] kizuita.
Hanako TOP Taroo TOP she NOM like COP C realize
'Hanako realized that. as for Taroo, he likes her.'

(41) a. Class C:
*Hanako wa [**sono hon wa** kodomo ga yonda *koto*] o hiteishita.
Hanako TOP that book TOP child NOM read C ACC denied
'Hanako denied that as for that book, her child read (it).'

be inherited by T (or both, or neither). In an Agreement-based language (such as English), only *phi* features are inherited by T: in a discourse-configurational language (such as Japanese), it is the **δ** features that are inherited, triggering obligatory topic movement in the narrow syntax. Strong Uniformity predicts two further language types: Class III, where both features are inherited by T (Spanish); and Class IV where both features remain in C (Dinka). Only the first two are relevant to the present discussion.

[40] *Cf*. Li & Thompson's (1976) distinction between 'Subject vs. Topic-prominent languages'.

[41] In (40-42)), I directly copy the glosses given by Miyagawa, which imply more structural parallelism than may be warranted. It is hard to ignore [the fact (!)] that *koto* functions like a nominalizing head—hence, the following ACC marker, which otherwise only attaches to nominal arguments. If this were the correct gloss, the impossibility of ET in D/E classes would be due to a simpler explanation, *viz*., that Topicalization is impossible inside complex noun phrases.

b. Class D:
*Hanako wa [**sono hon wa** zibun ga yonda *koto*] o kookai shita.
Hanako TOP that book TOP child NOM read C ACC regretted
'Hanako regretted that as for that book, her child read (it).' [42]

Having established that Japanese patterns with English in the default case ('Aboutness Topics'), Miyagawa then draws on work by Bianchi & Frascarelli (2010), demonstrating how the two languages diverge with respect to Contrastive Topics (Contrastive Focus) (see Kuno 1976, Büring 2003):[43] this is where the main empirical contribution lies. Miyagawa claims that Japanese diverges from English in allowing CT in all environments—including Class {C,D} contexts, where AT are blocked. Now compare the sentences in (41) with those in (42) below:[44]

(42) a. Hanako wa [**sono hon WA** kodomo ga yonda *koto*] o hiteishita
Hanako TOP that book CONTR.TOP child NOM read C ACC denied
'Hanako denied that as for that book, her child read (it).' (but not this book)

b. Hanako wa [**sono hon WA** self ga yonda *koto*] o kookai shita.
Hanako TOP that book CONTR.TOP child NOM read C ACC regretted
'Hanako regretted that as for that book, her child read (it).' (but not this one)

The implication is clear: if the unacceptability of the examples in (41) is due to the unavailability of TopP, then the Contrastive Topics in (42) must occupy a distinct pre-subject position. Rizzi & Bocci's cartography offers a direct candidate, namely, **Foc**.

…
(1) [Force [Top* [Int [Top* [**Foc** [Top* [Mod [Top* [Qemb [Fin [IP …]]]]]]]]]]
(From Rizzi & Bocci 2017)

The preceding discussion naturally raises the question of which class Vietnamese belongs to. Intuitively, since Vietnamese is normally classified as a "topic-prominent" language,[45] we would expect it to pattern with Japanese. Nevertheless, as was illustrated by the examples (30)-(33) above, *thì* is multifunctional in a way that makes it harder to tie it to one functional position.

The examples in (43)-(45) below make it clear that Vietnamese closely tracks the Japanese patterns: whereas no obvious difference is found between AT and CT topics with {A, B, E} predicates in (43), a contrast *is* found in the *{C,D} sets (44)-(45)—the same contrast as found in Japanese. There is some variation in judgments according to the

[42] It is noteworthy that Miyawa translates the Japanese sentences in (40) and (41)—even the grammatically acceptable cases—as Left Dislocation (LD) structures, rather than ET. No reason is provided for this shift.

[43] Japanese also allows object scrambling without a topic marker, what Miyagawa terms 'familiar topics' (FT); these display a different judgment pattern. This seems to be possible in Vietnamese also; however, space constraints preclude further discussion here.

[44] Notice that in the examples Miyagawa gives the contrast argument is not made explicit. Rather the claim seems to be that the string is only acceptable if the -WA argument, with contrastive focus, is interpreted as a Contrastive Topic. (In our Vietnamese examples, explicit alternatives are presented.)

[45] Li & Thompson (1976), Thompson (1978), Givón (1984), Hawkins (1986).

particular predicate, but where a contrast is observed as in (44b), for example, it always favours the CT alternant, just as Miyagawa's analysis predicts.

(43) a. Class A:
AT: Mary nói rằng **những cuốn sách đó thì** cô ấy sẽ đọc trong hôm nay.
CT: Mary nói rằng **những cuốn sách đó thì** cô ấy sẽ đọc, không phải những cuốn này. (nói = 'say')

b. Class B:
AT: Mary tin rằng **những cuốn sách đó** thì cô ấy có thể đọc trong hôm nay.
CT: Mary tin rằng **những cuốn sách đó thì** cô ấy có thể đọc, không phải những cuốn này. (tin = 'believe')

c. Class E:
AT: Mary nhận ra rằng những **cuốn sách đó thì** cô ấy có thể đọc trong hôm nay.
CT: Mary nhận ra rằng **những cuốn sách đó thì** cô ấy có thể đọc, không phải những cuốn này. (nhận ra = 'realize')

(44) a. Class C (1)
AT: ?Mary phủ nhận rằng những **cuốn sách đó thì** cô sẽ đọc trong hôm nay.
Mary deny COMP PL CLS book DEM TOP PRI FUT read in today
*'Mary denied that those books, she could read today.'
CT: ?Mary phủ nhận rằng **những cuốn sách đó thì** cô sẽ đọc, không phải những cuốn này.
Mary deny COMP PL CLS book DEM2 TOP PRN FUT read, not correct PL CLS DEM1
'Mary denied that those books, she could read today, but not these.'

b. Class C(2)
AT: *Không thể có chuyện **những cuốn sách đó thì** John sẽ đọc trước cuối tuần.
NEG fact ASR story PL CLS book DEM2 TOP John FUT read before end week.
'*It's impossible that those books, John will read by the end of the week.'
CT: Không thể có chuyện **những cuốn sách đó thì** John đọc, không phải những cuốn này.
NEG fact ASR story PL CLS book DEM2 TOP John FUT read, NOT correct PL CLS DEM1
'It's impossible that those books, John read, but not these.'

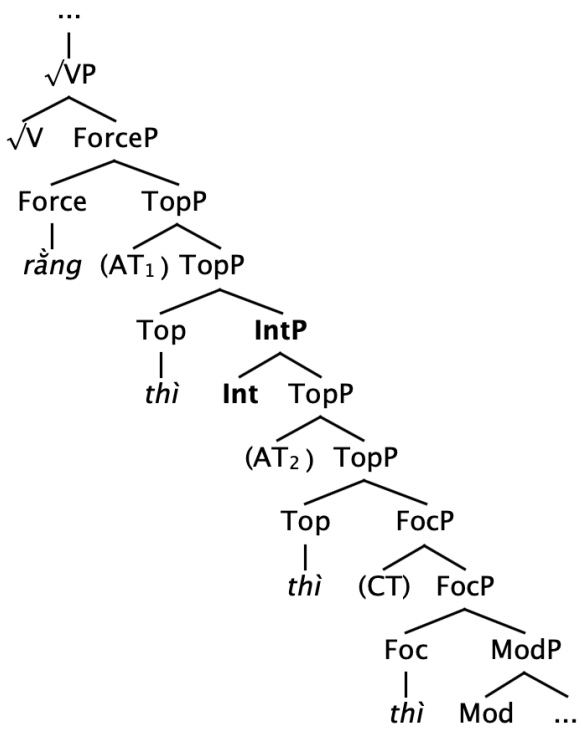
(45) a. Class D (1)
AT: *Mary bực tức rằng **những cuốn sách đó thì** John đọc trong kỳ nghỉ.
Mary resent COMP PL CLS book DEM2 TOP John read in vacation
*'Mary resents that those books, John read while on vacation.'
CT: Mary bực tức rằng **những cuốn sách đó thì** John đọc, không phải những cuốn này.
Mary resent COMP PL CLS book DEM2, John read, NEG correct PL CLS DEM1
*'Mary resents that those books, John read, but not these.'

b. Class D (2)
AT: ?Tôi tiếc rằng **những cuốn sách đó thì** John đọc **mà** không tham.khảo ý.kiến tôi.
I regret COMP PL CLS book DEM2, John read, CONJ NEG consult opinion I
?*'I regret that those books, John read without consulting me.'
CT: Tôi tiếc rằng **những cuốn sách đó thì** John đọc, không phải những cuốn này.
I regret COMP PL CLS book DEM2, John read, NEG correct PL CLS book DEM1
?*'I regret that those books, John read, but not these.'

The most crucial point to keep in mind is that these are judgments of only two native speakers: follow-up investigation is clearly necessary.[46] Nevertheless, supposing the patterns to be valid, the Rizzi & Bocci (2017) hierarchy offers a direct cartographic mapping: Contrastive Topics (CT) can be mapped to **Foc** across all predicate classes, whilst—in those constructions where TopP is allowed—AT constituents could in principle map to either the higher or lower TopP in (46), building on (21) above:

[46] There are two other points to observe here. First—which will be important later when we consider embedded *why*-questions—is the fact that the negative ellipsis in the CT examples seems to require more structure than in English: the direct translation of "not these" → "**không cuốn này*" is unacceptable (nominal ellipsis "*cuốn sách này*" is possible, but NEG cannot attach to a bare DP). Second, note the use of the 'relative pronoun' *mà* as a simple conjunction, in (37b): this shows that *mà*, like *thi*, can lexicalize several different positions.

(46)

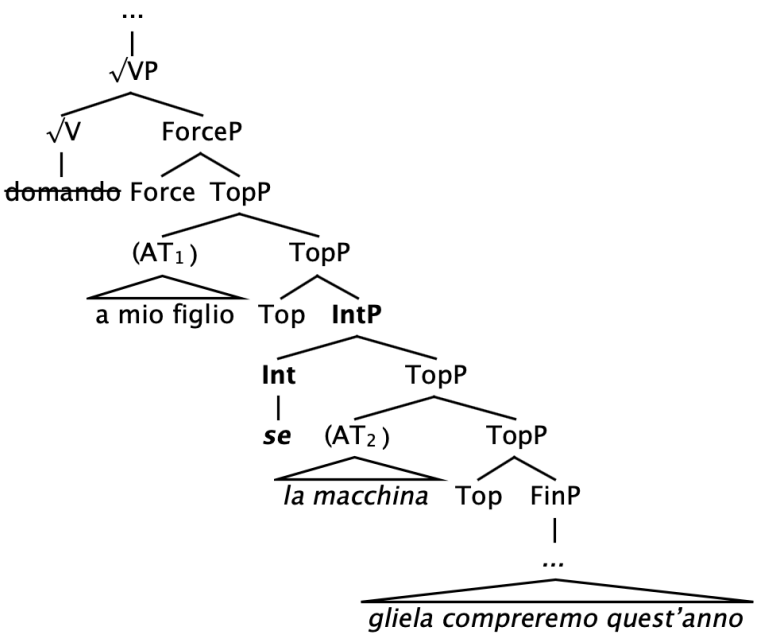


This diagram not only fits the data well,[47] it allows us to go further, to examine whether Vietnamese makes use of the higher or lower position for Aboutness Topics. For, unlike Italian, which allows for double topicalization in embedded clauses—compare (47) below—ET_{AT} in Vietnamese is restricted to one position, that is, only one single topicalization is allowed. The question is which one: AT_1 or AT_2?

(47) a. Mi domando, a mio figlio, **se**, la macchina, gliela compreremo quest'anno
'I wonder, to my son, if, the car, we will buy it to him this year'
(Rizzi & Bocci 2017:[9])

b.

[47] At least on first inspection: when we come to consider *tai sao*, we shall see that although Vietnamese patterns with Japanese with respect to the *availability* of ATs vs. CTs with [C, D] predicates, the data do not support the idea that CTs are projected to a distinct, lower, **Foc** position (*contra* Rizzi & Bocci): on the contrary Contrastive Topics occupy the same unique position as ATs; see below.



The chief diagnostic here is the presumed **Int** head *là*: if *là* can follow *thì*, this would provide evidence of the availability of the higher TopP position (AT_1); conversely, if *là* is unable to appear before the AT in ET contexts, we could exclude the lower position (AT_2) as a possible Topicalisation site. Finally, if *rằng* and *là* can co-occur preceding the AT-*thì*, and the other orders were excluded, this removes the possibility that Vietnamese disposes of a higher topic position.

The deciding data are given in (48) below, with the unacceptability of (48c) weighing strongly in favour of the third option:

(48) a. Mary nói **là** [$_{AT1}$ những cuốn sách đó] **thì** cô ấy sẽ đọc trong hôm nay.
Mary say INT PL CLS book DEM2 TOP PRN FUT read in today
'Mary says that those books, she'll read by the end of the day.'

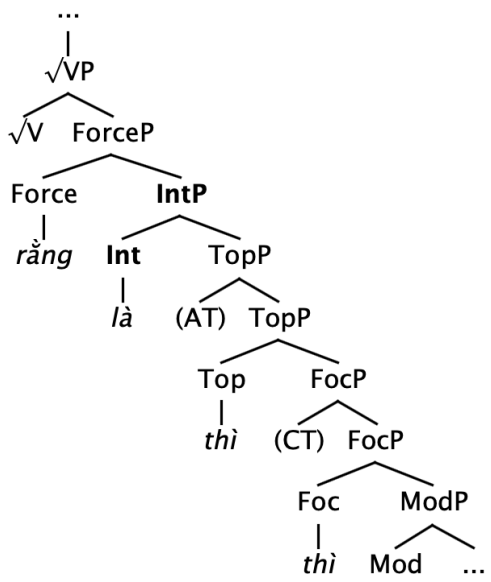
b. Mary nói **rằng là** [$_{AT2}$ những cuốn sách đó] **thì** cô ấy sẽ đọc trong hôm nay.
Mary say COMP INT PL CLS book DEM2 TOP PRN FUT read in today
(as a)

c. *Mary nói **rằn** [$_{AT1}$ những cuốn sách đó] **thì là** cô ấy sẽ đọc trong hôm nay.
Mary say COMP PL CLS book DEM2 TOP INT PRN FUT read in today
(as a)

These examples imply that the higher AT position (AT1) is unavailable in Vietnamese, and that the tree in (46) should be pruned to that in (49). If so, the AT/CT contrast in (43-45) then would vary only according to the availability of the lower (non-recursive) TopP:[48]

[48] There is indirect evidence that the opposite may be true of Japanese only the topmost Top position is available in this language. This is suggested by contrasts between Vietnamese and Japanese L2 learners of English, faced with instances of embedded **Foc** movement in English *wh*-questions "Who under no circumstances should you talk to" vs. "*Under no circumstances who

(49) AT in lower TopP only (*cf*. 40a):



To be sure, one can find examples in which *là* immediately follows *thì,* such as those in (50), but here *thì* is more plausibly analysed as an utterance-initial conjunction—indeed intonation suggests that *thì.là* may be a compound conjunction in such cases:[49]

(50) a. "không phải anh **thì.là** anh khác…"[50]
NEG right you CONJ-CONJ you different…
'If not you, then someone else…'

b. "ừ, **thì.là** em có nỗi buồn thật đẹp…" [song lyric][51]
well, CONJ-CONJ PRN have sadness beautiful
'Well, then I (?you) have a beautiful sadness.'

Before moving on to investigate positions below TopP, there is one further complication to examine, namely, a context in which a topicalized phrase is followed by non-copular *là*. Consider the minimal contrast in (51): if the sentential subject in (51a) counts as an embedded topic (AT), then so too should the apparently identical clause in (51b), where *là* replaces *thì,* yielding a shift in interpretation (*irrealis* to *realis*/finite):

(51) a. [Nó không học toán] **thì** tốt.
PRN NEG study maths ?? nice
'It would be nice for him not to study mathematics.'

---

should you talk to"—Vietnamese speakers replicated NS judgments, whereas Japanese learners strongly preferred the grammatically unacceptable option). See Duffield, Matsuo & Phan (2019), Duffield (in prep.).

[49] Data searches are further complicated by the fact that the string *thì…là* is homographic with the compound noun (cây) *thì là* (= the herb known as dill, in English).

[50] The whole citation is interesting and relevant. không phải anh thì là anh khác, tình là rác có sao phải buồn....| #shorts

[51] https://www.youtube.com/watch?v=-A9PW_X7kJs. At least, this example distinguishes *là* from *rằng*: the latter element would not be possible in this context.

b. [Nó không học toán] **là** tốt.
PRN NEG study maths ?? nice
'It's nice that he didn't study mathematics.'

The reason this presents a problem is that *là* cannot be a copula here since—as noted previously—copular *là* is prevented from appearing with adjectival predicates.[52] But if *là* is a coordinating conjunction, then it should not be found following any topicalized constituent, given (49).

Fortunately for our ongoing analysis, the superficial minimal contrast in (51) belies a significant structural difference, namely, that (51a-*thì*) is non-finite, whereas (51b-*là*) is finite (its own clause). This is demonstrated by the fact that the tense morpheme *đã* is excluded from (52a) below, while being fully acceptable in (52b). In other words, the string *[Nó không học toán]* has a different analysis in the two cases: in (51a), it really *is* a topicalized subject argument (≈ English small clause, or gerund); in (51b), by contrast, the same string is analyzed as a separate clause:[53]

(52) a. [Nó **(*đã)** không học toán] **thì** tốt.
PRN PAST not study maths TOP nice
'It's (a) nice (thing) that he didn't study mathematics.'

b. [Nó **(đã)** không học toán] **là** tốt.
PRN PAST not study maths CONJ nice
'It's (a) nice (thing) that he didn't study mathematics.'

This not only explains how conjunction *là* can follow *[Nó không học toán*] in (51a): it also accounts for the fact that neither *rằng* (nor *là)* can introduce sentential subjects in Vietnamese; compare the examples in (53) below (examples from Duffield 2013, 2019). Both restrictions follow if sentential topics cannot be full clauses, unlike in English.

I postpone to future work discussion of the clause following *là*.

(53) a. (*rằng) họ cười khúc khích] làm chúng em thẹn.
COMP PRN laugh giggle make PL PRN embarrassed
'[*(that) they giggled] embarrassed us.'

---

[52] A way around this would be to suppose that (51b) involves a nominal predicate with a covert nominal head (perhaps *điều*: [~~*điều*~~ *tốt*]), as in (i); the objection is that the overt version does not seem entirely natural:

(i) [Nó đã không học toán] là [$_{NP}$ ~~điều~~ tốt.]
PRN PAST not study maths COP thing nice
'It's (a) nice (thing) that he didn't study mathematics.'

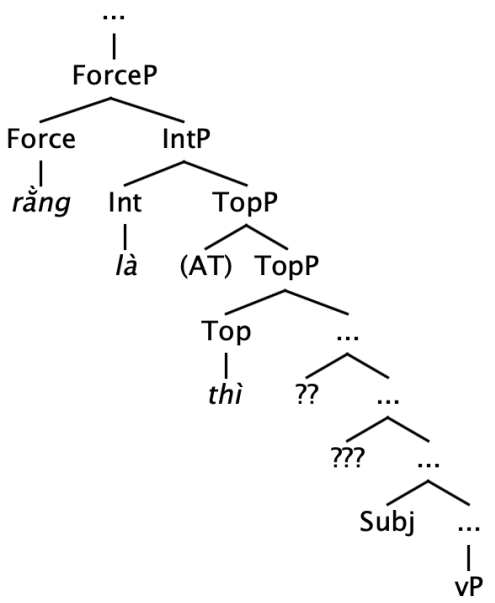
[53] This structure also reveals a contrast between pre-verbal tense/aspect morphemes, which are restricted to finite clauses, *vs*. post-verbal functional categories such as VP-final abilitative *được,* which may appear in topic position:

(i) Không cấp thiết, nhưng [làm được] thì tốt.
'Not urgent, but good to be able to do'
(https://baocamau.vn/khong-cap-thiet-nhung-lam-duoc-thi-tot-a35030.html)

b. (*là) cô ấy rời đi sớm làm tôi ngạc nhiên.
CONJ PRN.DEM leave early make I surprise
'[*(that) she left early] surprised me.'

Having thus established the upper-to-middle sections of the Vietnamese CP, we may turn attention to the remaining region of the left periphery—from the AT topic head to the subject position, schematized in (54) below—to consider three kinds of pre-subject phrase: speaker-oriented adverbials, weak indefinites interpreted as universal quantifiers, and the adjunct *wh*-expression *tại sao* ('why').

(54)



Left Peripheral Adverbials

Even before we come to universal quantifiers and *why* questions, sentences involving subject-oriented adverbials (specifically, the adverbial phrases *quả thật* 'indeed', and *nhất định* 'absolutely') provide evidence that the (AT) topic position is located significantly higher in the left periphery than the canonical subject position.

Consider the paradigm in (55) and (56), adapted from Phan & Duffield (2023). At first glance, the alternation in (55) appears to indicate that *quả thật* may either follow or precede the subject argument. However, given the optional pronunciation of the topic marker *thì*, it is necessary to ensure that there is true variability here: that (55a) is not a case of Top-ADV-(subject) order, such that (55b) represents the base order. The examples in (56) show clearly that this is not the case: first, the unacceptability of (56b) establishes that adverbials cannot appear to the left of an explicit topic; more crucially, the alternations in (57), which contain both an explicit topic [*những cuốn sách đó*] and an explicit subject [*cô ấy*]—compare (44)-(45) above—allow the adverbial to precede or follow the subject; indeed, the latter order seems to be preferred (57b)::

(55) a. Cô ấy **quả thật** đã nói với anh ấy.
PRN.DEM indeed ANT talk with PRN.DEM
'She indeed talked to him.'

b. ?**Quả thật** cô ấy đã nói với anh ấy.
indeed PRN.DEM ANT talk with PRN.DEM
'Indeed, she talked to him.'

(56) a. Cô ấy **thì** [**quả thật** [(pro) đã nói với anh ấy.]]
PRN.DEM TOP indeed ANT talk with PRN.DEM
'Speaking of her, she indeed talked to him.'

b. *__Quả thật__ [cô ấy **thì** [(pro) đã nói với anh ấy]]
Indeed PRN.DEM TOP ANT talk with PRN.DEM
intended (a)

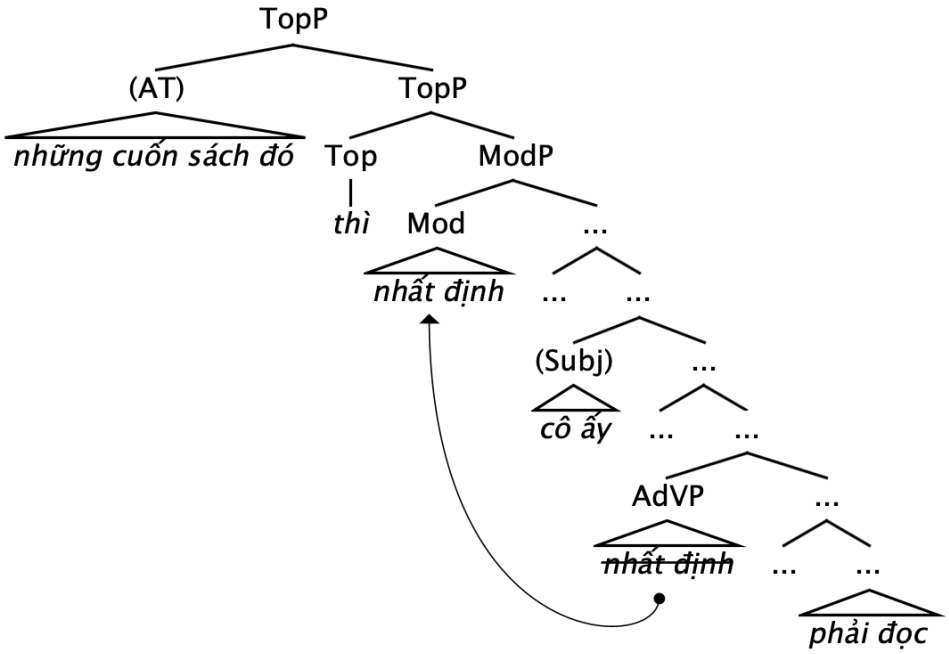
(57) a. ?[Những cuốn sách đó] **thì** **nhất định** cô ấy phải đọc.
PL CLF book DEM_2 TOP absolutely PRN.DEM must read
'(?)Those books absolutely she had to read.'

b. [Những cuốn sách đó] thì cô ấy **nhất định** phải đọc.
PL CLF book DEM_2 TOP PRN.DEM absolutely must read
'Those books, she absolutely had to read.'

From these data, it is reasonable to conclude that adverbials may be moved from a post-subject, pre-verbal slot to a pre-subject position, to the right of the Topic, as diagrammed in (58):[54]

[54] This tree only indicates movement of the adverb: I deliberately ignore all other likely displacements, especially of the subject, and—possibly of the object topic, as well as the issue of possible co-indexation.

(58)



Here, I am co-opting Rizzi & Bocci's **Mod** position for the moved adverbial phrases. This is in the letter of Rizzi & Bocci's cartography, but perhaps not in the spirit, since in their discussion R&B exemplify **Mod** with fronted adverbials such as *rapidamente* ('rapidly'), (59a) elements that can also appear post-verbally, as in (59b). By contrast, *quả thật* is only ever found pre-verbally: this is shown by the unacceptability of the examples in (58c,d) below (see Phan & Duffield 2023):

(59) a. **Rapidamente,** Gianni trovò la soluzione.
'Rapidly, Gianni found the solution.'

b. Gianni ha trovato **rapidamente** la soluzione. [Rizzi & Bocci 2017: 15a]
'Gianni found rapidly the solution.'

c. *Cô ấy đã nói **quả thật** với anh ấy.
PRN.DEM ANT talk indeed with PRN.DEM
'She indeed talked to him.'

d. *Cô ấy đã nói với anh ấy **quả thật.**
PRN.DEM ANT talk with PRN.DEM indeed
(as 58c).

Still, whatever the correct label may be, the distribution of LP-adverbials such as *quả thật* provides a useful diagnostic of the separation of topic and subject positions in Vietnamese.

Quantifier-Raising

Let us now consider fronted quantifier phrases ($QP_{\forall}$). As discussed in Duffield (2007, also 2014), Vietnamese disposes of a subset of universal quantifier expressions that must appear in front of the pre-verbal 'adverbial operator' *cũng*, even where these phrases are interpreted as thematic objects or as adjuncts that would normally appear postverbally. In the examples in (60), the displaced $QP_{\forall}$ appears to the right of the subject, yielding S-O_{QP}-*cũng*-V order in (60a,b)), S-$Adjunct_{QP}$-*cũng*-V, in (60c,d):

(60) a. Anh ấy **từ nào** *(cũng) nhớ. [√S-O_{QP}-cung-V order]
PRN.DEM word WH OP recall
'He remembers every word.'

b. Cô ấy **ai** [*(cũng) quen].
PRN.DEM WH OP know
'She knows everybody.'

c. Anh ấy **bao giờ** [*(cũng) đến muộn]. [√S-ADJ_{QP}-cung-V order]
PRN.DEM WH time OP come late
'He is always late.'

d. Tôi **ngày nào** [*(cũng) tập thể thao]
I day WH OP do exercise
'I do exercises every day.'

However, $QP_{\forall}$-S-V order is also found, as shown by the corresponding examples in (61): both thematic objects and $QP_{\forall}$-adjuncts may also precede the subject:

(61) a. **Từ nào** [anh.ấy [cũng nhớ].] [√O_{QP}-S-cung-V order]
word WH PRN.DEM OP recall
'He remembers every word.'

b. **Ai** [cô ấy [cũng quen]].
WH PRN.DEM OP know
'She knows everybody.'

c. **Bao giờ** [anh ấy [cũng đến muộn]]. [√ADJ_{QP}-S-cung-V order]]
WH time PRN.DEM OP come late
'He is always late.'

d. **Ngày nào** [tôi [cũng tập thể thao]].
day WH I OP do exercises
'I do exercises every day.'

A point to observe is that this $QP_{\forall}$-fronting is obligatory in the case of underspecified weak indefinites (*ai* ('who/anyone'), *nào* ('which/anything', *etc*.,): expressions that would

otherwise be interpreted as *wh*-questions if left in *situ*.[55] This state-of-affairs contrasts with regular, fully specified, universal quantifiers (*mọi~mỗi)*, such as those in (62), where overt $QP_{\forall}$ fronting is impossible, and where *cũng,* if present, only has its non-operator interpretation (= 'also'):

(62) a. *Anh ấy [cũng nhớ **từ nào**]. [*S-cung-V-O_{QP}]
PRN DEM OP recall word WH
'He remembers every word.'
(ok: 'Which word does he remember?')

b. Anh ấy [(cũng) nhớ **mọi từ**]. [√S-cung-V-O_{QP} (mọi)]
PRN DEM also remember each word
'He (also) remembers every word.'

c. *Anh ấy **mọi từ** [(cũng) nhớ]. [*S-O_{QP}-V: mọi]
PRN DEM each word also remember
*'He every word remembers.'

d. Tôi [tập thể thao **mỗi ngày**]. [√S-V-X_{QP}: mỗi]
I do exercise every day
'I do my exercises every day.'

e. ??Tôi **mỗi ngày** [tập thể thao]. . [??S-X_{QP}-V: mỗi]
I every day do exercise
??'I every day do my exercises.'

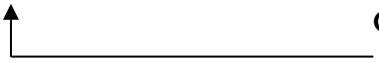
In Duffield (2007), I accounted for these patterns in terms of 'scope evasion':: $QP_{\forall}$-movement was driven by the functional need to avoid being interpreted as *wh*-expressions if they remain *in-situ* (alternatively, as negative polarity items (NPI), when in the scope of the negative operator *không*). In that paper, little attention was paid to the landing-sites involved, the issue of position being less important than that of motivation. Here, the converse is true: to establish the fine structure requires that we determine the precise position of the moved $QP_{\forall}$, in both paradigms ((60) and (61)). It is also necessary to determine the position of the subject argument in the two contexts: the optionality of the topic marker *thì* means that nothing can be taken for granted with respect to starting position.[56]

The most obvious move is to suppose that $QP_{\forall}$-raising involves raising to **Foc**, in the truncated (single topic position) version of Rizzi & Bocci's cartography, schematized in (63):

[55] Alternatively, as negative polarity items (NPI), when in the scope of negation.

[56] The situation is further complicated by the fact that QPs themselves can be topicalized in Class II (Japanese-type) languages; see Miyagawa (*ibid.*).

(63) [Force [Int [Top [**Foc** [[Mod [Qemb [Fin [IP …]]]]]]]]



Following the logic of the previous discussion of adverb placement, such an analysis could allow us to derive both SOV(60) and OSV(61)) order, provided that: (i), the *S* in S-O_{QP}-V is topicalized; equally, (ii), the *S* in O_{QP}-S-V must *not* be (topicalized).

These expectations are only partially borne out, as evidenced by the examples in (64). On one hand, it is clear—as it was with adverbial expressions—that $QP_{\forall}$ cannot raise past the Topic position. In (64a), where the (Aboutness) topic is a non-argument (*câu chuyện này*), which effectively precludes the subject topicalisation, the raised QP intervenes between the Topic and Subject position, as predicted by (63); the unacceptability of (64b) also implies that the QP cannot move beyond the topic position, again as predicted. What is more, the acceptabilty of TOP_{SUBJ}-O_{QP}-V in (64c) offers further support for QP-movement to an intermediate position between TOP and Subject:

(64) a. Câu chuyện này **thì từ nào** anh ấy cũng nhớ.
CLS story DEM_1 TOP word WH $PRN.DEM_2$ ALSO remember
'As for this story, he remembers every word (of it).'

b. ***Từ nào** [câu chuyện này **thì** anh ấy cũng nhớ.]
word WH CLS story DEM_1 TOP $PRN.DEM_2$ ALSO remember
'As for this story, he remembers every word (of it).'

c. "Câu chuyện này? Anh ấy **thì từ nào** cũng nhớ."
CLS story DEM_1 $PRN.DEM_2$ TOP word which ALSO remember
'This story? He's the one that remembers every word (of it).'

d. Câu chuyện này **thì** anh ấy **từ nào** cũng nhớ.
CLS story DEM_1 TOP $PRN.DEM_2$ word WH ALSO remember
'As for this story, he remembers every word (of it).'

All would be well for the analysis in (63), but for the acceptability of (64d): here, the topic position is filled, as in (64a), meaning that the subject cannot be topicalized, yet the $QP_{\forall}$ can still appear to its right {…TOP…SUBJ…$QP_{\forall}$…}. The acceptability of this word order is obviously inconsistent with the idea that $QP_{\forall}$-movement in Vietnamese involves raising to a dedicated **Foc** position, unless certain assumptions are revised.[57]

A more serious problem for the dedicated position analysis in (63) arises when we consider the interaction between fronted QPs and evaluative adverbs. The analysis of fronted adverbs in (58) predicts the acceptability of {TOP-QP-(ADVP)-Subj} order: the

[57] Were it not for the facts in (65), we could account for (64d), if we suppose that even non-topicalized subjects raise to a higher SUBJ position, outside of **FinP**. This would be consistent with ideas presented in Rizzi (2018). It should be clear, though, that any such claim, which could be independently true, requires much more extensive empirical investigation, beyond the scope of this paper.

same analysis, prevents adverbials from appearing between the Top and the $QP_{\forall}$ in **Foc**. Unfortunately, the data show the reverse: the predicted order, (65a) is marginal at best, while the excluded order (65b) is fully acceptable.

(65) a. ??[$_{AT}$Câu chuyện này] thì **từ nào quả thật** anh ấy cũng nhớ
CLF story DEM_1 TOP word WH indeed PRN.DEM ALSO recall
'As for this story, he remembers every word (of it).'

b. Câu chuyện này thì **quả thật từ nào** anh ấy cũng nhớ.
CLF story DEM_1 TOP indeed word WH $PRN.DEM_2$ ALSO recall
'As for this story, he remembers every word (of it).'

In short, the analysis in (63) faces challenges from two directions: one one hand, the acceptability of (64d) shows that the raised $QP_{\forall}$ can appear to the right of the subject, well below the hypothesized **Foc** node; conversely, the acceptability contrast in (65a) reveals that fronted QPs preferentially follow the position of fronted adverbials, to the left of the subject, but well below the hypotheszed **Foc** landing-site. Whilst all of these alternative orders could be derived by *ad hoc* stipulation, the more principled solution would be to revert to the original (2007) analysis, which freely adjoined QP to any projection within IP outside the scope (c-command) of *cũng*. This would correctly derive the variable word-orders, but would necessitate giving up on **Foc** as a unique landing-site for QP movement.

This leaves us with just one pre-subject constituent to place, and which is the *raison d'être* of this whole endeavour: *tại sao* ('*why*'), a *wh*–expression that diverges from other *wh*-phrases across a wide range of languages—also, within constructions in those varieties. As we shall see, sentences involving *tại sao* constitute a Pandora's box of interesting data, potentially jeopardizing the results obtained thus far, as well as as casting doubt on the universality of the Rizzi/Bocci hierarchy.

Where(fore) *Why*?

English *why*

Let us begin with English. The observation that *why* is anomalous *vis-à-vis* other English *wh*-expressions had been noted previously, especially in the child language acquisition literature: see de Villiers (1991) *et seq*.; *cf*. Rizzi (1990). However, until the advent of cartography and an articulated CP, it was not considered theoretically interesting, for want of mechanisms available to capture the distributional differences.

The following minimal contrasts describe the asymmetric lay of the land with respect to *why* (vs. other wh-phrases).. Firstl, the examples in (66) show that in a suitable discourse context *why* or *why (not)*—uniquely among *wh*-phrases—can merge with bare predicate phrases (VPs, PPs, DPs, ADVPs), though not with bare *v*Ps (thematic phrases containing a subject (67).[58] The examples in (68) show that other *wh*-expressions are

[58] These cases should be distinguished from cases of *Aux*-deletion in informal speech, which are only acceptable where an inverted auxiliary would appear in a formal register (√"what you

either excluded from the same environments, or else highly restricted (unless the sentence is parsed as a sluicing context):

(66) a. Why (not) [risk it]?
b. Why (not) [talk to a therapist]?
c. Why [fix the car]?
d.. Why [him]? Why [not her]?
e. Why [only yesterday]? Why not [two weeks ago]?
f. Why [with his brother]? Why not [with his partner]?[59]

(67) a. *Why (not) he/him go there?
b. *Why she/her take the money?
c. *Why (not) they/them came?

(68) a. *When/*who (not) [risk it]?
b. *When (not) talk to a therapist?
c. *How fix the car?
d. *Where him? Where not her?
e. *Who only yesterday? *Who not two weeks ago?
f. *Where with his brother? *How not with his partner?

Perhaps unsurprisingly, these bare *why*-XP fragments may not appear as embedded complements (69a,b), other than in quotative (direct speech) contexts (69c):

(69) a. *She wondered [why her]. *cf*. ("Why her?," she wondered)
b. *She knew [why with his brother].
c. She said: "Why not risk it?"

In contrast to the bare predicate examples in (66)-(68), where it is privileged over other *wh*-expressions, *why* is uniquely excluded from *non-finite* complement clauses (indirect questions), such as those in (70): this is despite being just as acceptable as other fronted *wh*-expressions in the corresponding *finite* contexts. Compare the examples in (70) with those in (71):

(70) a. She knew what/when/how/*why **to** tell him.
b. She wondered who to talk to/how best **to** talk to him/*why **to** stay any longer.

(71) a. She knew what/when/how/why **she should** tell him.
b. She wondered who to talk to/how best **she should** talk to him/why **she should** stay any longer.

doing?" "where you going?" "why you telling me this", *vs*. *where you go", "*what you did?", "Who you talk/*talked to?")

[59] *How come* diverges not just from *why* in these contexts, but also from other *wh*-expressions, in requiring a full and finite complement, though it does marginally allow embedding. Space constraints preclude further discussion.

This asymmetry obtains even in Sluicing contexts, where the elided portion is no less silent: compare (72a) *vs*. (72b).

(72) a. He was told *where* [**to** meet Jane], and *when…* *but not *why* [~~to meet Jane~~].
b. He was told *where* [**he should** meet Jane], and *when…* but not *why* [~~he should meet Jane~~].

These contrasts suggest that semantic interpretability is not at issue here: the constraint seems to be a purely grammatical one. This intuition is further supported by the contrast between the Sluicing contexts in (72) above, and the *Tough*-Construction complements in (73): in the latter context, once dissociated from the non-finite complement, [*why …* [+ non-finite clause]] is just as acceptable as any other embedded *wh*-expression.[60]

(73) a. She knew [what/*why* [it was best/tough [to tell him]]]
b. He wondered [who [it might be easier [to live with: Jane or Amy?]]]
c. He wondered [*why* [it might be easier [to live with Jane.]]]

Why once again diverges from other *wh*-adjuncts when it comes to LD *wh*-movement (that is, at the maximal extension of the clause). For many speakers—see Aoun, Hornstein et al (1987)—*why* exhibits '*that*-trace' effects on a par with *wh*-subjects, such that the presence of the embedded complementizer in (74a) blocks the 'downstairs' contrual of *why*, in what is otherwise an ambiguous string (74b). No similar constraint applies to other *wh*-adjuncts: if anything, the downstairs reading is favored elsewhere, regardless of the presence of *that*. (And, in (74d), where the adjunct is semantically selected by put, the lower reading is the only one available:)

(74) a. When/How did Justin say (that) he had finished the painting?
b. Why did Justin say he had finished the painting early?
c. ??Why did Justin say that he had finished the painting early?
d. Where did Justin say that he had put the painting?

Moreover, for speakers that share the judgments in (74), **why-to* effects can be ameliorated, as shown in (75)—just like *that*-trace effects in (76)—by intervening phrasal material:[61]

(75) a. She told you [why ??(not) to write about this problem]
b. She told you [why *(on no account) to write about this problem].

(76) a. Lee forgot which dishes Leslie had said *that* *(under normal circumstances) should be put on the table.

[60] Notice that whereas the object arguments must be construed with the lowest clause—given the selectional properties of *best, easy*, etc.—*why*, like other adjuncts, can only be construed with the immediately adjacent predicate: i.e., *why-best/tough, not *why-tell; why-easier, not *why-live with.*

[61] See Culicover (1993), Douglas (2017); Sobin (1987, 2002); Salzmann, Häusler, Bader, and Bayer (2013); also Duffield (2018).

b. Which kinds of drugs did you say *that* *(without proper testing) had been released on the market?

The connection with *wh*-subjects is probably not coincidental: notice that exactly these arguments pattern with *why* with respect to non-finite complements; compare (77a) with (70a) above, also with its finite, and non-subject counterparts (77b, c):

(77) a. *She wondered [who **to** tell Mary the news].
b. She wondered [who **had** told Mary the news].
c. She wondered [who (PRO) **to** tell the news to].

The traditional explanation of the contrast in (77) has it that the non-finite clause lacks the relevant features—standardly, Case features—to license an overt subject. The cartographic alternative proposed by Shlonsky & Soare (2011) extends this deficiency account to the *why* cases in (70), except that what is missing are not merely the features but the structural projections containing them: non-finite clauses are truncated versions of their finite counterparts. This is diagrammed in (78), from Shlonsky & Soare (2011): by hypothesis, the licit non-finite clauses involve splicing/deletion of all upper CP projections:

(78) Infinitival clauses are spliced at *Wh*P (Shlonsky & Soare 2011: [11[)
~~ForceP > IntP > TopP > FocP~~ > *Wh*P > FinP

On the explanation given by Shlonky & Soare (2011), the specific unacceptability of *'*why-to'* follows from the special licensing conditions on *why*: whereas regular *wh*-expressions in English are licensed through movement to ***Wh*P**—hence, splicing of the higher nodes has no effect on well-formedness—*why*, again by hypothesis, exceptionally requires movement to **IntP** for feature licensing. This, the authors claim, explains not only the distributional **why*-to constraint, but also the ambiguity of (79b): *why* may not be able to *appear* in the lower clause, but it does allow a "downstairs" reading (long construal), for some speakers, at least.[62]

(79) a. *She asked me [why to resign].
b. Why did you ask her to resign? [Shlonsky & Soare 2011:12]

i. What is the reason x, such that for x, you asked her to resign e.g.:
Because I didn't want to just tell her. (short construal)
ii. What is the reason x, such that you asked her to resign for that reason x?
e.g.: I asked her to resign because of her health, not because of her intelligence . . .(long construal)

If Shlonsky & Soare's analysis is correct, then *why* is not only licensed higher in CP than regular *wh*-expressions: it is also initially merged higher than these other expressions.

[62] Shlonsky & Soare (2012) refer to discussion in Cattell (1978) and Ko (2005:898–899.n31), citing David Pesetsky, pers. comm.).

Shlonsky & Soare's analysis of *why* aligns exactly with Rizzi & Bocci's treatment of Italian *perché*. The latter authors write:

> 'These considerations led to the postulation of an independent position **Int**(errogative), hosting *se* in the head position, and also *wh*-elements like *perché* ('why') and other reason adverbials in the specifier position, as they can also be surrounded by topics and can co-occur with a following focus position (see Rizzi 2001a, and the revision in Shlonsky and Soare 2011), both in main and embedded questions:
>
> (11) Italian
> A Gianni, perché, la macchina, gliela volete regalare?
> 'To Gianni, why, the car, you want to give it to him?
>
> (12) Perché LA MACCHINA/∗LA MACCHINA perché gli volete regalare, e non la moto?
> 'Why THE CAR/∗THE CAR why you want to give to him, and not the motorbike?'
>
> A topic can also occur between **Int** and **Foc**:
>
> (13) *Perché*, a Gianni, LA MACCHINA gli volete regalare, e non la moto?
> 'Why, to Gianni THE CAR you want to give to him, and not the motorbike?'
>
> The integration of **Int** thus gave rise to the following map:
>
> (14) [Force [Top∗ [**Int** [Top∗ [Foc [Top∗ [Fin [IP …]]]]]]]] '

Vietnamese *why*

This raises two kinds of questions: *where* is *why/perché,* initially? and *why* it it *(there)?* Vietnamese, as a *wh*-in situ language, should at least prove helpful in addressing the first question. Perhaps not surprisingly, the answer is more complicated than one might have wished, having come this far.

Before considering our final data sets, it is useful to summarize our interim progress, climbing down from the top of the left periphery. First, we have distinguished two head positions in the highest regions of CP: **Force**, typically lexicalized by the declarative complementizer *rằng*, and **Int**, lexicalized by the force-neutral linker *là*, just in case it functions as a conjunction; in interrogative (alternative question) contexts *liệu* may appear as a specifier of **Int.**

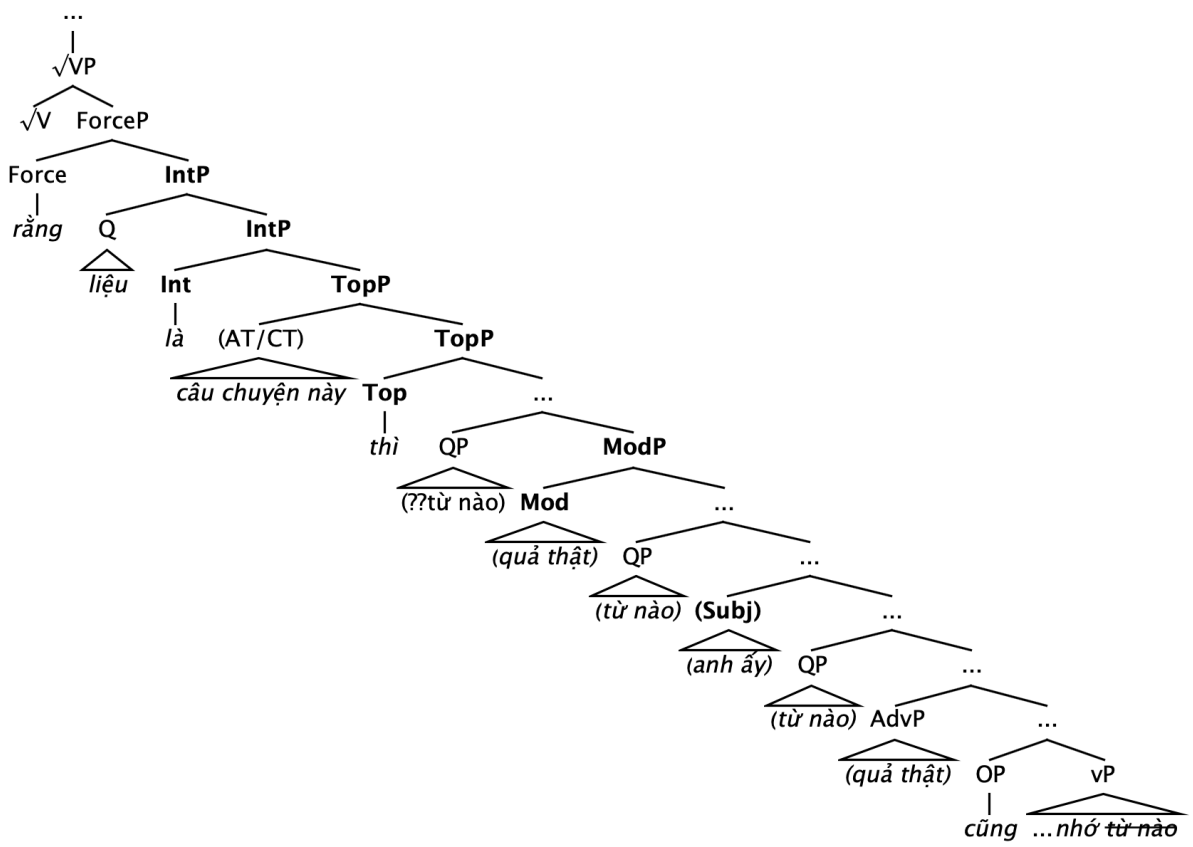
Second, we have established that Aboutness Topics invariably appear below **Int**, apparently ruling out the higher AT position, which is supposed to be available in Italian. We saw that Vietnamese also supports a *functional* distinction between Aboutness Topics and Contrastive Topics, as proposed by Miyagawa (2016, 2018). However, the evidence does not support any *positional* contrast (any more than is found in Japanese,

in fact): hence, the data implies only one Topic position, which may be filled either by an Aboutness topic or a Contrastive Topic, depending on context. This speaks against the availability of a distinct **Foc** node in Vietnamese, *contra* Rizzi & Bocci.

Finally, the case for a distinct **Foc** projection is further eroded when we consider how fronted QPs (weak indefinites with a universal interpretation) interact with the three other classes of left-peripheral elements. On one hand, the data clearly show that QPs move to a region *below* **Force** and **Int** *and* the "dual service" **Top** positions, but *above* the thematic subject position (making certain allowances for example (64d), see fn.). On the other hand, the more flexible interaction of QP with *quả thật*—specifically, the marginal acceptability of (65a)—casts additional doubt on the idea that QPs are moved to a dedicated position. It seems then, as proposed in Duffield (2007), that universal QPs are repelled, not attracted.

The diagram in (80) summarizes the resulting functional sequence for Vietnamese. (It should be clear that some of these elements cannot grammatically co-occur in any particular sentence: the object is to fix their relative positions in one tree.)

(80)



So now, finally, we come to the ultimate (where?) question, namely, where is *why*; more specifically, where is Vietnamese *why* (*tại sao*)? The first point to establish is that *tại sao*, like English *why* and Italian *perché*, only ever appears to the left of the subject. This is made clear by the paradigm in (81):

(81) a. ***Tại sao*** người ta không trồng nấm trên đất…?[63]
why people NEG grow mushrooms in soil
'Why do people not grow mushrooms in soil…?'

b. *Người ta ***tại sao*** không trồng nấm trên đất…?
People why NEG grow mushrooms in soil
'Why do people not grow mushrooms in soil…?'

c. *Người ta không trồng nấm trên đất ***tại sao***?
People NEG grow mushrooms in soil why
'Why do people not grow mushrooms in soil…?'

Second, we need to show that *tại sao*-questions can be embedded. The examples in (82) demonstrate that they can be moreover, that *tại sao*-complements can be introduced by the complementizer *rằng*:

(82) a. Anh đã nhiều lần hỏi em [*rằng* [**tại sao** …]] [song lyric]
PRN PAST many times ask em C why
'I have asked you why, so many times.'

b. Có nhiều bạn hỏi [*rằng* [**tại sao** thầy lại quay được các Video TiếngHàn…]][64]
ASR many friends ask C why teacher upload CAN CLF video lge. Korean
'I have many friends ask why my teacher can make Korean videos…'

c. [$_{DP}$ câu hỏi [*rằng* [**tại sao** bỗng nhiên thế giới lại có thang máy…]]][65]
CLF ask C why suddenly world again have elevator
'the question as to why the world suddenly has elevators…'

Taken together, the two example sets (81)/(82) provide clear evidence that *tại sao* occurs in our crucial region of interest, between **Force** (*rằng*), and the subject position. The question is, exactly where? By examining its position relative to the other left-peripheral elements we have encountered to this point, may allow us to pinpoint the location of reason phrases {*tại sao, why, perché* } prior to movement; we can also investigate whether Vietnamese manifests the **why*-to effect modeled by Shlonsky & Soare (2011).

As noted earlier, cartographic treatments of w*hy/perché* analyze the peripheral position of this element in Italian and English as the result of high external Merge to a Reason Phrase above the subject position, followed by further leftward movement: both Rizzi and Shlonsky & Soare assume that *why/perché* is attracted post-Merge to the {spec, **Int**}

[63] https://vietjack.com/khoa-hoc-tu-nhien-6-ct/tai-sao-nguoi-ta-khong-trong-nam- tren-dat-ma-phai.jsp

[64] https://www.instagram.com/ha_manh_trung/p/C9lzXkDyD7o/?img_index=1

[65] https://tapchithangmay.vn/tai-sao-thang-may-ra-doi/. Note that this sentence provides additional evidence of the availability of {C-tại sao-ADVP-subject} order.

for formal criterial (checking) reasons. Recall that this is necessary to explain the unavailability of **why to*. However, since WHY is required to move further than **Int** in both English and Italian, to satisfy selection—(2) diagrammed in (83)—(*She doesn't know (*that) why she came/Non sa (*che) perché è venuta*'), the intermediate movement (1) is hard to verify independently. However, as the examples in (82) show, *wh*-movement never satisfies c-selection in Vietnamese: hence, wherever *tại sao* is located, it is lower than the **Force** head. But is it as high as **Int?** Or does it remain *in situ*, once merged?

(83) ForceP … IntP > TopP > WhP > ReasonP > … FinP (S&S 2011:[35])
(2) (1) **tại sao**

The examples in (82) leave open the possibility that overt movement to **Int** has taken place. It is not entirely clear what we should expect to find here. On one hand, given that Vietnamese is a *wh*-in situ language, we might not expect to see overt movement of *tại sao*; on the other, we have seen that universal quantifiers *do* undergo obligatory A'-movement for interpretive reasons, and so we might suppose that *tại sao* could exhibit similar behavior.

Before considering the final data sets, it is useful to step back from the formal explanation of the position of *why*, to consider the broader *why*-question, namely, *why* is *why* in the left periphery? (More broadly, *why* is the left periphery (peripheral)?[66] Intuitively, the final answer is an interpretive one:[67] unlike other *wh*-expressions, which probe gaps in an incomplete propositions, *why* is interpreted as having semantic scope over a complete proposition, including specifications of tense and polarity: it 'makes sense' that this semantic scope is reflected in a high syntactic position.[68]

Moving down the tree in (80), the first point to establish is whether *tại sao* appears to the left of **Int**, which would imply overt movement, or to its right. The paradigm in (84) clearly supports the latter option: *tại sao* cannot appear between *rằng* and *là*.

(84) a. Cô ấy không biết (rằng) là [**tại sao** [anh ấy nhớ mọi từ]]
PRN.DEM NEG know C INT why PRN.DEM recall every word
'She doesn't know why he remembers every word.'

b. *Cô ấy không biết (rằng) **tại sao** [là [anh ấy nhớ mọi từ]]
PRN.DEM NEG know C why INT PRN.DEM recall every word
(as a)

[66] In fact, this question can be expanded further, applying to all constituent order. In Duffield (2022, 2024), I propose a semantically-driven theory of word-order involving the principle of Supervenience: constituents—including functional categories and adjunct modifiers—are invariably placed in the minimal syntactic domain to allow unique scope over the elements with which they are interpreted.

[67] I have discussed this problem with many prominent syntacticians and semanticists: the intuitive answer is always the same.

[68] Of course, languages don't always make sense in this way, otherwise there would be no languages with clause-medial sentential negation, but where they do, the intuition is supported. Hence, whether from a formal or intuitive perspective, the occurrence of *tại sao* above the propositional complex (FINP) as expected.

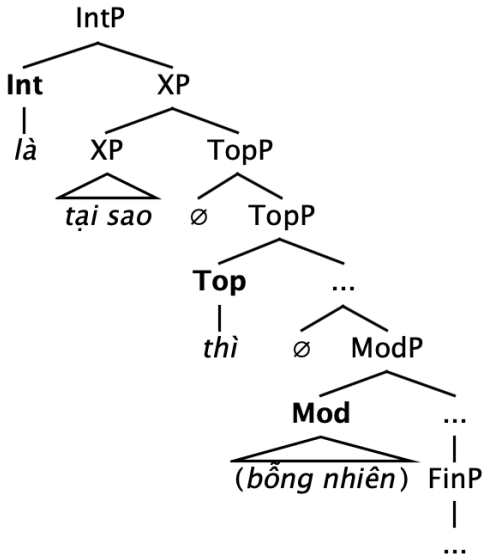
c. *Cô ấy không biết **(rằng) là** [anh ấy **tại sao** nhớ mọi từ]
PRN.DEM NEG know C INT PRN.DEM why recall every word (as a)[69]

d. *Cô ấy không biết (rằng) là [anh ấy nhớ mọi từ] **tại sao**.
PRN.DEM NEG know C INT PRN.DEM remember every word why (as a)

It is worth noting immediately that the unacceptability of (84b) is not due to any simple collocational restriction on {*tại sao…là*} sequences: we will consider grammatical instances of this directly.[70]

Moving down the tree, the next question concerns the position *tại sao* relative to the other left-peripheral elements diagrammed in (80) above: Topic constituents, POV-adverbials and (universal) QPs, respectively. The simplest hypothesis—on either formal or functionalist (semantic) motivation—would place *tại sao* in a dedicated projection having scope over all these elements, as in (85a), below. A plausible alternative might be that *tại sao* would sit between the topic position and the other nodes, as in (85b), on the grounds that *tại sao* would still have scope over all elements interpreted within the proposition; this would be true whether topicalization is derived via movement or co-indexation with a null pronominal. Whichever of these options is allowed, however, we would *not* expect *tại sao* to be able to appear to the right of the other two pre-subject constituents (ADVP, QP).

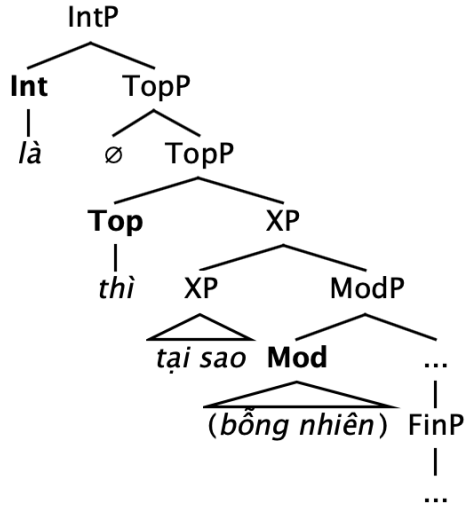
(85) a.



[69] (84c) is acceptable if the subject (*anh ấy*) is interpreted as a topic: as we have seen previously, the topic marker *thì* is often optional.

[70] *Cf*. Doubly-Filled-Comp-Filter (DFCF) that rules out {**wh**…*that*} in Standard English; cf. Bayer & Brandner (2008).

b.



We will again consider these projections in turn. The alternations in (86)[71] reveal that both affirmative hypotheses are borne out, with respect of topic constituents: that is to say, *tại sao* may appear either to the left (86a) or—marginally, to the right (86b)—of the topic:[72]

(86) Cô ấy không biết…

a. …[**tại sao** [*những cuốn sách đó*] **thì** cô ấy có thể đọc, không phải những cuốn này]
why PL CLF book DEM_2 TOP PRN.DEM possible read, NEG correct CLF DEM_1
'I don't know why he was able to read those books, but not these ones.'

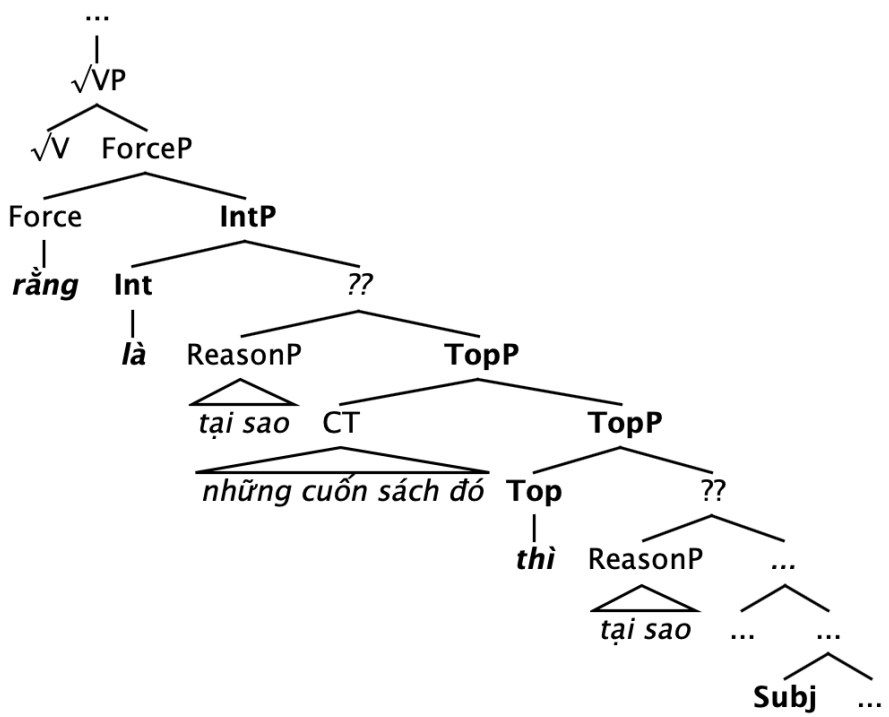
b. …?[*những cuốn sách đó*] **thì** **tại sao** cô ấy có thể đọc, không phải những cuốn này
PL CLF book DEM_2 TOP why PRN.DEM possible read, NEG correct CLF DEM_1
(as a.)

Since we have established that *tại sao* does not raise to **Int**, a formal approach to *tại sao* placement should keep it in its base position [ReasonP]: in (83), this position is unequivocally below the Topic position, closer to that of the surface Subject. Yet, this formal expectation is contradicted by the alternation in (86)—particularly given the preference for the (a) alternant, where *tại sao* precedes the Topic. This is unexpected for two reasons: not only is there no plausible probe between **Int** and **Top**, even if there were, standard feature-driven movement should then prohibit optionality, ruling out (86b). Consider (87). (Alternatively, if [ReasonP] were projected exceptionally high in Vietnamese, to capture (86a) without recourse to movement, (86b) should still be ruled out, given the ban on lowering).

[71] Compare the CT sentences in (44-45) above.

[72] In these examples, contrastive topics are illustrated, but aboutness topics exhibit the same alternation.

(87)



In this case, as was also true of QP placement earlier, the semantic supervenience approach makes better predictions than any appeal to feature-driven movement, or to a fixed functional sequence.

The same is true of adverbials. In a previous section we saw that adverbials such as can appear either to the right of the subject—their presumed base position—or in a position to the left of the subject, below the Topic phrase: following Rizzi & Bocci's cartography in (1), we analyzed this higher adverbial position as **Mod**. If we now insert this node into Shlonsky & Soare's functional sequence in (83), we obtain the sequence in (88):

(88) ForceP … IntP > TopP > **ModP** > WhP > ReasonP > … FinP (S&S 2011:[35])
(2) (1) *tại sao*

The formal prediction is clear: if *tại sao* remains *in situ* in (88), then it should follow higher adverbs {TOP-ADV-*tại sao*-SUBJ}, there being no plausible position between **TopP** and **ModP** for it to occupy.

Once again, the facts are clear: the predicted order (89a) is unacceptable, while the underivable order (89b) is (largely) unexceptional. The most preferred order, (89c) is where the adverb remains in its base position, so this sentence does not speak directly to the question at hand.

(89) *Mary hỏi rằng…

a. …*[những cuốn sách đó] thì **nhất định** *tại sao* [cô ấy phải đọc]
PL CLF book DEM_2 TOP absolutely why PRN must read
'Mary asked why that book she absolutely had to read.'

b. …?[những cuốn sách đó] thì *tại sao* **nhất định** [cô ấy phải đọc]
PL CLF book DEM_2 TOP why absolutely PRN must read
(as a)

c. …[những cuốn sách đó] thì *tại sao* [cô ấy **nhất định** phải đọc.]
PL CLF book DEM_2 TOP why PRN absolutely must read
(as a)

For completeness, let us consider the position of *tại sao* relative to moved $QP_{\forall}$s. Since we have just seen that *tại sao* invariably appears to the left of the higher adverbs, and since we observed previously in (65) that moved $QP_{\forall}$ appear to the right of the same adverbs—apart from the highly marginal (65c) above—it should be logically impossible for a moved $QP_{\forall}$ to appear to the left of *tại sao*, at least given normal assumptions about how syntax works. Fortunately for syntax, {TOP-$QP_{\forall}$-*tại sao*} is just as unacceptable as it should be, as revealed by the examples in (90) below. Here, the paradigm aligns exactly with that in (89), *viz.*: example (90a) shows that {$QP_{\forall}$-*tại sao*} order is unacceptable; example (90b) shows the acceptable left-peripheral order {*tại sao*-$QP_{\forall}$}, while (90c) shows that $QP_{\forall}$, like pre-verbal adverbs, can appear within FinP (unlike *tại sao*, which is restricted to the higher CP-domain).

(90) a. *Cô ấy không biết [**từ nào** [*tại sao* [anh ấy [cũng nhớ]]]
PRN.DEM NEG know word WH why PRN.DEM OP recall
'She doesn't know why he remembers every word.'.

b ?Cô ấy không biết [*tại sao* [**từ nào** [anh ấy [cũng nhớ]]]
PRN.DEM NEG know why word WH PRN.DEM OP recall
'She doesn't know why he remembers every word.'.

c. Cô ấy không biết [*tại sao* [anh ấy **từ nào** [cũng nhớ]]]
PRN.DEM NEG know why PRN.DEM word WH OP recall
'She doesn't know why he remembers every word.'

These results are theoretically significant, in as much as they reinforce the conclusion that *tại sao* is merged/adjoined relatively high in the left periphery: *below* **Top**, but *above* all other pre-subject positions. The implication, for those that subscribe to a cartographic account of the Left Periphery, must be *either* that ReasonP is ordered differently in Vietnamese relative to other CP projections, so that the functional sequence is subject to parametrization, **or**—taking a universalist stance—that ReasonP is generally projected significantly higher than supposed by Rizzi and Shlonsky & Soare, something like (91) perhaps:

(91) ForceP > IntP > TopP > ReasonP > **ModP** > WhP > … FinP (cf. S&S 2011:[35])

Although I am unconvinced of the viability of this approach, especially given the variability of *tại sao* placement with respect to Topics and Adverbials, it could still derive the finite *vs*. non-finite asymmetry (**why-to*), with which we started this section: if ReasonP was projected higher than WhP, one could still derive **why-to* constraint, by splicing the top three nodes (92a), as originally proposed (see 78), assuming covert movement to **Int** in Vietnamese; alternatively, the splicing could be more radical, entirely removing ReasonP and ModP (as in 92b):

(92) a. ~~ForceP > IntP > TopP >~~ ReasonP > **ModP** > WhP > … FinP
b. ~~ForceP > IntP > TopP > ReasonP > ModP >~~ WhP > … FinP

Note that the latter analysis removes the need to assume raising to **Int** (either overt or covert): **why-to* would be blocked *ab initio*, for want of an initial merger site. (The alternate possibilities in (92) could also explain the inter-speaker variation in judgments in (79) above: speakers that accept long-distance readings would have the splicing in (92a); those for whom long-distance readings are blocked would have the full splice in (92b).)

Which brings us, almost, to the end of our descriptive journey. The one question remaining is whether *tại sao* displays the same finiteness restriction as observed in English and Italian, despite its not raising overtly to **Int**, even in finite clauses. If not, all bets are off, so to speak: Vietnamese is simply different from English and Italian with respect to *why* complements. On the other hand, if Vietnamese does exhibit the same constraint, lack of overt *wh*-movement notwithstanding, then some further consideration must be given, either to better accommodate the distributional facts of Vietnamese within a revised cartography, or to find alternative, non-feature driven, explanations for the constraint. The choice, I suggest, depends as much on one's theoretical aesthetics as on any empirical argument.[73]

The key contrasts are to be found in the comparison between the examples in (93) and (94); compare the English examples in (70) and (71) above.

(93) a. Chị ấy biết chị nên nói cái *gì*.
PRN.DEM know PRN should say CLF what
'She knew what she should say.'

b. Chị ấy biết chị nên đi *đâu*.
PRN.DEM know PRN should go where
'She knew where she should go.'

c. Chị ấy không biết chị nên giải thích vấn đề như *thế nào*.
PRN.DEM NEG know PRN should explain problem as how
'She doesn't know how she should explain the problem.'

[73] This may not be too problematic: after all, many other Germanic languages disallow *all* wh-non-finite complements, not just with *why*.

d. Chị ấy tự hỏi chị nên rời đi *khi nào*.
PRN.DEM wonder PRN should leave when
'She wonders when she should leave.'

e. Chị ấy tự hỏi tại sao chị nên rời đi.
PRN.DEM wonder why PRN should leave
'She wonders why she should leave.'

(94) a. Chị ấy biết nói cái *gì*.
PRN.DEM know say CLF what
'She knew what to say.'

b. ?Chị ấy biết chị nên đi *đâu*.
PRN.DEM know PRN should go where
'She knew where to go.'

c. Chị ấy không biết giải thích vấn đề như *thế nào*.
PRN.DEM NEG know explain problem as how
'She doesn't know how to explain the problem.'

d. ?Chị ấy tự hỏi rời đi *khi nào*.
PRN.DEM wonder leave when
'She wonders when to leave.'

e. ?Chị ấy tự hỏi tại sao rời đi.
PRN.DEM wonder why leave
''*She wonders why to leave.'

Although once again not as clear as one might wish, the results are consistent with formal expectations: whereas all of the finite options in (93) are grammatically acceptable, the pattern of results among the non-finite complements (94)—especially the contrast between (94a) and (94c) on one hand, and (94e, *tại sao*) on the other, is consistent with the splicing analysis in (92b), which would prevent *tại sao* from being merged, whilst allowing other *wh*-elements.

Nevertheless, there remains the two-fold problem: (i), that (94b) and (94d) are less than fully acceptable; (ii), that (94e) is more acceptable (for the speakers consulted) than it should be, given the theory. On the first count, it is worth noting, with respect to (94b), that other licit instances of *đi đâu* ('to go where') *are* attested, such as the example in (95a) below. As for the marginality of *khi nào* in (94d), other *when*-adjuncts are possible with non-finite complements ((95b); furthermore, *khi nào* becomes more acceptable if fronted (95c):

(95) a. Cậu Abdullah đồng ý dẫn tôi đi…. và chỉ cho tôi biết *đi đâu*.
PRN Abdullah agree take me go and let me know go where
'Abdullah agreed to take me…and to let me to know where to go.'

b. Chị không biết [rời đi *lúc nào*]
PRN NEG know leave time WH
'She doesn't know when to leave.'

c. Chị không biết [khi nào [rời đi]]
PRN NEG know time WH leave
'She doesn't know when to leave.'

As for the marginal acceptability of (94e), I have no clear explanation other than to note that even some English speakers find '*why-to' sentences acceptable. Possibly, then, this could be a sampling error; alternatively, these speakers may exceptionally allow the shorter splicing in (92a). These wrinkles notwithstanding, the general pattern is in line with formal expectations.

Summary

Aiming for a better understanding of the fine structure of the left periphery in Vietnamese, we have examined ordering constraints and interactions among five different kinds of constituent appearing to the left of the core proposition: (i) complementizers/relativizers {*rằng, liệu, (mà), (là)*}: (ii) topicalized expressions, optionally marked by *thì*; (iii) fronted adverbials appearing in the left periphery; (iv) fronted $QP_{\forall}$ (arguments and adjuncts) with universal interpretations; (v) the *wh*-expression *tại sao* (*why*). Except for QPs, all these elements appear not only to the left of the proposition, but also to the left of the non-topicalized subject: that is, they appear above FinP. Hence, details of their interaction and relative orderings provide important clues to the fine structure of the Left Periphery.

With respect to complementizers, the rigid ordering of *rằng vs*. the multifunctional element *là* illustrated in (19) above provides evidence of at least two positions in the highest layer of the clause:

(96) ForceP > IntP >

With respect to embedded topicalization, our investigation reveals an important difference between Italian, with its proliferation of Topic positions (above and below **Int**) and Vietnamese, having a unique lower slot below **Int**:

(97) ForceP > IntP > TopP >

What is more, despite displaying the same *functional* contrast between Aboutness Topics and Contrastive Topics (43-45) as found in Japanese, there is no evidence for any *positional* distinction between the two topic types; in fact, there is no direct evidence from any source for a distinct focus position (**Foc**). Pending further study, therefore, I suppose that this position is not a component of the Vietnamese left periphery.

Moving further down the clausal spine, the marginal possibility of placing *tại sao* below the topic in embedded clauses (86b) is consistent with the existence of a dedicated ReasonP projection—from which *tại sao* typically raises to have scope over

the Topic constituent (86a), as in (98). Alternatively, *tại sao* may be a pure adjunct, as it is in bare non-integrated structures in English (*Why not go? Why say that?*), its position determined by interpretive scope requirements; in that case there is no need to posit a dedicated Merger site.

(98) ForceP > IntP > TopP > (ReasonP) >

Finally, the interactions between adverbials and fronted quantifier expressions ($QP_{\forall}$) reveal an important pre-subject/post-subject distinction: whereas either order is acceptable where both constituents appear to the right of the subject, fronted adverbs must appear to the left of universal QPs when both appear in the left periphery (65). This is consistent with Rizzi & Bocci's proposal of a distinct **Mod** position for fronted adverbials (albeit no special interpretive effect applies to this position in Vietnamese).[74]

(99) ForceP > IntP > TopP > (ReasonP) > ModP > (WhP) >... FinP

[74] Example (95c), involving a fronted *when*-phrase, is consistent with the postulation of a distinct WhP; however, in the absence of example sentences involving complementizers, embedded topics and fronted *khi nao,* this is only suggestive.

Conclusion (Nusquam sunt dracones)

As implied in the Introduction, cartography is a curiously unsatisfying product of exploration: whether considered literally or figuratively, a tremendous amount of legwork is required only to demystify uncharted territory, especially in the higher ground; this is true whether one is hiking up (in typical physical surveys, before the age of drones) or climbing down (in syntactic ones).

It is nonetheless necessary work, since one can only ask interesting *why*-questions (of the broad range kind) once one has a clear picture of the lay of the land: where *what* is, or—more pertinently—where *why?* is. Hence, though much of the resulting chart of the Vietnamese left periphery may be unsensational—in contrast to medieval maps of uncharted territory, there are no dragons here—what has been uncovered allows us to pose more useful research questions. These include, but are not limited to, the following:

- If, as seems to be the case for Vietnamese, the availability of particular constituent orders within the left periphery is determined more by semantic considerations—scope and scope evasion—than by the featural properties of dedicated positions, is this also true of other languages, such as Italian, or are we dealing with a formal typological contrast? (Consider, for example, the difference between Contrastive and Aboutness Topics, the variable positioning of adverbs and universal QPs, the pre-/post-Topic positioning of *tại sao*: in each case, observed alternations are more consistent with functional-interpretive demands than with a fixed cartography.)
- Vietnamese also challenges the idea of a sharp delineation between main and subordinate clauses: at least in some contexts what would be unambiguous subordinators in English and Italian have a much looser linking function; other

constructions—passives, causatives, other serial verb structures—exhibit similar main~subordinate indeterminacies. Once again, this raises the question of whether such differences imply a typological contrast between hypotactic and paratactic languages (Clark 1992), or whether they simply urge a more flexible and nuanced view of functional categories.

- More narrowly, why no "Why-to"? Splicing of non-finite clauses (Shlonsky & Soare 2011), Truncation (Hooper & Thompson 1973), Exfoliation (Pesetsky 2021) typically operates to facilitate extraction and binding, thus increasing effability. In the case of *why* questions, however, splicing reduces effability. Whether considered from a formal or functional position, this is a theoretical mystery: why should this be, even in as permeable a language as Vietnamese?

So, whilst the fact-finding may be nearly complete, for now, the real discoveries still lie in front of us, once we begin to engage with the theory. And at that point, it may well turn out that the map is wrong. Maybe we should just head out, *sans savoir pourquoi.*

Naturally, Baudelaire said it better:

> *…Mais les vrais voyageurs sont ceux-là seuls qui partent*
> *Pour partir; coeurs légers, semblables aux ballons,*
> *De leur fatalité jamais ils ne s'écartent,*
> *Et, sans savoir pourquoi, disent toujours: Allons*!
>
> But real travellers are just those for whom departure
> Is its own reward; [who leave], hearts light as air
> Not to evade their fate/[but] always declaring
> – without knowing why – 'Let's go!'
>
> Charles Baudelaire, 'Le Voyage' (Les Fleurs du Mal, 1861)

Abbreviations
The following abbreviations are used in this manuscript:

ANT	Anterior marker
ASR	Assertion Marker
C	Complementizer
CLF	Noun Classifier
CONJ	Conjunction
COP	Copular/Linker
DEM	Demonstrative Marker
INF	Infinitive Affix
LOC.COP	Locative copula
NEG	Negation Morpheme
NEG^{Q}	Final Negation (interrogative brace)
OP	Scopal Operator
PAST	Past Tense
PL	Plural morpheme
PRN	Pronoun (Vietnamese equivalent)
PROG	Progressive

QP	Quantified Expression
$QP^{\forall}$	Universal Quantifier
REL	Relativizer
SUBJNC	Subjunctive
TOP	Topic Marker
WH	*Wh*-expression

References

Aboh, E., Haeberli, E., Puskás, G., & Schönenberger, M. (Eds.). (2017). *Elements of Comparative Syntax: Theory and Description*. Berlin, Boston: De Gruyter Mouton.

Bayer, J., & Brandner, E. (2008). 'On Wh-Head-Movement and the Doubly-Filled-Comp Filter', in C. B. Chang & H. J. Haynie (eds), *Proceedings of the 26th West Coast Conference on Formal Linguistics*, Somerville, MA: Cascadilla Press.

Bianchi, V., & Frascarelli, M. (2010) 'Is Topic a Root Phenomenon?', *IBERIA: An International Journal of Theoretical Linguistics 2*(1). Retrieved from https://revistascientificas.us.es/index.php/iberia/article/view/86

Bresnan, J. (1970) 'On complementizers: towards a syntaxric theory of complement types', *Foundations of Language 6*: 297-321.

Büring, D. (2003) 'On D-Trees, Beans, and B-Accents', *Linguistics and Philosophy 26*: 511–545.

Cattell, R. (1978) 'On the source of interrogative adverbs', *Language 54*: 61–77.

Chomsky, N. (1957) *Syntactic structures*, s -Gravenhage,: Mouton.

Chomsky, N. (1965) *Aspects of The Theory of Syntax*, Cambridge, MA: MIT Press.

Chomsky, N. (1977) *Essays on form and interpretation*, New York: North-Holland.

Chomsky, N. (1981) *Lectures on Government and Binding: The Pisa Lectures* (Vol. 9), Dordrecht: Foris.

Chomsky, N. (1985) *Knowledge of language: its nature, origin, and use*, New York: Praeger.

Chomsky, N. (1989). 'Some Notes on Economy of Derivation and Representation', in I. Laka & A. Mahajan (eds), *MIT Working Papers in Linguistics #10* (pp. 43-74), MIT.

Chomsky, N. (1993). 'A Minimalist Program for Linguistic Theory', in K. Hale & S. J. Keyser (eds), *The View from Building 20: Essays in honor of Sylvain Bromberger.*, Cambridge, MA: MIT Press.

Chomsky, N. (1998) *Minimalist Inquiries: the framework.* (ms.). MIT.

Chomsky, N. (2000). 'Minimalist Inquiries: the framework', in R. Martin, D. Michaels, & J. Uriagareka (eds), *Step by Step: Essays on Minimalist Syntax in honor of Howard Lasnik* (pp. 89-155), Cambridge, MA: MIT Press.

Chomsky, N. (2005) 'Three Factors in Language Design', *Linguistic Inquiry 36*: 1-22.

Chomsky, N. (2008). 'On Phases', in R. Freidin, C. Otero, & M. L. Zubizarreta (eds), *Foundational Issues in Linguistic Theory, Essays in Honor of Jean-Roger Vergnaud* (pp. 133-166), Cambridge, Mass: MIT Press.

Chomsky, N., & Lasnik, H. (1977) 'Filters and control', *Linguistic Inquiry 8*: 425-504.

Chung, S., & McCloskey, J. (1987) 'Government, Barriers and Small Clauses in Modern Irish', *Linguistic Inquiry 18*: 173-237.

Cinque, G. (1994). 'On the evidence for partial N movement in the Romance DP', in G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi, & R. Zanuttini (eds), *Paths Toward Universal Grammar* (pp. 85-110), Georgetown: Georgetown University Press.

Cinque, G. (1999) *Adverbs and functional heads: a cross-linguistic perspective*, New York: Oxford.

Cinque, G., & Rizzi, L. (2015). '65The Cartography of Syntactic Structures', in B. Heine & H. Narrog (eds), *The Oxford Handbook of Linguistic Analysis* (pp. 0): Oxford University Press.

Citko, B. (2014) *Phase Theory: An Introduction*, Cambridge: Cambridge University Press.

Clark, M. (1992) 'Conjunction as Topicalizer in Vietnamese', *Mon-Khmer Studies 20*: 91-110.
Culicover, P. (1993) 'Evidence against ECP accounts of the *that*-trace effect', *Linguistic Inquiry 24*: 557-561.
Culicover, P. (1999) *Syntactic Nuts: hard cases, syntactic theory and language acquisition*, Oxford: Oxford University Press.
de Villiers, J. (1991) 'Why questions', *University of Massachusetts Occasional Papers in Linguistics (special editions): Papers in the Acquisition of WH; Proceedings of the UMASS round table 17*: Article 8.
Deprez, V., & Pierce, A. (1993) 'Negation and Functional Projections in early grammar.', *Linguistic Inquiry 24*(1): 25-67.
Douglas, J. (2017) 'Unifying the that-trace and anti-that-trace effects.', *Glossa: a journal of general linguistics 2(1): 60*: 1–28.
Duffield, N. (2007) 'Aspects of Vietnamese clause structure: separating tense from assertion', *Linguistics 45*(4): 765-814.
Duffield, N. (2013). 'Head-First: On the head-initiality of Vietnamese clauses', in D. Hole & E. Löbel (eds), *Linguistics of Vietnamese: an international survey* (pp. 127-155), Berlin: de Gruyter Mouton.
Duffield, N. (2014) 'Minimalism and Semantic Syntax: Interpreting Multifunctionality in Vietnamese'. Paper presented at the International Conference on The Linguistics of Vietnam in the Context of Renovation and Integration, Hanoi.
Duffield, N. (2018). 'Parsing Out in English and in Vietnamese', in H. T. Bao, L. A. Cuong, H. V. Khuong, & B. T. Lam (eds), *Proceedings of 5th NAFOSTED Conference on Information and Computer Science (NICS)* (pp. 311-316), Ho Chi Minh City, Vietnam: Science and Technics Publishing House.
Duffield, N. (2022) 'Whales, and other Mammals: A Naturalist's Take on Syntactic Variation (with particular reference to Vietnamese)'. Paper presented at the JSEALS (Special Publication Number 9).
Duffield, N. (2024). 'Nominative Case and X'-Theory Revisited: "We bade it a tedious returning"', in C. Sevdali, D. Mertyris, & E. Anagnostopoulou (eds), *The Place of Case in Grammar* (pp. 388-426), Oxford: Oxford University Press.
Duffield, N., & Matsuo, A. (2019) ''The young man knows the rules, but the old man knows the exceptions': L2 learners' sensitivity to partial rules in English wh-questions', *EICE Technical Report TL2019-11, 7-11.*
Duffield, N., & Phan, T. (2021). *Introducing Vietnamese Syntax: A Generative-Typological Approach*. Paper presented at the 6th international conference on Vietnamese Studies (Linguistics Session), Vietnam Academy of Social Sciences, Hanoi.
Emonds, J. (1970) *Root and structure-preserving transformations.* (PhD. dissertation). MIT. Cambridge.
Emonds, J. (1976) *A Transformational Approach to English Syntax: Root, Structure-preserving, and Local Transformations*, New York: Academic Press.
Emonds, J. (1980) 'Word Order in Generative Grammar', *Journal of Linguistic Research 1*: 33-54.
Endo, M., Shibuya, M., & Hirakawa, M. (2016). 'Explicit Instruction vs. Natural Exposure in L2 Acquisition of Adjective Ordering in English', in D. Stringer, J. Garrett, B. Halloran, & S. Mossman (eds), *Proceedings of the 13th Generative Approaches to Second Language Acquisition Conference (GASLA 2015)* (pp. -71): Cascadilla Proceedings Project.
Fraser, B. (2009) 'An account of discourse markers', *International Review of Pragmatics 1*: 293-320.
Givón, T. (1984) *Syntax: A Functional-Typological Introduction* (Vol. Volume 1), Amsterdam: John Benjamins.
Guilfoyle, E., & Noonan, M. (1992) 'Functional categories and language acquisition', *Canadian Journal of Linguistics 37*: 241-272.

Haegeman, L. (2006) 'Conditionals, factives and the left periphery', *Lingua 116*(10): 1651-1669. doi:https://doi.org/10.1016/j.lingua.2005.03.014
Haegeman, L. (2010) 'The internal syntax of adverbial clauses', *Lingua 120*(3): 628-648. doi:https://doi.org/10.1016/j.lingua.2008.07.007
Hạo, C. X. (1991) 'Some preliminaries to the syntactic analysis of the Vietnamese sentence', *Mon-Khmer Studies 20*: 137-151.
Hawkins, J. A. (1986) *A Comparative Typology of English and German: Unifying the Contrasts*, Austin: University of Texas Press.
Hawkins, J. A. (2001) 'Why are categories adjacent', *Journal of Linguistics 37*: 1-34.
Hyams, N. (1986) *Language acquisition and the theory of parameters*, Dordrecht: Reidel.
Kayne, R. S. (1994) *The Antisymmetry of Syntax*, Cambridge, Massachusetts: MIT Press.
Ko, H. (2005) 'Syntax of why-in-situ: Merge into [SPEC,CP] in the overt syntax', *Natural Language and Linguistic Theory 23*: 867–916.
Kuno, S. (1976). 'Subject, Theme, and the Speaker's Empathy: A Reexamination of Relativization Phenomena', in C. N. Li (ed), *Subject and Topic* (pp. 417-444), New York: Academic Press.
Lasnik, H., & Uriagareka, J. (2022) *Structure: Concepts, Consequences, Interactions*, Cambridge, MA: MIT Press.
Li, C. N., & Thompson, S. A. (1976). 'Subject and Topic: a new typology of language', in C. N. Li (ed), *Subject and Topic* (pp. 457-489), New York: Academic Press.
McCawley, J. D. (1970) 'English as a VSO Language', *Language 46*: 286-299.
McCloskey, J. (2017). 'Ellipsis, polarity and the cartography of verb-initial orders in Irish', in E. Aboh, E. Haeberli, G. Puskás, & M. Schönenberger (eds), *Elements of Comparative Syntax: Theory and Description* (pp. 99-152), Berlin: Walter de Gruyter.
Miyagawa, S. (2009) *Why Agree? Why Move?: Unifying Agreement-Based and Discourse-Configurational Languages*: The MIT Press.
Miyagawa, S. (2017) 'Topicalization', *Gengo Kenkyu (Linguistic Research) 152*: 1-29.
Moltmann, F. (2020). 'Abstract Objects and the Core-Periphery Distinction in the Ontological and the Conceptual Domain of Natural Language', in J. L. Falguera & C. Martínez-Vidal (eds), *Abstract Objects: For and Against* (pp. 255-276), Cham: Springer International Publishing.
Nguyen, M. C. (2021) *The particle thì in Vietnamese.* (Masters). Liberty University.
Ouhalla, J. (1991) *Functional Categories and Parametric Variation*: Croom Helm, London.
Park, S. (2006) 'The development of complementizer and relativizer *that* in the History of English', *English Language and Linguistics 22*: 197-216.
Pesetsky, D. M. (2021) 'Exfoliation: towards a derivational theory of clause size', *Lingbuzz lingbuzz/004440*.
Phan, T. (2013) *Syntax of Vietnamese Aspect.* (PhD dissertation). University of Sheffield. Sheffield.
Phan, T., & Duffield, N. (2022). 'A Roadmap to Vietnamese Phrase Structure', in C. Shei & S. Li (eds), *The Routledge Handbook of Asian Linguistics*, London: Taylor & Francis.
Rizzi, L. (1997). 'The fine structure of the left periphery', in L. Haegeman (ed), *Elements of Grammar*, Dordrecht: Kluwer.
Rizzi, L. (2014). 'Some Consequences of Criterial Freezing', in P. Svenonius (ed), *Functional Structure from Top to Toe* (pp. 19–45), New York: Oxford University Press.
Rizzi, L. (2017). 'Locality and the functional sequence in the left periphery', in A. Enoch, H. Eric, P. Genoveva, & S. Manuela (eds), *Elements of Comparative Syntax* (pp. 319-348), Berlin, Boston: De Gruyter Mouton.
Rizzi, L., & Bocci, G. (2017). 'Left periphery of the clause: primarily illustrated for Italian', in M. Everaert & H. C. van Riemsdijk (eds), *The Wiley Blackwell Companion to Syntax* (Second Edition ed.): John Wiley and Sons.
Roberts, I. G. (2019) *Parameter Hierarchies and Universal Grammar*: Oxford University Press.

Saito, M. (2015). 'Cartography and selection: Case studies in Japanese', in U. Shlonsky (ed), *Beyond Functional Structure: The Cartography of Syntactic Structures* (Vol. 10, pp. 255-274), Oxford: Oxford University Press.
Salzmann, M., Häusler, J., Bader, M., & Bayer, J. (2013). '*That*-trace effects without traces. An experimental investigation', in S. Keine (ed), *Proceedings of the 42nd Annual Meeting of the North East Linguistic Society* (Vol. 2, pp. 149-162), Amherst, MA: GLSA.
Sheehan, M., Biberauer, T., Roberts, I. G., & Holmberg, A. (2017) *The final-over-final condition: a syntactic universal*: The MIT Press.
Shlonsky, U., & Soare, G. (2011) 'Where's 'Why"?', *Linguistic Inquiry 42*(4): 651–669. doi:10.1162/LING_a_00064
Sobin, N. (1987) 'The variable status of Comp-trace phenomena', *Natural Language & Linguistic Theory 5*: 33-60.
Sobin, N. (2002) 'The Comp-Trace Effect, the Adverb Effect and Minimal CP', *Journal of Linguistics 38*(3): 527-560. Retrieved from http://www.jstor.org/stable/4176763
Sproat, R., & Shih, C. (1991). 'The Crosslinguistic Distribution of Adjective Ordering Restrictions', in C. Georgopoulos & R. Ishihara (eds), *Interdisciplinary Approaches to Language: Essays in honor of S.-Y. Kuroda*, Dordrecht: Kluwer.
Teodorescu, A. (2006). 'Adjective Ordering Restrictions Revisited', in D. Baumer, D. Montero, & S. Scanlon (eds), *Proceedings of the 25th West Coast Conference on Linguistics* (pp. 399-407), Somerville, MA: Cascadilla Press.
Thompson, L. C. (1987 [1965]) *A Vietnamese Reference Grammar*, Honolulu: University of Hawaii Press.
Thompson, S. A. (1978) 'Modern English from a typological point of view: some implications of the functions of word-order', *Linguistische Berichte 54*: 19-35.
Tran, T. (2016). 'Lone contrastive topic construction: A puzzle from Vietnamese', in M. Grubic & A. Mucha (eds), *Proceedings of The Semantics of African, Asian and Austronesian Languages* (Vol. 2, pp. 52–64), Potsdam: Potsdam University Press.
Tran, T. (2024). 'Syntax-information structure interface in Vietnamese', in E. McCready, H. Nomoto, & Y. Takubo (eds), *Discourse Particles in Asian Languages*, Oxford: Routledge.
Travis, L. (1984) *Parameters and Effects of Word Order Variation.* (PhD dissertation). MIT.